Health technology assessment of diagnostic strategies for Alzheimer's disease

Health technology assessment of diagnostic strategies for Alzheimer's disease

2

Health technology assessment of diagnostic strategies for Alzheimer's disease

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de Universiteit Maastricht, op gezag van Rector Magnificus, Prof. Dr. L.L.G. Soete volgens het besluit van het College van Decanen, in het openbaar te verdedigen op donderdag, 11 september 2014 om 14:00 uur

door

Ron Leonardus Hubertus Handels

Geboren op 15 maart 1986 te Kerkrade

Promotores:

Prof. Dr. F.R.J. Verhey

Prof. Dr. J.L. Severens (Erasmus Universiteit Rotterdam)

Co-promotores:

Dr. M.A. Joore

Dr. C.A.G. Wolfs

Beoordelingscommissie:

Prof. Dr. C. Dirksen, voorzitter

Prof. Dr. M. van Buchem

Dr. L. Jönsson (Karolinska Institutet, Stockholm)

Prof. Dr. J. Kleijnen

Prof. Dr. R. Ponds

"Diagnosis is not an end in itself. In general, medicine is directed toward the goal of improved health outcome."

Harvey V. Fineberg

6

Contents

"In God we trust; all others must bring data."

W. Edwards Deming

CHAPTER 1

GENERAL INTRODUCTION

With a spectrum of available diagnostic test outcome measures that do not reflect a patient's health state it is difficult to see wood for the trees: should biomarkers for Alzheimer's disease (AD) be adopted in clinical practice for testing subjects with cognitive disorders?

Burden of disease

AD causes the cognitive disorder dementia which has a tremendous impact on a human's capacity to live independent. It is characterised by multiple cognitive deficits (memory impairment and at least one other cognitive disturbance of aphasia, apraxia, agnosia, or a disturbance in executive functioning) that cause social or occupational dysfunction and constitute a decline from a previous performance level [1]. Furthermore, dementia affects the caregiving role of families. This role tends to advance from support for household, financial and social activities to personal care and almost constant supervision and surveillance [2]. Prince et al. [3] estimated that 35.6 million people lived with dementia worldwide in 2010 and expected this to double in the next 20 years. Worldwide it caused 0.8% of all years lost due to death or equivalent healthy years lost due to disability in 2004 [4]. This proportion increased to 4.1% in persons aged 60 or older and on average dementias caused 7.4 years lived with disability in high income countries [2]. The associated societal economic impact was estimated at US$604 billion in 2010 [5].

Prior to developing dementia, people can suffer from the cognitive disorder Mild Cognitive Impairment (MCI), often defined by an objective memory impairment for one's age, preserving general cognitive function, and with intact functional activities [6]. Each year, 9.6% of the patients diagnosed with MCI in a clinical setting progress to dementia [7]. However, the timing is uncertain and a proportion improves or dies before dementia develops making an accurate individual prognosis difficult.

Alzheimer's disease

AD is considered the most prevalent cause of dementia syndromes. It is pathologically defined by an abundance of Amyloid plaques consisting of aggregated Amyloid-Beta proteins and neurofibrillary tangles consisting of hyperphosphorylated tau (p-tau) proteins. Evidence suggests that the build-up of these proteins is associated with neuronal injury and eventually cognitive failures and dementia [8,9].

Clinical diagnostic criteria, set up by the NINCDS-ADRDA [10], define the clinical diagnosis of probable AD by the presence of dementia and a progressive worsening of memory and other cognitive functions. AD can only be concluded after other causes of these symptoms are ruled out (especially non-neurodegenerative conditions such as cerebrovascular disease, Huntington's disease, vitamin B12 deficiency, alcohol, Major Depressive Disorder, or delirium). A definite diagnosis of AD requires confirmation of probable AD upon neuropathological examination. Current treatment with Cholinesterase inhibitors and Memantine targets dementia symptoms in AD but the underlying pathological progression remains unaffected.

Advances in diagnostic criteria

Since the publication of the NINCDS-ADRDA criteria in 1984 [10] much research has focussed on identifying Amyloid plaques and Amyloid-Beta proteins in vivo in the cognitive disorders MCI and dementia using biomarkers in Cerebrospinal Fluid (CSF), Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI). These markers are considered important for the development of new therapies to alter the hypothesised fundamental causes of AD before conversion to dementia limits a patient's independent functioning.

The possible applications of biomarkers vary, such as a risk factor screening test, prognostic test or test to decide therapy or therapy response. In the past decade, new diagnostic research criteria have been proposed to make a diagnosis of MCI due to AD [11–13]. They have positioned biomarkers in addition to the current clinical diagnostic steps in practice diagnostic (i.e. history taking and neuropsychological examination) to increase or decrease the likelihood of AD pathology as the cause of the cognitive disorder. The research criteria hypothesise that there will be added value in the prognosis of cognitive and functional decline or in deciding on therapies when effective treatments become available in MCI. Despite the research status of these biomarkers they are increasingly being applied in clinical practice, especially CSF [14]. It is therefore important to evaluate the clinical utility of these biomarkers before adoption in clinical practice can be decided.

Comparative options under evaluation

The prognostic value of CSF in MCI has been well established in cohort studies measuring the hazard ratio or odds ratio, or the probability of AD type dementia conversion [15]. However, most studies did not research the added diagnostic value of CSF to current clinical tests. Furthermore, the studies were limited regarding their reflection of the clinical practice by a few single measures such as the Mini-Mental State Examination (MMSE) [16–18] or a neuropsychological test or combination of such tests [16,19–21]. This provides a high diagnostic window of opportunity because the information available in standard clinical practice is not used optimally, which increases the probability of overestimating the value of CSF. This poor methodology, alongside with a lack of reporting blinding and dealing with intermediate values, was highlighted in a systematic review by Noel-Storr et al. [22]. This urges the need for studies that compare the current clinical diagnostic workup to a diagnostic workup with CSF explicitly positioned within it, either to triage patients, to replace a current test, or as an add-on [23]. Furthermore, it should rely on a reproducible independent diagnostic reference test. To ensure generalizability to clinical practice, such research should be performed within a clinical sample consisting of cognitive disorders with various neurodegenerative (e.g. AD, vascular, Lewy Bodies or mixed pathologies) and non-neurodegenerative causes.

Patient-important outcomes of non-medical consequences

A high diagnostic accuracy of a biomarker does not necessarily prove that it is ready for translation into practice as a clinical test. It should demonstrate improved patient-important outcomes such as survival, cognitive or functional abilities, reduced anxiety, or increased health-related quality of life. The entire test-treatment pathway must be evaluated to measure how a

test outcome guides clinical decision-making and treatment management in order to preserve or restore a patient's health [24–26]. Aside from medical outcomes, the test result information itself can be of value to improve a patient's non-medical health consequences at a cognitive, emotional, social, or behavioural level [27,28]. Iliffe et al. [29] have described such consequences qualitatively as the advantage of reducing uncertainty, excluding remediable causes, planning support, avoiding crisis and opportunities to make appropriate legal arrangements. However, anxiety, depressive reaction, labelling and stigma have been identified as hazards of an early diagnosis. The patient's health benefits can only be established by comparing the medical and non-medical consequences of a biomarker-driven diagnostic pathway to its best alternative. A diagnostic test is no different to other medical technologies and its benefit to medical and non-medical health outcomes must be proven before adoption in practice is appropriate.

Early health technology assessment of biomarkers

Health Technology Assessment (HTA) takes the evaluation of biomarkers beyond the level of individual patients [25]. It provides a decision making framework to evaluate the health impact on society as a whole as well as the societal economic, social and ethical implications of a medical technology in order to support decision making regarding reimbursement of the use of biomarkers in health practices using societal resources. Cost-effectiveness analysis, being one part of HTA, can aid in the efficient allocation of the limited care resources available to maximise society's health. It expresses resource use in monetary terms and effectiveness in terms of medical and non-medical consequences for a patient's health, often quantified as health-related quality adjusted life years (QALYs). QALYs represent a person's remaining life years adjusted for the quality of life, which may be affected by a disease. This quality of life is expressed as a utility score which represents the preference of a (general) population for a certain health state, often measured by a scale that measures the ability to perform daily actions and disease symptoms. The choice between alternative courses of action is based on the most efficient ratio between QALYs and the costs that could be allocated to other opportunities. Such evidence is generated by cost-utility analyses and allows the comparison of interventions for various diseases.

The increasing use of CSF in MCI patients in clinical practice, in the absence of treatment options and with non-medical consequences that have not yet been quantified, urges the need for an early technology assessment [30] to estimate the societal cost-effectiveness of CSF in order to control the process of diffusion towards efficient organisation of clinical practice. This early assessment can also set incentives for research to focus on particular biomarker scenarios with potential cost-effectiveness.

Decision-analytic modelling

A randomised controlled trial with a follow-up period that captures all the relevant consequences of the test-treatment pathway including treatment effects would generate the highest level of evidence. Such studies are rarely performed due to the high demand on resources to monitor patients for a long time, during which the withholding of possible beneficial treatment is unethical. In such case, decision modelling is an alternative tool that can be used to retrieve cost-effectiveness estimates. This can also facilitate early evaluation by exploring the consequences of

hypothetical likely future scenarios, for example with an AD-specific disease-modifying treatment available of which the efficacy is currently unknown. A decision model combines various sources of evidence using a set of mathematical relationships. It quantifies the links between test accuracy, medical and non-medical treatment outcomes and the related consequences for resource use and impact on QALYs. Such links can be based on various sources of evidence, including randomised controlled trial reports, reviews or expert opinion. By simulating patient cohorts, a decision model makes it possible to estimate the likelihood of each consequence and its corresponding costs and effects [31,32].

Various AD disease decision models were built mainly to evaluate treatment interventions in dementia [33,34]. Decision models that capture the consequences of AD diagnostics in MCI are few in numbers and less extensive. Data on long-term disease progression, starting in the MCI phase are scarce, and studies have mostly relied on selective trial population estimates while future treatment will likely be applied in MCI patients from the general population. Furthermore, the consequences for care use, informal care and quality of life show high variation over a spectrum of patient disease characteristics as well as the social context characteristics for both patient and caregiver. This urges the need to improve functions that describe disease progression, resource use and quality of life estimates in order to develop decision models that are able to evaluate the consequences of MCI biomarker interventions. Diagnostic accuracy studies that represent clinical practice, in combination with health technology assessment results generated by an improved decision model, can reveal wood for the trees when deciding whether or not biomarkers for AD should be adopted in clinical practice.

THESIS AIM AND OUTLINE

The general aim of this thesis is to evaluate the prognostic accuracy of CSF and the cost-utility of CSF in the context of diagnostic workup and treatment in cognitive disorders.

The research was performed within the LeARN study in which patients who visited a memory clinic for their memory complaints were followed for 2 years. The design, methodological considerations and choices of measurement scales to enable the studies of prognostic accuracy and cost-utility are described in **Chapter 2**. Two thesis research questions are answered by the LeARN study:

RESEARCH QUESTION PART I: WHAT IS THE PROGNOSTIC VALUE OF CSF BIOMARKERS IN COGNITIVE DISORDERS?

Recent methodological studies in the field of diagnostic research expose the limited reference test in current studies, which urges the need for an improved reproducible independent reference standard before the prognostic value of an AD biomarker can be established.

> a) *How can the effectiveness of an AD biomarker be assessed in absence of a gold standard?*

A reference standard protocol was developed in **chapter 3**. It describes the methodological considerations and the design of a procedure to reach consensus in an expert panel to deal with the relatively short 2-year follow up period of the LeARN study and to optimise its relevance for

clinical practice. This was based on a pilot study using 11 subjects from the memory clinic and resulted in a recommended protocol for the evaluation of biomarkers from a clinical perspective.

> b) *Has CSF added prognostic value to the current clinical practice prognosis?*

The recommended protocol was applied in the LeARN subjects in **chapter 4**. The results of this chapter compare the current clinical diagnostic workup to a version of the current diagnostic workup with CSF explicitly positioned within it as an add-on test. This study optimally reflects current practice in contrast to current studies.

An early cost-effectiveness study is required to decide whether the diagnostic accuracy gained by using CSF from part I improves the balance between (medical and non-medical) patient-important outcomes and the care resources required.

RESEARCH QUESTION PART II: HOW CAN A COST-EFFECTIVENESS MODEL TO EVALUATE THE COST-UTILITY OF A CSF BIOMARKER AND SUBSEQUENT TREATMENT IN COGNITIVE DISORDERS BE DEVELOPED, POPULATED AND ANALYSED?

Before the model was built the general and methodological characteristics of current available health economic evidence were evaluated.

> c) *How can the cost-effectiveness of biomarkers for AD be studied?*

The scarcity of knowledge on cost-effectiveness studies for AD diagnostic tests was explored in a systematic review in **chapter 5**. Several recommendations aided in the development of the decision model structure as the basis for the cost-utility analysis. Evidence currently available on disease progression and the related resource use was not optimal to populate the decision model.

> d) *What is the disease progression in cognitive disorders?*

Long-term symptomatic disease progression of MCI and demented persons in terms of mathematical functions for the decision-analytic model was developed in **chapter 6** on the basis of data from the population-based Kungsholmen Project (Sweden) cohort.

> e) *What care resources do subjects with cognitive disorders consume?*

Variables beyond cognition that explained the variation in the health care costs of cognitive disorders at various stages of the disease progression were determined in **chapter 7**. The data from 219 memory clinic patients from the MEDICIE study who were followed for 1 year were used.

> f) *What is the incremental cost-utility of CSF added to clinical practice when disease-modifying treatment is available?*

The development and results of the cost-utility model are described in **chapter 8**. This chapter brings together the results of the previous 3 chapters and combines them with the diagnostic accuracy evidence generated in the LeARN cohort from part I. In this chapter, the pathway of progression and the corresponding consequences for care costs and quality adjusted life years were simulated for 2000 MCI subjects, for both current clinical practice and when CSF was added to current practice.

The main findings and implications of the various chapters are discussed in **chapter 9**.

REFERENCES

[1]   American Psychiatric Association, editor. DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association; 1994.

[2]   Martin P, Jim J. World Alzheimer Report 2009. London: 2009.

[3]   Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013;9:63–75.e2.

[4]   World Health Organization. The global burden of disease 2004 update. World Health Organization; 2008.

[5]   Wimo A, Jönsson L, Bond J, Prince M, Winblad B. The worldwide economic impact of dementia 2010. Alzheimers Dement 2013;9:1–11.e3.

[6]   Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004;256:183–94.

[7]   Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia-- meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand 2009;119:252–65.

[8]   Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010;9:119–28.

[9]   Jack CR, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol 2013;12:207–16.

[10]   McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS--ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;77:939–44.

[11]   Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment 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 2011;7:270–9.

[12]   Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007;6:734–46.

[13]   Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol 2010;9:1118–27.

[14]   Spies PE, Slats D, Ramakers I, Verhey FRJ, Olde Rikkert MGM. Experiences with cerebrospinal fluid analysis in Dutch memory clinics. Eur J Neurol 2011;18:1014–6.

[15]   Van Rossum IA, Vos S, Handels R, Visser PJ. Biomarkers as predictors for conversion from mild cognitive impairment to Alzheimer-type dementia: implications for trial design. J Alzheimers Dis 2010;20:881–91.

[16]   Vos S, Van Rossum I, Burns L, Knol D, Scheltens P, Soininen H, et al. Test sequence of CSF and MRI biomarkers for prediction of AD in subjects with MCI. Neurobiol Aging 2012.

[17]   Buerger K, Teipel SJ, Zinkowski R, Blennow K, Arai H, Engel R, et al. CSF tau protein phosphorylated at threonine 231 correlates with cognitive decline in MCI subjects. Neurology 2002;59:627–9.

[18]   Palmqvist S, Hertze J, Minthon L, Wattmo C, Zetterberg H, Blennow K, et al. Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up study. PLoS One 2012;7:e38639.

[19]   Richard E, Schmand BA, Eikelenboom P, Van Gool WA. MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer's disease in patients with mild cognitive impairment: a diagnostic accuracy study. BMJ Open 2013;3.

[20]   Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR, et al. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol Aging 2012;33:1203–14.

[21]   Heister D, Brewer JB, Magda S, Blennow K, McEvoy LK. Predicting MCI outcome with clinically available MRI and CSF biomarkers. Neurology 2011;77:1619–28.

[22]   Noel-Storr AH, Flicker L, Ritchie CW, Nguyen GH, Gupta T, Wood P, et al. Systematic review of the body of evidence for the use of biomarkers in the diagnosis of dementia. Alzheimers Dement 2013;9:e96–e105.

[23]   Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ 2006;332:1089–92.

[24]   Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ 2012;344:e686.

[25]   Thornbury JR, Fryback DG. Technology assessment--an American view. Eur J Radiol 1992;14:147–56.

[26]   Bossuyt PMM, Reitsma JB, Linnet K, Moons KGM. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem 2012;58:1636–43.

[27]   Bossuyt PMM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Med Decis Making 2009;29:E30–8.

[28]   Lee DW, Neumann PJ, Rizzo JA. Understanding the medical and nonmedical value of diagnostic testing. Value Health n.d.;13:310–4.

[29]   Iliffe S, Manthorpe J, Eden A. Sooner or later? Issues in the early diagnosis of dementia in general practice: a qualitative study. Fam Pract 2003;20:376–81.

[30]   Hartz S, John J. Public health policy decisions on medical innovations: what role can early economic evaluation play? Health Policy 2009;89:184–92.

[31]   Briggs AH, Claxton K, Sculpher MJ. Decision Modelling for Health Economic Evaluation. Oxford University Press; 2006.

[32]   Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. Pharmacoeconomics 2006;24:1043–53.

[33]   Cohen JT, Neumann PJ. Decision analytic models for Alzheimer's disease: state of the art and future directions. Alzheimers Dement 2008;4:212–22.

[34]   Green C. Modelling disease progression in Alzheimer's disease: a review of modelling methods used for cost-effectiveness analysis. Pharmacoeconomics 2007;25:735–50.

CHAPTER 2

DIAGNOSTIC AND ECONOMIC EVALUATION OF NEW BIOMARKERS FOR ALZHEIMER'S DISEASE: THE RESEARCH PROTOCOL OF A PROSPECTIVE COHORT STUDY

Handels RL, Aalten P, Wolfs CA, Olde-Rikkert M, Scheltens P, Visser PJ, Joore MA, Severens JL, Verhey FR.

Abstract

Background: New research criteria for the diagnosis of Alzheimer's disease (AD) have recently been developed to enable an early diagnosis of AD pathophysiology by relying on emerging biomarkers. To enable efficient allocation of health care resources, evidence is needed to support decision makers on the adoption of emerging biomarkers in clinical practice. The research goals are to 1) assess the diagnostic test accuracy of current clinical diagnostic work-up and emerging biomarkers in MRI, PET and CSF, 2) perform a cost-consequence analysis and 3) assess long-term cost-effectiveness by an economic model.

Methods/design: In a cohort design 241 consecutive patients suspected of having a primary neurodegenerative disease are approached in four academic memory clinics and followed for two years. Clinical data and data on quality of life, costs and emerging biomarkers are gathered. Diagnostic test accuracy is determined by relating the clinical practice and new research criteria diagnoses to a reference diagnosis. The clinical practice diagnosis at baseline is reflected by a consensus procedure among experts using clinical information only (no biomarkers). The diagnosis based on the new research criteria is reflected by decision rules that combine clinical and biomarker information. The reference diagnosis is determined by a consensus procedure among experts based on clinical information on the course of symptoms over a two-year time period. A decision analytic model is built combining available evidence from different resources among which (accuracy) results from the study, literature and expert opinion to assess long-term cost-effectiveness of the emerging biomarkers.

Discussion: Several other multi-centre trials study the relative value of new biomarkers for early evaluation of AD and related disorders. The uniqueness of this study is the assessment of resource utilization and quality of life to enable an economic evaluation. The study results are generalizable to a population of patients who are referred to a memory clinic due to their memory problems.

Trial registration: NCT01450891

BACKGROUND

Alzheimer's disease (AD) and other dementing disorders are common in the elderly, with a worldwide prevalence estimated in 2010 at 35.6 million, which will double every 20 years to 115.4 million in 2050. AD has a substantial impact on the person who suffers from the disease, his or her family and society [1]. AD affects a person's cognition, behavior and functional ability, and it is one of the leading causes of disability in older people living in developed countries [2].

The NINCDS-ADRDA criteria [3] are currently applied in diagnostic guidelines [4,5] to determine AD aetiology. Scientific knowledge, advanced imaging techniques and cerebrospinal fluid analyses have evolved since the publication of these criteria in 1984. This has led to much debate and the proposition of new clinical and research criteria to enhance diagnostic accuracy, even at the stage of early clinical symptoms [6-10]. These criteria distinguish between the AD pathophysiological process and the clinically observable syndrome to enable determination of AD in a pre-dementia state; e.g. mild cognitive impairment (MCI). In the end the criteria are meant to support therapy decision making (when effective treatments are available) or to determine the likelihood of cognitive and functional progression to a more severe disease state. Emerging biomarkers are attributed a more prominent role in the diagnostic criteria; Amyloid β42, total tau and phosphorylated-tau in the cerebrospinal fluid (CSF), Amyloid tracer uptake and fluorodeoxyglucose (FDG) in positron emission tomography (PET), hippocampal volume and medial temporal atrophy in structural magnetic resonance imaging (MRI) and single photon emission tomography (SPECT) perfusion imaging. However, validation of the criteria is needed before adoption of the proposed role of new biomarkers in clinical practice [9].

The ultimate goal of diagnostic testing is to guide disease management in order to improve patient outcomes and patient well-being. Tests that lack this potential should be regarded obsolete [11,12]. This has raised an urgent need for health technology assessment to address the direct, intended consequences of technologies as well as the indirect, unintended consequences for the evaluation of the value of diagnostic strategies including biomarker for AD compared to current clinical practice. Evidence is needed to support decision makers on the adoption of new diagnostic tests in clinical practice to enable efficient allocation of health care resources.

Study aim

The general aim of the study is to assess the clinical and economic value of current, emerging and novel (to be developed) techniques for an early diagnosis of AD and related disorders. In this paper the methodology is described.

The research goals are:

1. To assess the diagnostic test accuracy of the current clinical standard diagnostic work-up and emerging diagnostic biomarkers in MRI, PET and CSF

2. To assess costs and effects for the follow-up period to perform a cost-consequence and cost-effectiveness analysis

3. To develop a preliminary economic model to assess the uncertainty surrounding long-term cost-effectiveness of diagnostic strategies

METHODS/DESIGN

Study design

A cohort design was chosen because an assessment of test combinations within a randomized controlled trial would require the evaluation of many diagnostic strategies for which the number of subjects needed would exponentially increase [13]. To determine the diagnostic value of emerging biomarkers for AD and related disorders both a clinical diagnosis and diagnosis based on emerging biomarkers (index tests) are compared with a reference diagnosis. Due to limited ability of biopsy (which is unethical) or autopsy (which requires follow up until death) a two-year follow up of the clinical course is used as a proxy to obtain information on the state of the disease at baseline; so-called delayed-type cross-sectional accuracy study design [12]. Four academic memory clinics (Leiden University Medical Centre, Maastricht University Medical Centre, Radboud University Nijmegen Medical Centre and VU University Medical Centre) specialized in the diagnosis and treatment of memory disorders participate in the study. Two memory clinics are settled within a department of geriatrics, one within neurology and one within psychiatry. The study is performed within the framework of CTMM, the Center for Translational Molecular Medicine (http://www.ctmm.nl), a Dutch public-private partnership; project LeARN (grant 02 N-101).

Subjects

For the study, 241 consecutive patients of the participating memory clinics who were suspected of having a primary neurodegenerative disease were included for participating in the study from October 2009 to May 2011; this included all patients with subjective and/or objective memory complaints. Eligibility criteria were chosen to represent the current clinical situation and enable generalisability to clinical practice (see Table 1). Informed consent was obtained from both the patient and the informal caregiver. Gender, age and reason for refusal were obtained for patients unwilling to participate.

Table 1: Eligibility criteria for subject selection

Eligibility criteria
Inclusion criteria:
All new consecutive patients of the participating memory clinics who are suspected of having a primary neurodegenerative disease.
Mini-Mental State Examination: ≥ 20.
Clinical Dementia Rating: 0 – 1.
Availability of a reliable informer or proxy (who visits or contacts the patient at least once a week).
Exclusion criteria:
Normal Pressure Hydrocephalus, Huntington's disease.
Less than two years ago a transient ischaemic attack (TIA) or cerebral vascular accident (CVA) or TIA/CVA followed (within three months) by cognitive impairment.
Psychiatric history (schizophrenia, bipolar disorder or psychotic problems (not otherwise specified), less than 12 months ago).
Major Depression according to the DSM-IV, less than 12 months ago.
Cognitive problems due to excessive alcohol use (based on clinical judgement).
Brain tumour, epilepsy, encephalitis.
Probably not available for follow-up.

Data assessment

Each centre collects a minimum dataset of clinical information based upon the dataset protocol used for The String of Pearls Initiative – Pearl Neurodegenerative Diseases (http://www.string-of-pearls.org), cost data and data on emerging biomarkers. Table 2 provides an overview of all patient and informal caregiver assessments. Assessments take place at baseline and at 12 and 24 months follow up during a visit (from both patient and informal caregiver) to the memory clinic. Furthermore, several questionnaires were composed in a booklet to measure resource consumption and quality of life. This is filled out by the informal caregiver at baseline, 3, 12 and 24 months.

Table 2: Overview of patient and informal caregiver assessments at baseline and follow-up

Outcome measure	Operationalization / type of instrument	B	T3	T12	T24
Clinical data					
Demographic data	History taking	P		P	P
Cognitive impairment	Mini-mental State Examination	P		P	P
Dementia severity	Clinical Dementia Rating	P		P	P
Functional disability	Disability Assessment for Dementia	P		P	P
Neurological and physical examination	Neurological assessment and co-morbidities	P		P	P
Neuropsychiatric problems	Neuropsychiatric Inventory	P		P	P
Depression	Geriatric Depression Scale 15	P		P	P
Cerebral atrophy & white matter lesions	Structural MRI (T1 & T2-weighted and FLAIR)	P			
Neuropsychological Assessment	Rey's Verbal Learning Test, Visual Association Test, Digit-Span, Letter Digit Substitution Test, Stroop Color-Word Test, Trail Making Test	P		P	P
Quality of life data					
Patient generic quality of life	EQ-5D	I*/P	I*	I*/P	I*/P
Patient disease specific quality of life	QoL-AD	I*	I*	I*	I*
Caregiver generic quality of life*	EQ-5D	I	I	I	I
Caregiver disease specific quality of life*	QoL-AD	I	I	I	I
Caregiver burden*	Sense Of Competence	I	I	I	I
Care-related quality of life*	Carer Quality of Life	I	I	I	I
*Cost data**					
Resource utilization and caregiver time	RUD-Lite	I	I	I	I
Work productivity and absence	Productivity and Disease Questionnaire	I	I	I	I
Consequences of informal caregiving on paid or unpaid work	Health and Labour Questionnaire	I	I	I	I
Other resource use		I	I	I	I
Emerging biomarker data					
Functional connectivity	Resting state functional MRI	P			
White matter integrity	Diffusion tensor imaging	P			
Hippocampal volume	Structural MRI	P			
Glucose metabolism	Fluorodeoxyglucose PET†	P			
Amyloid plaque deposition	Pittsburgh compound B PET†, CSF Aβ1-42	P			
Tau	CSF total tau, CSF phosphorylated tau	P			

*These items were assessed by the informal caregiver by means of a booklet in which several questionnaires were composed for resource consumption and quality of life.

†Data on this item is only collected at the memory clinic of the VU University Medical Centre Amsterdam

Abbreviations: B, Baseline; FLAIR, Fluid attenuation inversion recovery; I, Information retrieved from informal caregiver; MRI, Magnetic Resonance Imaging; P, Information retrieved from patient; PET, positron emission tomography; QoL-AD, Quality of Life Alzheimer's Disease scale; RUD-Lite, Resource Utilization In Dementia; T3, 3 month follow-up measurement; T12, 12 month follow-up measurement; T24, 24 month follow-up measurement.

Clinical data

Demographic and medical information is retrieved from an open interview with both patient and informal caregiver and physical examination by a clinician.

The Mini-Mental State Examination is used to detect cognitive impairment, to assess its severity and to monitor cognitive changes over time [14]. The Clinical Dementia Rating scale (CDR) [15,16] provides a global rating of dementia severity. The Geriatric Depression Scale-15 [17] is applied to detect depression. Patient's behavioural and psychological problems are measured by the Neuropsychiatric Inventory (NPI) [18]. The Disability assessment for Dementia (DAD) is assessed to evaluate basic and instrumental activities in daily activities [19]. The information of both the NPI and DAD is obtained from a caregiver familiar with the patient's behaviour by means of a semi structured interview. Caregiver's burden of care is assessed using the disease specific Sense of Competence Questionnaire (SoCQ) [20,21].

Neuroimaging markers include medial temporal lobe atrophy measurements and white matter lesions which are qualitatively scored based on 3 T MRI scan images.

Neuropsychological examination consists of a standardized battery of cognitive tests performed by a (neuro)psychologist. Tests include Rey's Verbal Learning Test [22,23], Visual Association Test [24], and Digit-Span [25] to assess memory; Letter Digit Substitution Test [26] to assess mental processing rate; and Stroop Color-Word Test [27] and Trail Making Test [28,29] to assess attention, concentration and interference. Raw scores were converted to z-scores, adjusting for age, education and gender.

Quality of life data

Patient's generic quality of life is measured by the EQ-5D instrument. It was developed and validated in a number of European countries including the Netherlands [30-32] and it has been validated in patients with dementia [33,34]. The EQ-5D describes health status according to five three-level dimensions, which yields 243 potential combinations of health states. Each combination leads to a utility score by means of an additive function derived from the UK general population [35,36].

Patient disease specific quality of life is measured by the validated Quality of Life – Alzheimer's Disease scale (QoL–AD) [37,38]. It has 13 items covering the domains of physical health, energy, mood, living situation, memory, family, marriage, friends, self as a whole, ability to do chores around the house, ability to do things for fun, money, and life as a whole. Scale scores range from 13 to 52 with higher scores indicating greater QoL. An improvement of 3 points on the QoL-AD is judged as clinically relevant as this indicates a change of well-being on one of the domains from very poor to excellent [39].

Patient EQ-5D is assessed by the patient during the visit to the memory clinic. Furthermore, the informal caregiver judges the EQ-5D and QoL-AD for the situation of the patient and for his/her own situation and fills this out in the booklet of questionnaires.

Care-related quality of life of informal caregivers is assessed by the CarerQol [40]. It combines seven important burden dimensions with a valuation component (a visual analogue scale (VAS)) for happiness. The seven burden dimensions are 1) fulfilment; 2) relational problems; 3) mental problems; 4) problems with daily activities; 5) financial problems; 6) support; and 7) physical

problems. The CarerQol-VAS ranges from 0 ("completely unhappy") to 100 ("completely happy") and has been validated in a Dutch sample of heterogeneous caregivers.

Cost data

Cost data are retrieved by the composed booklet of questionnaires. Patient resource utilization and caregiver time, which often contains productivity losses, are assessed by means of the short version of the Resource Utilization in Dementia-questionnaire (RUD-lite). This instrument has been validated and proved to register over 95% of the costs involved in AD-care [41]. Work status, income, and productivity losses of both the patient and caregiver are assessed by the adjusted PRODISQ (PROductivity and DISease Questionnaire) [42]. The consequences of informal caregiving on paid or unpaid work are assessed by the Health and Labour Questionnaire on a two-week scale on which is indicated whether one was ill, ill by caregiving or not ill [43]. Additional questions are asked referring to the number of visits to various health care professionals, resources or aids that are bought and other out-of-pocket costs.

Emerging biomarker data

Biomarkers can be divided into two categories, one reflecting the presence of beta-Amyloid protein (Aβ) and one reflecting neuronal degeneration or injury. A pathological cascade is hypothesized in which basically Aβ markers become abnormal first, followed by neuronal injury [9,44].

The biomarkers included in this project are outlined in Table 3. CSF is collected and Amyloid β42, total tau and phosphorylated-tau are analyzed using a standardized quantitative method. FDG uptake and Pittsburgh compound B binding (PiB) on PET are both qualitatively rated by a radiologist and quantitatively analyzed by standardized methods. Whole brain and hippocampal volume, white matter integrity and functional connectivity derived by MR imaging are quantitatively analyzed by a researcher. These tests are not part of the current routine clinical diagnostic procedure. They are judged and analyzed independently and blindly. The outcome of each test is dichotomous or continuous, both for AD aetiology and progression of cognitive decline.

Table 3: Included biomarkers in the project categorized by reflecting Aβ or neuronal injury

Technique	Aβ accumulation	neuronal dysfunction
CSF	Amyloid β42 in the CSF	total tau and phosphorylated-tau in CSF
PET	Amyloid tracer uptake in PET	fluorodeoxyglucose PET
MRI		White matter integrity (DTI)
		Functional connectivity (rsfMRI)
		Whole brain volume
		Hippocampal volume

Abbreviations: Aβ, beta-Amyloid protein; CSF, cerebrospinal fluid; DTI, diffusion tensor imaging; MRI, magnetic resonance imaging; PET, positron emission tomography; PiB, Pittsburgh compound B

Baseline clinical diagnosis and reference diagnosis

Current clinical practice diagnosis is reflected by a consensus procedure among experts using baseline clinical information only, excluding any information of emerging biomarkers to prevent underestimating their accuracy.

The reference diagnosis is also determined by a consensus procedure among experts based on clinical information on the course of symptoms over a two-year time period and applying the core clinical criteria for the diagnosis of dementia due to AD [8] and core clinical criteria for the diagnosis of MCI due to AD [7]. Experts are kept blind for any information of the emerging markers under evaluation. Evaluating all cases by expert panel discussion meetings is highly time consuming. Therefore, first expert raters will assess all cases by means of an internet based form and if consensus is not reached the case will be discussed by a panel discussion meeting. The consensus diagnosis during the expert panel meetings is based on a modified Delphi method in which face-to-face discussions are held [45,46].

Analyses

Several diagnostic procedures based on current practice and emerging biomarker information (index tests) are compared to the reference diagnosis. First, AD aetiology based on the core clinical criteria [7,8] is evaluated. This is reflected by the baseline diagnosis as determined by the consensus procedure. No information of any emerging biomarker is included. Second, AD aetiology based on the research criteria as established by the National institute on aging and the Alzheimer's Association are evaluated [9]. At last, several explorative decision rules are applied including clinical information and biomarker information to determine AD aetiology and suspected progression of cognitive decline within two years.

Research goal 1: Diagnostic test accuracy

Diagnostic test accuracy is determined by relating index test results to the reference diagnosis (reference test). All diagnoses consist of a dichotomous outcome value on underlying pathology of the clinical syndrome and on (expected) progression of cognitive decline within two years. Separately for aetiology and progression of cognitive decline each index test result for each case is indicated as either true positive, true negative, false positive or false negative based on the reference diagnosis (reference test). This enables the calculation of accuracy estimates: sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio and (an increase of) the Area Under the Curve of a Receiver Operating Characteristic curve.

Finally, novel innovative diagnostic tests are developed during the course of this study by other cooperating researchers for which proof of principle does not yet exist. These tests will be evaluated as soon as evaluation in clinical subjects is possible. To assess diagnostic accuracy the tests will be applied in subgroups of the cohort retrospectively (for CSF samples) or using a case–control design.

Research goal 2: Cost-consequence and cost-effectiveness analysis

A cost-consequence analysis is performed listing all relevant costs and effects without aggregating it into a ratio [47] allowing decision-makers to choose the outcome of particular interest to include in an economic analysis. Average costs and consequences of the whole cohort are compared with the subgroup of patients who have received a correct diagnosis according to the reference test. This enables the comparison of the current diagnostic practice costs and effects and the costs and effects of an ideal situation (a costless diagnostic test with 100% accuracy). The difference indicates the maximum possible achievable benefit of new biomarkers for AD and related disorders in terms of costs and health effects. Furthermore, a cost-effectiveness analysis is performed comparing current practice with diagnostic procedures that include emerging biomarkers. The change in costs is compared to the change in diagnostic accuracy to obtain a cost-effectiveness ratio in terms of costs per correctly diagnosed patient.

Research goal 3: Decision analytic model

Finally a decision analytic model is built which provides a framework combining available evidence from different resources among which (accuracy) results from the study, literature and expert opinion. A decision analytic model can be defined as a set of mathematical relationships that form a structure reflecting the natural progression of a disease. By simulating patients or fractions of a population, these models enable the estimation of the likelihood of each consequence and its corresponding costs and effects [48]. It is applied to evaluate the short-term cost-effectiveness in terms of cost per correct diagnosis and long-term cost-utility in terms of cost per quality adjusted life year gained of diagnostic strategies under evaluation. Utility scores will be used to calculate Quality Adjusted Life Years (QALY). Patients who pass away during the year covered by the evaluation will be given a utility score of zero from the exact time of death. QALYs will be derived using the trapezium rule. A societal viewpoint will be adopted including the evaluation of all relevant costs and effects to calculate the societal benefits.

The cohort design facilitates the evaluation of many different diagnostic procedures by varying the place of a new diagnostic marker in the clinical pathway. Each procedure generates specific proportions of correct or incorrect diagnoses. Average costs and effects of correct and incorrect diagnoses are applied to calculate the total costs and effects of each procedure.

Sensitivity analysis will be performed taking into account both first order uncertainty regarding variation between patients in a homogeneous group and second order uncertainty regarding the true value of the parameters included in the model. This also enables the evaluation of an earlier diagnosis, different test sequences and the effect of possible new disease modifying drug treatments.

Sample size and missing data

A telephone interview is performed for patients who refuse follow-up assessments to determine the reason for refusal, possible cognitive decline and interference with daily activities, and to assess the CDR. Incomplete data will be imputed by means of a regression model. Complete missing data or data missing covariates will be imputed using Rubin's multiple imputation (MI) procedure.

Sample size is based on an 80% accuracy of current clinical practice to determine correct aetiology [4] and 70% in non-demented patients [49]. Applying a type I error ($\alpha = 5\%$), type II error ($\beta = 80\%$), drop-out rate of 10% and minimum clinically relevant difference of 10% accuracy increase requires 219 patients to be included in the study.

Contrary to clinical studies, economic evaluations are not based on testing hypotheses. Their goal is to assess decision uncertainty. Therefore, economic evaluations are restricted to the estimation of the uncertainty surrounding cost-effectiveness (expressed in a statistical confidence interval). Within this Bayesian framework, classical inference (and therewith a power analysis) is irrelevant [50].

Ethical considerations

According to the medical ethics committee "MedischEthischeCommissieazM/UM" the research protocol complies with the Declaration of Helsinki (October 2008, http://www.wma.net, ref.nr.: MEC 09-3-038) and with the Medical Research Involving Human Subjects Act and codes on 'good use' of clinical data and biological samples as developed by the Dutch Federation of Medical Scientific Societies.

DISCUSSION

This research protocol describes the methods used to assess the clinical and economic value of new diagnostic approaches for the diagnosis of AD. A delayed-type cross-sectional accuracy study design is chosen because a randomized clinical trial comes with ethical issues, long follow-up time and limited power. Two hundred forty one consecutive patients suspected of having a primary neurodegenerative disease are followed up for two years and a reference diagnosis is determined by an independent consensus expert panel. Eligibility criteria are chosen to maximally reflect a patient cohort within clinical practice.

Several other multi-centre trials study the relative value of new biomarkers for early evaluation of AD and related disorders. The Alzheimer's Disease Neuroimaging Initiative (ADNI) in North America is aimed to identify neuroimaging measures and biomarkers associated with cognitive and functional changes in healthy elderly subjects and in subjects who have MCI and AD [51]. Furthermore, the 'Development of Screening Guidelines and Clinical Criteria for Predementia AD' (DESCRIPA) study is aimed to develop screening guidelines for predementia AD in the general population [52]. Both include markers in PET, MR imaging and CSF. The uniqueness of this study is the assessment of resource utilization and quality of life to enable an economic evaluation. Furthermore, the decision analytic model enables the evaluation of the optimal diagnostic strategy and the evaluation of diagnostic techniques to be developed during the study in sub-cohorts of the study population. At last, without a disease modifying treatment, the added value of biomarkers is uncertain. Therefore, the availability of such treatment is explored in the sensitivity analysis.

The study has some limitations. It focuses on applying new tests for diagnostic or prognostic goals. Screening and treatment monitoring are outside the scope of this study. A follow-up period of two years was taken as a compromise to maximise the time for the disease to express symptoms of progression (to prevent false negative reference diagnoses) and to minimize the

time to prevent the start of a new disease episode after the baseline assessment (to prevent false positive reference diagnosis). It may take up to 10 years before all symptoms of dementia come to expression in subjects with AD pathology [53].

The study results are generalizable to a population of patients who are referred to a memory clinic of a university medical centre due to their memory problems.

COMPETING INTERESTS

The author(s) declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

CW designed the study and assisted RH in drafting the manuscript. FV (project leader) designed the study. JS (project leader) designed the study. MOR designed the study. PA (project coordinator) designed the study. PJV designed the study. PS designed the study. RH designed the study and drafted the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

REFERENCES

1. Prince M, Jackson J: World Alzheimer Report. 2009. http://www.alz.co.uk/research/worldreport/.
2. Wimo A, Winblad B, Jonsson L: The worldwide societal costs of dementia: Estimates for 2009. Alzheimers Dement 2010, 6(2):98–103.
3. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984,34(7):939–944.
4. Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, Small GW, Miller B, Stevens JC: Practice parameter: diagnosis of dementia (an evidence-based review), Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001, 56(9):1143–1153.
5. Waldemar G, Dubois B, Emre M, Georges J, McKeith IG, Rossor M, Scheltens P, Tariska P, Winblad B: Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. Eur J Neurol 2007, 14(1):e1–e26.

6.   Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger- Gateau P, Delacourte A, Frisoni G, Fox NC, Galasko D, et al: Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol 2010, 9(11):1118–1127.

7.   Albert MS, Dekosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, et al: The diagnosis of mild cognitive impairment 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 2011, 7(3):270–279.

8.   McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, et al: The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011, 7(3):263–269.

9.   Jack CR Jr, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, Thies B, Phelps CH: Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011, 7(3):257–262.

10. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr, Kaye J, Montine TJ, et al: Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011, 7(3):280–292.

11. Van den Bruel A, Cleemput I, Aertgeerts B, Ramaekers D, Buntinx F: The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. J Clin Epidemiol 2007, 60(11):1116–1122.

12. Knottnerus JA, Buntinx F: The evidence base of clinical diagnosis: theory and methods of diagnostic research. 2nd edition. Oxford Hoboken, NJ: Wiley- Blackwell Pub./BMJ Books; 2009.

13. Schaafsma JD, van der Graaf Y, Rinkel GJ, Buskens E: Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness. J Clin Epidemiol 2009, 62(12):1248–1252.

14. Cushman LA, Scherer MJ: Psychological assessment in medical rehabilitation. 1st edition. Washington, DC: American Psychological Association; 1995.

15. Morris JC: The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993, 43(11):2412–2414.

16. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL: A new clinical scale for the staging of dementia. Br J Psychiatry 1982, 140:566–572.

17. Almeida OP, Almeida SA: Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. Int J Geriatr Psychiatry 1999,14(10):858–865.

18. Cummings JL: The neuropsychiatric inventory: Assessing psychopathology in dementia patients. Neurology 1997, 48(5):S10–S16.

19. Gelinas I, Gauthier L, McIntyre M, Gauthier S: Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. Am J Occup Ther 1999, 53(5):471–481.

20. Vernooij-Dassen MJ, Felling AJ, Brummelkamp E, Dauzenberg MG, van den Bos GA, Grol R: Assessment of caregiver's competence in dealing with the burden of caregiving for a dementia patient: a Short Sense of Competence Questionnaire (SSCQ) suitable for clinical practice. J Am Geriatr Soc 1999, 47(2):256–257.

21. Vernooij-Dassen MJ: Dementia and Homecare: Determinants of the sense of Competence of Primary Caregivers and the effect of Professionally Guided Caregiver Support. Lisse: Swets & Zeitlinger; 1993.

22. Brand N, Jolles J: Learning and retrieval rate of words presented auditorily and visually. J Gen Psychol 1985, 112(2):201–210.

23. Rey A: L'examen clinique en psychologie. Oxford England: Presses Universitaries De France; 1958.
24. Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C: Visual association test to detect early dementia of the Alzheimer type. J Neurol Neurosurg Psychiatry 2002, 73(2):126–133.
25. Wechsler D: Wechsler Memory Scale – 3rd edition: Administration and Scoring Manual. San Antonio: The Psychological Corporation; 1997.
26. Natu MV, Agarwal AK: Digit Letter Substitution Test (DLST) as an alternative to Digit Symbol Substitution Test (DSST). Human Psychopharmacology: Clinical and Experimental 1995, 10(4):339–343.
27. Stroop JR: Studies of interference in serial verbal reactions. J Exp Psychol 1935, 18(6):643–662.
28. Reitan RM: Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958, 8:271–276.
29. Reitan RM: Trail Making Test: Manual for administration, scoring, and interpretation. Bloomington: Indiana University; 1956.
30. Brooks R: EuroQol: The current state of play. Health Policy 1996,37(1):53–72.
31. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ: Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. Ned Tijdschr Geneeskd 2005, 149(28):1574–1578.
32. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ: The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006, 15(10):1121–1132.
33. Jonsson L, Andreasen N, Kilander L, Soininen H, Waldemar G, Nygaard H, Winblad B, Jonhagen ME, Hallikainen M, Wimo A: Patient- and proxyreported utility in Alzheimer disease using the EuroQoL. Alzheimer Dis Assoc Disord 2006, 20(1):49–55.
34. Wolfs CA, Dirksen CD, Kessels A, Willems DC, Verhey FR, Severens JL: Performance of the EQ-5D and the EQ-5D + C in elderly patients with cognitive impairments. Health Qual Life Outcomes 2007, 5:33.
35. Dolan P: Modeling valuations for EuroQol health states. Med Care 1997,35(11):1095–1108.
36. Brooks RG, Rabin R, De Charro F: The measurement and valuation of health status using EQ-5D: a European perspective: evidence from the EuroQol BIOMED Research Programme. Dordrchet; Boston: Kluwer Academic Pub; 2003.
37. Logsdon RG, Gibbons LE, McCurry SM, Teri L: Assessing quality of life in older adults with cognitive impairment. Psychosom Med 2002,64(3):510–519.
38. Thorgrimsen L, Selwood A, Spector A, Royan L, de Madariaga Lopez M, Woods RT, Orrell M: Whose quality of life is it anyway? The validity and reliability of the Quality of Life-Alzheimer's Disease (QoL-AD) scale [comment]. Alzheimer Dis Assoc Disord 2003, 17(4):201–208.
39. Spector A, Thorgrimsen L, Woods B, Royan L, Davies S, Butterworth M, Orrell M: Efficacy of an evidence-based cognitive stimulation therapy programme for people with dementia: randomised controlled trial [comment]. Br J Psychiatry 2003, 183:248–254.
40. Brouwer WB, van Exel NJ, van Gorp B, Redekop WK: The CarerQol instrument: a new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. Qual Life Res 2006,15(6):1005–1021.
41. Wimo A, Winblad B: Resource utilization in dementia. Brain & Aging 2003,3:48–60.
42. Koopmanschap MA, Meerding WJ, Evers S, Severens J, Burdorf A, Brouwer W: Handleiding voor het gebruik van PRODISQ versie 2.1 (PROductivity and DISease Questionnaire). Een modulaire vragenlijst over de relatie tussen ziekte en productiviteitskosten. Toepasbaar bij economische evaluaties van gezondheidszorgprogramma's voor patiënten en werknemers. Rotterdam/Maastricht. 2004.
43. van Roijen L, Essink-Bot ML, Koopmanschap MA, Bonsel G, Rutten FF: Labor and health status in economic evaluation of health care. The Health and Labor Questionnaire. Int J Technol Assess Health Care 1996, 12(3):405–415.

44. Jack CR Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ: Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010, 9(1):119–128.

45. Gabel MJ, Foster NL, Heidebrink JL, Higdon R, Aizenstein HJ, Arnold SE, Barbas NR, Boeve BF, Burke JR, Clark CM, et al: Validation of consensus panel diagnosis in dementia. Arch Neurol 2010, 67(12):1506–1512.

46. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE: A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care 1986, 2(1):53–63.

47. Mauskopf JA, Paul JE, Grant DM, Stergachis A: The role of costconsequence analysis in healthcare decision-making. Pharmacoeconomics 1998, 13(3):277–288.

48. Briggs AH, Claxton K, Sculpher MJ: Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

49. Visser PJ, Scheltens P, Verhey FR: Do MCI criteria in drug trials accurately identify subjects with predementia Alzheimer's disease?J Neurol Neurosurg. Psychiatry 2005, 76(10):1348–1354.

50. Claxton K: The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ 1999,18(3):341–364.

51. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack C, Jagust W, Trojanowski JQ, Toga AW, Beckett L: The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin N Am 2005, 15(4):869–877. xi-xii.

52. Visser PJ, Verhey FR, Boada M, Bullock R, De Deyn PP, Frisoni GB, Frolich L, Hampel H, Jolles J, Jones R, Minthon L, Nobili F, Olde Rikkert M, Ousset PJ, Rigaud AS, Scheltens P, Soininen H, Spiru L, Touchon J, Tsolaki M, Vellas B, Wahlund LO, Wilcock G, Winblad B: Development of screening guidelines and clinical criteria for predementia Alzheimer's disease. The DESCRIPA Study. Neuroepidemiology 2008, 30(4):254–265.

53. Visser PJ, Kester A, Jolles J, Verhey F: Ten-year risk of dementia in subjects with mild cognitive impairment. Neurology 2006, 67(7):1201–1207.

PART I

PROGNOSTIC VALUE OF CSF BIOMARKERS IN COGNITIVE DISORDERS

CHAPTER 3

OPTIMIZING THE USE OF EXPERT PANEL REFERENCE DIAGNOSES IN DIAGNOSTIC STUDIES OF COGNITIVE DISORDERS

Ron L.H. Handels, Claire A. G. Wolfs, Pauline Aalten, Patrick M.M. Bossuyt, Manuela A. Joore, Albert F.G. Leentjens, Johan L. Severens, Frans R. J. Verhey

Abstract

Background: In the absence of a gold standard, a panel of experts can assign a reference standard diagnosis for use in research. Available literature offers limited guidance on assembling and working with an expert panel for this purpose. This study aimed to develop a protocol for an expert panel consensus diagnosis, and evaluate its applicability in a pilot project.

Methods: An adjusted Delphi method was used, which started with the assessment of clinical vignettes by each expert individually, followed by a consensus discussion to solve diagnostic discrepancies. A panel facilitator ensured that all participants were enabled to express their views, and encouraged the use of argumentation to arrive at a specific diagnosis, until consensus was reached by all experts. Eleven vignettes of patients suspected of having a primary neurodegenerative disease were presented to the experts. Clinical information was provided stepwise and included medical history, neurological, physical and cognitive function, brain MRI scan, and follow-up assessments over two years. After the consensus meeting, the procedure was evaluated by the experts.

Results: The average degree of consensus for the reference standard increased from 52% after individual assessment of the vignettes to 94% after the consensus discussion. Average confidence in the diagnosis was 85% (after individual assessment) and did not increase after the consensus discussion. The process evaluation led to several recommendations for improvement of the protocol.

Conclusion: A protocol for attaining a reference diagnosis based on expert panel consensus was shown feasible in research practice.

BACKGROUND

Evidence on diagnostic test accuracy is often produced in cross-sectional studies by comparing the results of the index test with the actual presence or absence of a target condition [1]. Ideally, the presence of the target condition is evaluated using a gold standard, a test that provides an error-free classification in all patients, regardless of their index test result, within a short interval of time [2]. There are, however, many diseases for which such a gold standard test does not exist. The alternative is then to rely on a clinical reference standard: the best available method for evaluating the presence of the condition of interest.

Examples of such a reference standard include watching for clinical manifestations of the disease, using structured follow-up information and using the outcome of post-mortem pathological examination. Yet there are conditions for which these reference standards do not apply. An option for diseases defined by multiple dimensions, for which no single measure or procedure exists, is to use a panel of experts, to identify those with the target condition among the persons being tested [2, 3].

Several studies have reported on the use of such an expert panel to assign a final diagnosis [4–12]. Unfortunately, there are no standardized guidelines on how to assemble and use such an expert panel. We report on a case study in which we developed a protocol for an expert consensus diagnosis, and evaluated its feasibility in a pilot project. Based on our findings, we provide a number of recommendations for other researchers considering the use of a panel diagnosis.

METHODS

Study context

The study was performed in the context of diagnostic research into Alzheimer's disease (AD) [13]. New diagnostic research criteria [14–16] enable a diagnosis of AD in which biomarkers are given a prominent role. There is, however, no gold standard for AD to evaluate the diagnostic accuracy of these biomarkers, and several reference standards have been applied and criticized. A post-mortem neuropathological examination has been criticized for imperfect interobserver reliability and imperfect association with cognitive impairment or dementia [17, 18]. Following up a patient until a clinical diagnosis of AD can be made [19, 20] requires a long time, especially in the pre-dementia or pre-clinical phase, to ensure that all patients with a neurodegenerative disease at baseline decline to the level of dementia within the follow-up period. AD is defined by the patient being demented (interference with the ability to function in daily activities), a gradual onset of symptoms over months to years, deterioration of cognition and no evidence of another neurological or non-neurological medical comorbidity or use of medication that could affect cognition [15]. These elements are mainly based on clinical judgement. The concept of AD is essentially multi-dimensional for which no single measure or pre-defined decision rules exist. Such conditions make an expert panel diagnosis the best available alternative.

Gabel et al. [21] validated an AD expert panel consensus diagnosis and concluded that it provides sufficient accuracy when neuropathological data is unavailable, and is more accurate than individual expert diagnoses. However, several choices made when applying this method in research practice remain underexposed, making it difficult to reproduce. These choices refer to the basic approach, the number and choice of experts that should be invited, the information that

must be supplied to enable expert consensus, the specific questions to be asked, and how to arrive at consensus.

Study design

Questions relevant for the development of a consensus diagnosis protocol were identified by a group of experts (RH, FV, MJ, JS and CW). Guidance on these questions was searched for non-systematically in the PubMed database. A protocol was drafted based on the recommendations from the literature, and tested in a pilot study.

A single expert panel was used to assign both the 'care as usual' clinical diagnosis and the reference diagnosis. The panel consisted of 3 experts with complementary expertise on neurology, geriatrics and psychiatry, with clinical experience ranging from 1 to more than 10 years. An adjusted Delphi method was applied which started with the assessment of each case by each expert individually, followed by a group discussion meeting to resolve all diagnostic discrepancies. The overall process was the 4-step approach graphically presented in Figure 1.

Figure 1: Process flow of the consensus protocol.
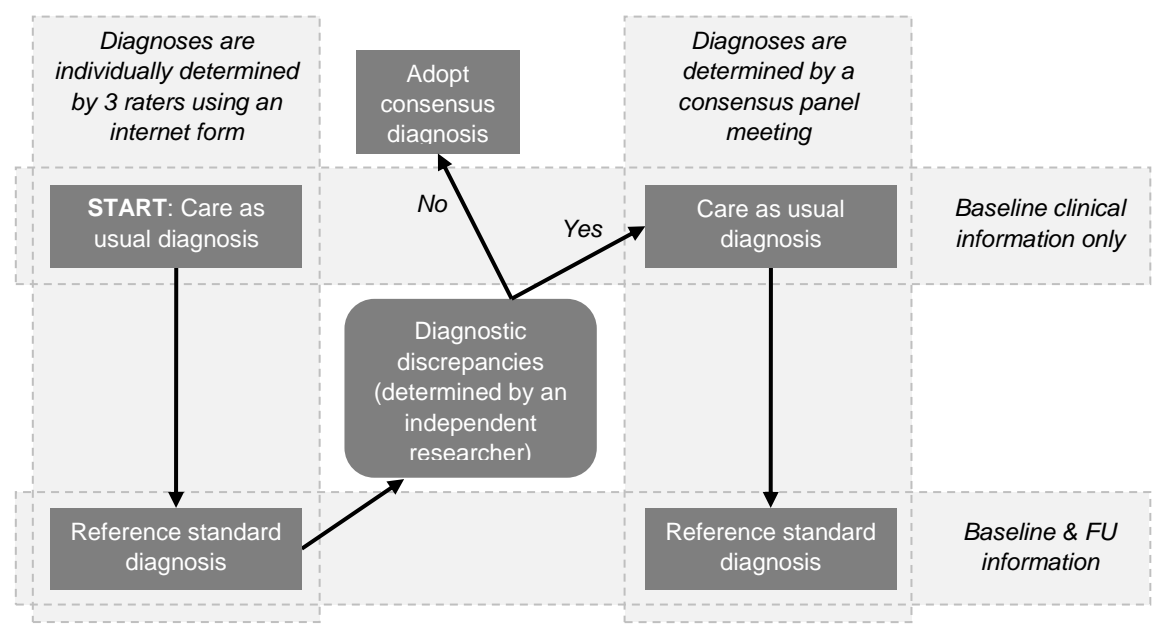
Abbreviations: FU, follow-up.

First, experts were invited to assess each case individually by logging on to a web-based questionnaire. Information on each case was provided in a stepwise fashion. In the stage 1, medical history, neurological and physical examination findings, psychiatric and clinimetric assessments were summarized in a clinical vignette and presented as concisely as possible in tabular format. In stage 2, neuropsychological test results were added to the information. In stage 3, results of an MRI scan of the brain were added. No information on other biomarkers, such as cerebrospinal fluid markers or positron emission tomography scans, was provided, as this could result in context bias [22]. In each stage, the expert was asked to individually rate the syndromal diagnosis, the aetiological diagnosis, and the expected course of the syndrome (improvement,

stable, or decline). The level of diagnostic certainty had to be indicated for each question (see table 1). After each stage, the answers were frozen; these could not be adjusted during the next stages. In the fourth stage, clinical, clinimetric and neuropsychologic information on the course of the symptoms was provided. In this fourth stage, the experts were asked to establish a diagnosis based on all the information, including the course of the symptoms. The diagnostic questions asked were the same, except for the last one 'What would be the most likely course of decline', which was changed to 'what was the course of decline'. This fourth diagnosis was considered the reference diagnosis.

In the second step, consensus was identified by an independent researcher if all experts had separately reached the same conclusion for syndromal diagnosis, aetiological diagnosis and prognosis for both the 'care as usual' and reference diagnoses. Levels of certainty were averaged. If a discrepancy occurred, all 3 experts were invited to express their arguments and to reach consensus during a group discussion at a face-to-face panel meeting. Only the diagnoses in which a discrepancy occurred were taken to this next step; if there was only a discrepancy in the aetiology of the baseline diagnosis, the other diagnostic items (syndrome and decline) were not discussed at the panel meeting.

In the third step, a summary of the diagnostic conclusion reached by the 3 experts was presented to the group during an expert panel meeting (either from the baseline diagnosis or the reference diagnosis). All experts were then once again provided individually with all relevant clinical information (identical to the internet form) and their own diagnostic conclusions. The panel members were asked to consider whether, in the light of their colleagues' assessments, they would like to alter their conclusion. A panel facilitator ensured that all participants were enabled to express their views and encouraged the use of argumentation to arrive at a specific diagnosis, until consensus by all experts was reached. No time limit was set for the discussion.

In the fourth and final step, the experts were asked to complete a questionnaire in which they were asked about their experiences regarding the assessment of cases and the consensus discussion, and were asked to provide feedback and suggestions to improve the protocol. This questionnaire is provided as appendix 1.

Table 1: Questionnaire used for rating the vignettes.

Question	Response options										
1a) What is the most probable syndrome for this patient?	☐ Subjective cognitive impairment ☐ Mild cognitive impairment ☐ Dementia										
1b) How certain are you of this?	Completely uncertain Completely certain										
	0% ☐	10% ☐	20% ☐	30% ☐	40% ☐	50% ☐	60% ☐	70% ☐	80% ☐	90% ☐	100% ☐
2a) What is the most probable aetiology for this patient?	☐ Alzheimer ☐ Vascular ☐ Frontotemporal ☐ Lewy Bodies ☐ Parkinson ☐ Other neurodegenerative disease, namely _____ ☐ No neurodegenerative disease, namely _____										
2b) How certain are you of this?	Completely uncertain Completely certain										
	0% ☐	10% ☐	20% ☐	30% ☐	40% ☐	50% ☐	60% ☐	70% ☐	80% ☐	90% ☐	100% ☐
3a) In your opinion, what would be the most likely course of cognitive and/or daily functioning within 2 years?	☐ Decline ☐ Stable ☐ Improvement										
3b) How certain are you of this expectation?	Completely uncertain Completely certain										
	0% ☐	10% ☐	20% ☐	30% ☐	40% ☐	50% ☐	60% ☐	70% ☐	80% ☐	90% ☐	100% ☐

For the reference diagnosis, question 3a was phrased differently: "In your opinion, what was the course of cognitive and/or daily functioning during the 2-year follow-up?
Is there evidence of progression of cognitive decline compared to baseline?"

Patient population

The cases for evaluation consisted of a sample of 11 patients who had visited the memory clinic of the Maastricht University Medical Centre in the Netherlands in 2009 and 2010 and were suspected of having a primary neurodegenerative disease according to the following eligibility criteria: Mini-Mental State Examination (MMSE) [23] score 20 or higher, Clinical Dementia Rating (CDR) [24, 25] between 0 and 1, and availability of a reliable informer or proxy. Subjects were excluded if they had normal pressure hydrocephalus, Huntington's disease, transient ischaemic attacks (TIAs) or cerebral vascular accidents (CVAs) less than two years ago, or a previous psychiatric history. Informed consent was obtained from both the patient and the informal caregiver. Subjects without any follow-up assessment (due to refusal or other reasons) were excluded from this research. The sample was selected such that it included similar proportions of patients with subjective memory complaints, mild cognitive impairment (MCI) and dementia. According to the medical ethics committee "Medisch Ethische Commissie azM/UM" this study complies with the Declaration of Helsinki (www.wma.net).

Clinical Information

The clinical information was based on guidelines from the American Academy of Neurology [26] and European Federation of the Neurological Societies [27] recommending tests to be performed for the diagnosis of AD dementia.

Patient and informant history (medical, family history, education, co-morbidities, behavioural and psychological symptoms, and activities of daily living) were retrieved from an open interview with both patient and informal caregiver. A neurological and physical examination, and assessment of co-morbidities was performed by a clinician. Clinical tests included the MMSE, CDR, The Geriatric Depression Scale-15 (GDS-15) [28], the Neuropsychiatric Inventory [29] and the Disability Assessment for Dementia (DAD) [30]. Atrophy measurements and white matter lesions were assessed on 3T MRI scan images by a neuroradiologist. Medial temporal lobe atrophy (MTA) scores, as well as Fazekas scores were used to quantify hippocampal atrophy and the severity of white matter lesions. Neuropsychological examination consisted of a standardized battery of cognitive tests administered by a psychologist neuropsychologist. Tests included Rey's Verbal Learning Test [31, 32], Visual Association Test [33], and Digit-Span [34] to assess memory; Letter Digit Substitution Test [35] to assess mental processing rate; and Stroop Color-Word Test [36] and Trail Making Test [37, 38] to assess attention, concentration and interference. Raw scores were converted to z-scores, adjusted for age, education level and gender. All assessments took place at baseline and at 12 and 24 months follow-up at the memory clinic, except the MRI scan which was performed only at baseline.

Statistical Analyses

The primary outcome consisted of the expert panel consensus diagnoses for the 11 vignettes and the level of confidence in the diagnostic conclusions. The secondary outcome was the result of a process evaluation questionnaire that was filled in by the expert panellists.

RESULTS

The patient sample included 8 male and 3 female patients, with a median age of 79 years (range: 49-87). The median MMSE score was 28 (range: 20-30), CDR scale was 0.5 (0-2) and DAD was 94% (25-100).

The panel's diagnoses for the 11 cases are presented in Table 2. The final reference syndrome diagnoses after the panel meeting were dementia in 5 cases (45%), MCI in 3, and subjective complaints in 3 others (with 100% consensus). Consensus on aetiology was 91%, with 8 AD (73%), 2 no neurodegenerative disease and 1 without consensus (2 experts indicated no neurodegenerative disease and 1 expert indicated a vascular aetiology). Cognitive and functional decline were assessed as improvement in 1 case, stable in 2 cases and decline in 7 cases (64%), while in 1 case there was no consensus.

Table 2: Percentage agreement among experts during individual assessment and consensus discussion of 11 cases.

Item	Care as usual		Reference standard	
	Internet Form	Panel meeting	Internet Form	Panel meeting
Degree of consensus (average)	70%	91%	52%	94%
Consensus on syndrome	55%	100%	55%	100%
Consensus on aetiology	82%	100%	64%	91%*
Consensus on disease course	73%	73%	36%	91%
Confidence in the diagnoses (average)	76%	76%	85%	85%

* Mixed Alzheimer aetiology and vascular aetiology was scored as either of the two, to facilitate consensus with other experts

The average degree of consensus (regarding syndrome, aetiology and disease course) for the care as usual diagnosis improved from 70% (internet form) to 91% (panel meeting) and that for the reference standard from 52% to 94%. The average confidence in the individually established diagnosis was 76% for the care as usual diagnosis and 85% for the reference diagnosis, and these were not increased after the discussion meeting.

It took the experts an average of 6 minutes and 6 seconds (not including reading the instructions) to assess each case via the internet form, and 8 minutes and 38 seconds to discuss a discrepancy in the assessment of a case during the expert meeting.

Table 3 presents the results of the process evaluation questionnaire. Instructions, procedure and diagnostic questions were clear, except for the difference between the diagnostic question about the expected decline that was asked for the care as usual diagnosis and the question about the actual decline that was asked for the reference diagnosis (indicated by 1 expert). Insufficient clinical information was reported to be available for several reasons: one expert would have preferred information on clinical history at follow-up, two experts stated that they would have liked to have a 'real' clinical picture or to see the patient in real life. The experts indicated that their reference diagnosis was partly influenced by the concluded baseline care as usual diagnosis (diagnostic review bias). Although none of the panel members felt impeded in expressing their opinion during the discussion meeting, one expert thought that not all members had had an equal share in the discussion. The experts also stated that a two year follow-up period is sometimes insufficient; it is 'a compromise between desirable and feasible' and in an ideal situation they would prefer a longer period.

Table 3: Results of the evaluation questionnaire.

Item	Result
Instructions, procedure and diagnostic questions were clear	92%
Estimated time per case to fill in the internet form	10 minutes
Diagnosis reflects clinical practice	100%
Consensus procedure considered valid to determine a reference diagnosis (scale 0-10)	7.1
Sufficient information available to determine a diagnosis	33%
Influence of baseline diagnosis on reference diagnosis (scale 0-10)	5.7
All panel members had equal shares in the discussion*	50%
Felt impeded in expressing their opinion	0%
Years of experience needed to participate in an expert panel (average, range)*	3, 1-5
Two-year follow-up on disease course is sufficient to determine a reference diagnosis	33%
Three experts is enough	100%

* One expert answered 'don't know' to the this question

During the meeting, the experts indicated that they considered the patient's history information to be decisive if it contradicted test outcomes of clinical scales such as MMSE or DAD. The experts also discussed with each other what they considered to constitute sufficient cognitive and functional decline to regard a patient as 'declining'. They indicated that a patient did not necessarily have to decline to a more severe syndrome, but could also decline to a clinically relevant lower state of cognition and/or functioning.

DISCUSSION

We tested a 4-step consensus diagnosis protocol in a pilot study using clinical vignettes based on 11 cases. The use of an expert panel to attain consensus on a reference diagnosis was considered feasible in research practice.

Two possible forms of bias can occur in the design of this type of cohort-based diagnostic test research. Incorporation bias occurs if result of the index test is used to establish the reference diagnosis [39]. If the index test is a biomarker, then making the biomarker result available to an expert panel when determining the reference diagnosis could lead to overestimation of its accuracy, because the association between biomarker and reference test result is artificially inflated. This could also occur when using the diagnosis established in clinical practice at follow-up as the reference standard, as many biomarkers are used in clinical practice.

Many reports on studies applying a longitudinal reference standard provide insufficient information on the methods used to blind participants for biomarker results when assessing the reference diagnosis [20]. Incorporation bias can also occur when the index test is the baseline clinical information (required to determine the care as usual diagnosis) and it is used to determine cognitive and functional decline at follow-up for the reference diagnosis. In contrast to biomarker results, however, this information can hardly be omitted, since a point of departure is required to determine decline. As it is likely that the baseline clinical tests provide a 'piece of information' and the biomarkers would receive 'much weight in the consensus judgement' it can be recommended to incorporate the baseline clinical information but no biomarker information in the reference diagnosis [40].

The second form of bias is review bias. Review bias could occur when the reference standard result is known while the index test is being interpreted (test review bias) or when the index test result is known when the reference standard is being interpreted (diagnostic review bias) [39]. Figures 2a and 2b graphically represent both situations as they may occur in a procedure to arrive at the three required diagnoses by a single panel of experts (e.g. to compare a biomarker diagnosis with the care as usual diagnosis using a reference diagnosis). In an ideal situation, all three diagnoses are assessed by three independent expert panels. Reasons of limited time and resources may result in decision rules being applied that combine the baseline standard clinical diagnosis with the diagnostic information under evaluation (e.g. biomarker) using cut-off values, and both the baseline and reference diagnoses are then established by an expert panel, as was evaluated in our study (see figure 2c for an overview) [2]. In our assessment of diagnostic review bias in the evaluation questionnaire, the experts rated the influence of the baseline diagnosis on the reference diagnosis at an average of 5.7 on a scale of 0 to 10.

*Figure 2: Single panel approach (A and B) and partly independent approaches (C) to evaluate diagnostic tests for AD. * = possible diagnostic review bias; † = possible test review bias. Control = standard clinical practice, Intervention = add biomarkers to clinical practice.*

A

Control
Information: clinical data (no biomarkers) retrieved at baseline (no follow-up)
Method: expert panel

→ *

Reference standard
Information: baseline clinical data and follow-up clinical data (no biomarkers)
Method: expert panel

→ †

Intervention
Information: clinical data & biomarkers all retrieved at baseline (no follow-up)
Method: expert panel

B

Control
Information: clinical data (no biomarkers) retrieved at baseline (no follow-up)
Method: expert panel

→

Intervention
Information: clinical data & biomarkers all retrieved at baseline (no follow-up)
Method: expert panel

→ *

Reference standard
Information: baseline clinical data and follow-up clinical data (no biomarkers)
Method: expert panel

C

Control
Information available: clinical data (no biomarkers) all retrieved at baseline (no follow-up). Method: expert panel

Intervention
Information: clinical data & biomarkers all retrieved at baseline (no follow-up)
Method: pre-defined decision rules to combine the standard practice diagnostic conclusion (as determined by the consensus method for the control group) with the biomarker test results

→ *

Reference standard
Information: baseline clinical data and follow-up clinical data (no biomarkers)
Method: expert panel

As has been explained above, follow-up until a diagnosis of AD is established in clinical practice is a widely applied reference standard to validate biomarkers [20]. A false-negative diagnosis might occur when the follow-up period is too short to detect patients who suffer from cognitive and functional decline, but the decline is not severe enough to develop dementia within that time period. Conversely, a false-positive diagnosis might occur when the follow-up is too long, which means that some of the control patients at baseline develop AD later on during the follow-up period. This is why we adopted the concept of decline itself, instead of decline to an absolute level of cognition and functioning. It was unclear how long patients should be followed up to determine a valid reference diagnosis. In our opinion, the optimal follow-up period is between 3 and 5 years, as the experts indicated that 2 years may not be sufficient, especially when treatment of cognitive symptoms, e.g. with cholinesterase inhibitors, is started. A follow-up longer than 5 years was considered clinically irrelevant.

The use of a heterogeneous panel composed of experts with different backgrounds, though within the area of interest, has been recommended in the literature [41, 42]. Gabel et.al [21] emphasized the importance of selecting experts who are likely to make different types of errors of judgment. Heterogeneity of backgrounds could also help prevent domination by a particular expertise. Other studies applying a consensus diagnosis have used a variety of expertise for their expert panel [9, 10]. Gabel et al. [21] found no difference in diagnostic accuracy between a trainee panel and an expert panel though did not recommend about a specific 'amount' of required experience. In our study the experts indicated in the process evaluation that 1-5 years of experience in the particular field of expertise was necessary to establish a valid diagnosis. The limited clinical experience of one of our experts might be the reason why one of the other panel members indicated that not everyone had had an equal share in the discussion. Perhaps requiring

a larger number of years of clinical experience might have helped prevent some of the panel members dominating the discussion. Previous studies showed a large variety regarding the number of experts in the panel, though no specific number was recommended [41, 43]. In our study all experts shared the opinion that having 3 panel members was sufficient to establish a valid diagnosis. The qualities of the experts were generally considered more important than the size of the expert group. An uneven number of experts could facilitate the decision process if a majority vote is used [3]. Studies reported on in the literature used 2 to 6 experts to determine a consensus diagnosis [5–10, 21].

The Delphi procedure was adjusted by not blinding the experts for each other's opinion in the second round (panel discussion) [7, 42]. Since this might cause bias towards dominant experts, a panel facilitator ensured that all participants were enabled to express their views [21, 42]. An alternative, less time–consuming, method could be to adopt a majority decision after individual assessment [5, 8, 9], which has been reported to have similar diagnostic accuracy as forced consensus [21]. Most studies we consulted suggested that experts should apply clinical practice diagnostic guidelines, though without imposing strict decision rules.

Video recordings, which could reveal valuable subtle information on patient history [3], were not provided in our study, due to limited time and resources. Instead, we provided a written summary of the clinical history in which an independent researcher had highlighted the most important aspects.

Recommendations for adjustments to the protocol

Based on the results of the evaluation questionnaire, 7 adjustments to the consensus protocol could be recommended:

- Information on clinical history at follow-up should be included to arrive at a reference diagnosis.
- A 2-year follow-up period was considered too short, though the experts could not provide a specific period required.
- Experts invited to the panel should have a minimum of 3 years of clinical experience.
- In some cases the expert panel concluded after the group discussion that no consensus could be reached. As proposed in the literature [5, 9], a majority decision could be adopted in these cases to prevent inefficient use of discussion time.
- One expert recognized 1 case, so we recommend inviting only experts who have had no direct interaction with the patients under evaluation.
- Initially we included several non-neurodegenerative diagnostic options, which resulted in irrelevant discussion. These were therefore replaced by the question: 'No neurodegenerative disease, namely…'.
- The login procedure consisted of several steps. When implementing this procedure in practice, the required time and the complexity of the procedure should be minimized to maximize the willingness of experts to devote their time.

The final protocol can be found in appendix 2.

Limitations

Several limitations apply to this study. Not imposing strict decision rules allowed for different views within the panel on how to determine a diagnosis. For example, different sources of information were used to determine an objective memory deficit to distinguish between subjective memory complaints and MCI, and some experts always expected a decline if a neurodegenerative disease was identified. A preparatory discussion among all experts might have reduced discussion time and could increase our understanding of the concept being assessed by the experts.

Another limitation is that the 3-day period that elapsed between filling in the internet form and the panel meeting may have been too short. The experts may have remembered the follow-up information from the internet-based questionnaire when discussing the care as usual diagnosis during the panel meeting, which may have resulted in test review bias.

Although the protocol we developed was a practical and transparent method to determine a reference diagnosis, it must be kept in mind that it represents a compromise between the available time and resources and a minimisation of bias. The optimal design to evaluate a diagnostic test would be a randomised controlled trial to determine the prognostic effects of the treatment initiated by the test. Although no disease-modifying therapies in the pre-dementia phase are available for AD, there is still an interest in the validity of new biomarkers to distinguish disease from non-disease or to enable future planning for patients. Also, when such treatments should become available, evidence on the level of diagnostic accuracy is required to enable correct treatment decisions to be made [44, 45].

Conclusion

Our aim was to assess the feasibility of establishing a consensus expert diagnosis, for the purpose of studies into cognitive decline and AD, and to establish a protocol for such a consensus diagnosis. Such a protocol was designed based on findings from the literature as well as the evaluation of this pilot study. The results indicate that this protocol was feasible in research practice.

COMPETING INTERESTS

The author(s) declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

AL designed the study and was one of the three clinical experts. CW designed the study and was panel facilitator for the expert panel discussion meeting. FV (project leader) designed the study and was one of the three clinical experts. JS (project leader) designed the study. MJ designed the study. PA designed the study. PB designed the study. RH designed the study, organized the expert panel meeting, performed the analysis, and drafted the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENT

REFERENCES

1. Knottnerus JA, Buntinx F: The Evidence Base of Clinical Diagnosis : Theory and Methods of Diagnostic Research. 2nd edition. Oxford ; Hoboken, NJ: Wiley-Blackwell Pub./BMJ Books; 2009:xiii, 302 p.
2. Reitsma JB, Rutjes AWS, Khan KS, Coomarasamy A, Bossuyt PM: A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. J Clin Epidemiol 2009, 62:797–806.
3. Rutjes AWS, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PMM: Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol Assess 2007, 11:iii, ix–51.
4. Laméris W, Van Randen A, Dijkgraaf MG, Bossuyt PM, Stoker J, Boermeester MA: Optimization of diagnostic imaging use in patients with acute abdominal pain (OPTIMA): Design and rationale. BMC Emerg Med 2007, 7:9.
5. Cowie MR, Struthers AD, Wood DA, Coats AJ, Thompson SG, Poole-Wilson PA, Sutton GC: Value of natriuretic peptides in assessment of patients with possible new heart failure in primary care. Lancet 1997, 350:1349–1353.
6. Maisel AS, Krishnaswamy P, Nowak RM, McCord J, Hollander JE, Duc P, Omland T, Storrow AB, Abraham WT, Wu AHB, Clopton P, Steg PG, Westheim A, Knudsen CW, Perez A, Kazanegra R, Herrmann HC, McCullough PA: Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med 2002, 347:161–7.
7. McKeith I, O'Brien J, Walker Z, Tatsch K, Booij J, Darcourt J, Padovani A, Giubbini R, Bonuccelli U, Volterrani D, Holmes C, Kemp P, Tabet N, Meyer I, Reininger C: Sensitivity and specificity of dopamine transporter imaging with 123I-FP-CIT SPECT in dementia with Lewy bodies: a phase III, multicentre study. Lancet Neurol 2007, 6:305–313.
8. Reeves MJ, Wehner S, Organek N, Birbeck GL, Mullard AJ, Jacobs BS, Kothari R, Hickenbottom S: Accuracy of Identifying Acute Stroke Admissions in a Michigan Stroke Registry. Prev Chronic Dis 2011, 8:A62.
9. Rutten FH, Moons KGM, Cramer M-JM, Grobbee DE, Zuithoff NPA, Lammers J-WJ, Hoes AW: Recognising heart failure in elderly patients with stable chronic obstructive pulmonary disease in primary care: cross sectional diagnostic study. BMJ 2005, 331:1379.
10. Whiteley WN, Wardlaw JM, Dennis MS, Sandercock PAG: Clinical scores for the identification of stroke and transient ischaemic attack in the emergency department: a cross-sectional study. J Neurol Neurosurg Psychiatry 2011, 82:1006–1010.
11. Van Dijk N, Boer KR, Colman N, Bakker A, Stam J, Van Grieken JJM, Wilde AAM, Linzer M, Reitsma JB, Wieling W: High diagnostic yield and accuracy of history, physical examination, and ECG in patients with transient loss of consciousness in FAST: the Fainting Assessment study. J Cardiovasc Electrophysiol 2008, 19:48–55.

12. Carnero-Pardo C, Espejo-Martinez B, Lopez-Alcalde S, Espinosa-Garcia M, Saez-Zea C, Vilchez-Carrillo R, Hernandez-Torres E, Navarro-Espigares JL: Effectiveness and costs of phototest in dementia and cognitive impairment screening. BMC Neurol 2011, 11:92.
13. Handels RL, Aalten P, Wolfs CA, Olde Rikkert M, Scheltens P, Visser PJ, Joore MA, Severens JL, Verhey FR: Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease: the research protocol of a prospective cohort study. BMC Neurol 2012, 12:72.
14. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH: The diagnosis of mild cognitive impairment 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 2011, 7:270–279.
15. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH: 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 2011, 7:335–337.
16. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster M V, Phelps CH: Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011, 7:280–292.
17. Scheltens P, Rockwood K: How golden is the gold standard of neuropathology in dementia? Alzheimers Dement 2011, 7:486–489.
18. Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C: Age, neuropathology, and dementia. N Engl J Med 2009:2302–2309.
19. Mason SE, McShane R, Ritchie CW: Diagnostic tests for Alzheimer's disease: rationale, methodology, and challenges. Int J Alzheimers Dis 2010, 2010:1–7.
20. Noel-Storr AH, Flicker L, Ritchie CW, Nguyen GH, Gupta T, Wood P, Walton J, Desai M, Solomon DF, Molena E, Worrall R, Hayen A, Choudhary P, Ladds E, Lanctôt KL, Verhey FR, McCleery JM, Mead GE, Clare L, Fioravanti M, Hyde C, Marcus S, McShane R: Systematic review of the body of evidence for the use of biomarkers in the diagnosis of dementia. Alzheimers Dement 2013, 9:e96–e105.
21. Gabel MJ, Foster NL, Heidebrink JL, Higdon R, Aizenstein HJ, Arnold SE, Barbas NR, Boeve BF, Burke JR, Clark CM, Dekosky ST, Farlow MR, Jagust WJ, Kawas CH, Koeppe RA, Leverenz JB, Lipton AM, Peskind ER, Turner RS, Womack KB, Zamrini EY: Validation of consensus panel diagnosis in dementia. Arch Neurol 2010, 67:1506–1512.
22. Egglin TK, Feinstein AR: Context bias. A problem in diagnostic radiology. JAMA 1996, 276:1752–5.
23. Cushman L, Scherer MJ: Psychological Assessment in Medical Rehabilitation. 1st edition. Washington, DC: American Psychological Association; 1995:471.
24. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL: A new clinical scale for the staging of dementia. Br J Psychol 1982, 140:566–572.
25. Morris JC: The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993, 43:2412–2414.
26. Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, Small GW, Miller B, Stevens JC: Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001, 56:1143–53.
27. Hort J, O'Brien JT, Gainotti G, Pirttila T, Popescu BO, Rektorova I, Sorbi S, Scheltens P: EFNS guidelines for the diagnosis and management of Alzheimer's disease. Eur J Neurol 2010, 17:1236–1248.

46

28. Almeida OP, Almeida SA: Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. Int J Geriatr Psychiatry 1999, 14:858–865.

29. Cummings JL: The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. Neurology 1997, 48(5 Suppl 6):S10–S16.

30. Gélinas I, Gauthier L, McIntyre M, Gauthier S: Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. Am J Occup Ther 1999, 53:471–481.

31. Rey A: L' Examen Clinique En Psychologie. Oxford, England: Presses universitaires de France; 1958.

32. Brand N, Jolles J: Learning and retrieval rate of words presented auditorily and visually. J Gen Psychol 1985, 112:201–210.

33. Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C: Visual association test to detect early dementia of the Alzheimer type. J Neurol Neurosurg Psychiatry 2002, 73:126–133.

34. Wechsler D: WMS-III: Wechsler Memory Scale Administration and Scoring Manual. Psychological Corporation; 1997:212.

35. Natu M V, Agarwal AK: Digit Letter Substitution Test (DLST) as an alternative to Digit Symbol Substitution Test (DSST). Hum Psychopharmacol 1995, 10:339–343.

36. Stroop JR: Studies of interference in serial verbal reactions. Can J Exp Psychol 1935, 18:643–662.

37. Reitan R: Trail Making Test: Manual for Administration, Scoring and Interpretation. Bloomington: Section of Neuropsychology, Department of Neurology, Indiana University Medical Centre; 1956.

38. Reitan RM: Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958, 8:271–276.

39. Bossuyt PM: The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration. Clin Chem 2003, 138:7–18.

40. Moons KGM, Grobbee DE: When should we remain blind and when should our eyes remain open in diagnostic studies? J Clin Epidemiol 2002, 55:633–6.

41. Powell C: The Delphi technique: myths and realities. J Adv Nurs 2003, 41:376–382.

42. Jones J, Hunter D: Consensus methods for medical and health services research. BMJ 1995, 311:376–380.

43. Keeney S, Hasson F, McKenna H: Consulting the oracle: ten lessons from using the Delphi technique in nursing research. J Adv Nurs 2006, 53:205–212.

44. Thornbury JR, Fryback DG: Technology assessment--an American view. Eur J Radiol 1992, 14:147–156.

45. Van Den Bruel A, Cleemput I, Aertgeerts B, Ramaekers D, Buntinx F: The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. J Clin Epidemiol 2007, 60:1116–1122.

APPENDIX 1: EVALUATION QUESTIONNAIRE

1. Were the instructions for the internet form clear? Yes / no, because… / don't know

2. Was your intended role in the expert panel clear? Yes / no, because… / don't know

3. Was the overall procedure clear (combination of internet form and panel discussion)? Yes / no, because… / don't know

4. Was the final reference diagnosis a reflection of clinical practice? Yes / no, because… / don't know

5. Please indicate how much time you spent filling in the internet form: ___ minutes / don't know

6. Was the login procedure time-consuming or difficult? Yes / no, because… / don't know

7. Was the consensus procedure valid to establish the syndrome? 10-point scale: Completely invalid – completely valid

8. Was the consensus procedure valid to establish the aetiology? 10-point scale: Completely invalid – completely valid

9. Was the consensus procedure valid to establish the prognosis? 10-point scale: Completely invalid – completely valid

10. Were all diagnostic questions clear? Yes / no, because… / don't know

11. Was sufficient clinical information available to establish the diagnosis? Yes / no, because… / don't know

12. Was all information transparently presented (both in the internet form and the forms used during the panel discussion)? Yes / no, because… / don't know

13. Was the sequence of information logical? Yes / no, because… / don't know

14. To what extent did the preselection of only the most important information from the patient's history and the neuropsychological conclusion influence your diagnosis? 10-point scale: Not at all – completely

15. Was it useful to have only the most relevant parts of the history and neuropsychological summary preselected? Yes / no, because… / don't know

16. When you were asked to establish the reference diagnosis, your conclusion on the baseline diagnosis was available. To what extent did this influence your reference diagnosis? 10-point scale: Not at all – completely

17. Did you use the internet during the individual diagnostic scoring? Yes, namely… / no/ don't know

18. Did you recognize any of the cases (and thus had foreknowledge about the patient)? Yes / no / don't know

19. Each case was assessed by experts from 3 disciplines (neurologist, geriatrician and psychiatrist). In your opinion, did the experts have sufficient expertise to establish a valid diagnosis? Yes / no, because… / don't know

20. Did all experts have an equal share in the discussion? Yes / no, because… / don't know

21. How many years of experience are required for an expert to establish a valid diagnosis? ___ years / don't know

22. Was a follow-up period of two years sufficient to establish a valid reference diagnosis? Yes / no, because… / don't know

23. Was the number of experts (3) sufficient to establish a valid reference diagnosis? Yes / no, because… / don't know

24. Should we have applied stricter decision rules to arrive at a diagnosis (e.g. ask the experts to strictly apply NINCDS-ADRDA or NIA / AA criteria)? Yes, namely… / no / don't know

25. Did you feel impeded in sharing your opinion in the group discussion? Yes, namely… / no / don't know

26. Do you have any suggestion for improvement of the protocol? Yes, namely… / no / don't know

27. Would you be willing to evaluate 50 more cases in a second expert panel? Yes/no, because…/don't know

28. Name (optional)

APPENDIX 2: FINAL PROPOSED PROTOCOL

Use a single panel to determine both the care as usual diagnosis and the reference diagnosis.

Do not involve the tests under evaluation in this process.

Compose a panel of three experts with complementary expertise who have had no direct interaction with the patients under evaluation.

Panellists should have three or more years of clinical experience.

Apply an adjusted Delphi method that starts with the assessment of each case by each expert individually, followed by solving all diagnostic discrepancies during a group discussion meeting.

The overall process is a 3-step approach:

- First, experts are invited to assess each case individually by logging on to a web-based questionnaire. All relevant clinical information (medical history, neurological and physical examination, cognitive assessment, imaging and assessment of co-morbidities at baseline) is presented as concisely as possible in tabular format. A written summary of the clinical history in which an independent researcher has highlighted the most important aspects is provided. Any diagnostic interpretations (diagnostic conclusions written in the clinical history or neuropsychological examination and biomarker information) are filtered from the information provided to the experts. Each expert individually answers the three diagnostic questions from table 1 and indicates their level of diagnostic certainty. The diagnosis is frozen and cannot be changed during the next steps. Next, the internet form is updated with clinical information from three- to five-year follow-up on the same aspects. The same diagnostic questions are asked except the last (what would be the most likely course in terms of decline), which becomes 'what was the course of decline'.

- Second, consensus is determined by an independent researcher if all experts separately come to the same conclusions as to the syndrome, aetiology and prognosis, for both the care as usual and reference diagnoses. Levels of certainty are averaged. If a discrepancy occurs, all three experts are invited to express their arguments and reach consensus during a panel meeting with group discussion. Only the diagnoses for which a discrepancy occurred are carried forward to the next step (e.g. if there is only a discrepancy on the aetiology of the baseline diagnosis, the other diagnostic items, syndrome and decline, are not discussed in the panel meeting).

- Third, during the expert panel discussion meeting, a summary of the diagnostic conclusions of the three experts is presented to the group (either from the baseline diagnosis or the reference diagnosis). Then, all experts are provided individually with all relevant clinical information (identical to the internet form) and their own diagnostic conclusions. The panel members are asked to consider whether, in the light of their colleagues' assessments, they would like to alter their answers. A panel facilitator ensures that all participants are enabled to express their views and encourages the use of argumentation to arrive at a specific diagnosis, until consensus by all experts is reached. No time limit is set for the discussion. If consensus cannot be reached, the majority diagnosis is adopted.

CHAPTER 4

ADDED DIAGNOSTIC VALUE OF CEREBROSPINAL FLUID IN PREDICTING DECLINE IN MEMORY CLINIC SUBJECTS IN CLINICAL PRACTICE

Ron L.H. Handels, Manuela A. Joore, Stephanie Vos, Pauline Aalten, Inez Ramakers, Marcel Olde-Rikkert, Philip Scheltens, Bart M.N. van Berckel, Albert F.G. Leentjens, Claire A. G. Wolfs, Johan L. Severens, Frans R. J. Verhey

Abstract

Introduction: An accurate prognosis of decline is essential to reduce the dementia related burden by timely access to formal advice and care. The purpose of this study is to determine the added diagnostic value of cerebrospinal fluid (CSF) to a clinical judgment for the prediction of cognitive or functional decline.

Methods: In a prospective cohort study memory clinic patients were evaluated at baseline according to current clinical standard diagnostic evaluations (physical, clinical and neuropsychological examination, history, and MRI) and received a CSF test. A diagnostic pathway 'clinic without CSF' was reflected by the judgment of patient vignettes, containing clinical information from baseline only, by three clinical experts who classified each patient as 'AD-related decline expected', 'other cause-related decline expected' or 'no decline expected'. As a comparison pathway, a CSF test was positioned as an add-on test to reclassify the clinical judgments if an abnormal Beta Amyloid(1-42)/total tau ratio was found in subjects clinically judged as 'no decline expected' or a normal ratio was found in subjects judged as 'AD-related decline'. In each subgroup (subjective memory complaints, mild cognitive impairment and dementia) reclassifications were verified by a reference diagnosis set by an adjusted Delphi consensus procedure using 2 year follow up data to observe which patients actually declined.

Results: The 'clinic without CSF' pathway correctly predicted decline in 55 out of 58 subjects (95%) who declined after 2 years and correctly predicted no decline in 34 out of 53 subjects (64%) who remained stable. When adding CSF, 25 subjects were reclassified of whom 6 were correctly and 19 incorrectly reclassified. The 6 correct reclassifications all concerned patients that were initially diagnosed as having no AD. Of the 19 incorrect reclassifications 11 were from changed 'no decline expected' to 'AD-related decline expected' and 8 AD-related to no decline expected. In all subgroups the proportion correctly reclassified due to CSF was lower than the proportion incorrectly reclassified.

Discussion: Adding the CSF Beta Amyloid total tau ratio information to the clinical judgment by a decision rule did not improve the prediction of 2-year cognitive or functional decline based on clinical evaluation in memory clinic subjects with subjective memory complaints, mild cognitive impairment or dementia. New data are expected in 2014 using a panel of experts to combine CSF to clinical information instead of relying on the decision rule that is limitedly generalizable to clinical practice.

52

INTRODUCTION

With a global prevalence of 35.6 million people and an economic impact of US$604 billion, dementia has a substantial impact on societies worldwide [1,2]. Timely disease management in demented subjects, with adequate support for patients and their caregivers, is needed to reduce dementia related burden. An accurate diagnosis is considered essential for timely access to formal advice and care [3]. In the last decades much research has focussed on in vivo identification of Alzheimer's disease (AD) pathology by Beta Amyloid and Tau markers in the Cerebrospinal Fluid (CSF) in pre-dementia stages such as mild cognitive impairment (MCI) where 5-10% [4] converts to dementia. New diagnostic research criteria were proposed to diagnose AD using CSF and markers in Positron Emission Tomography and Magnetic Resonance Imaging in dementia [5] and pre-dementia phases [6,7], but argued that much additional work is needed to validate the application of biomarkers for diagnostic purposes in clinical practice. In particular in absence of treatment possibilities in pre-dementia states effective use of biomarkers in the clinical routine requires accurate individual predictions of decline over a specific time interval. With much debate around the added value of an early biomarker testing and early testing in pre-dementia stages in general [8] it is important to reveal the prognostic added value from a clinical perspective.

So far, the studies that investigated the added value of CSF markers in addition to clinically available measures have limitedly reflected clinical practice by a single or a few measures such as the Mini-Mental State Examination (MMSE) [9–11] or a combination of neuropsychological tests [9,12–14]. This provides a high diagnostic window of opportunity because the available information in standard clinical practice is not optimally used which increases the probability to overestimate the value of CSF. Furthermore, most of these studies relied on an odds ratio outcome which is difficult to translate into practice. At last, the use of biomarkers in different clinical subgroups (i.e. subjective memory complaints (SMC), MCI and dementia) remains to be compared. The purpose of this study is to determine the added prognostic value of CSF markers as addition to the standard clinical routine that consists of patient history, physical and neurological examination, neuropsychological tests, and MRI for a 2-year prediction of overall cognitive or functional decline in a memory clinic population.

METHODS

In a cohort study, 251 consecutive patients from 3 Dutch academic memory clinics (Maastricht University Medical Centre, VU University Medical Centre, and Radboud University Medical Centre) were included if they were referred for a suspicion of a cognitive disorder between 2009-2013; MMSE >= 20; Clinical Dementia Rating (CDR) <= 1; availability of a reliable informer or proxy; and absence of neurological or psychiatric pathology other than neurodegenerative disorder associated with cognitive impairment (such as Normal Pressure Hydrocephalus, Huntington's disease, tumour, stroke, schizophrenia, major depression, excessive alcohol use) [15].

Informed consent was obtained from both the patient and the informal caregiver. The Medical Ethics Committee of Maastricht University Medical Centre approved the study (NCT01450891 [15]) and confirmed that it complied with the Declaration of Helsinki.

Clinical assessments

Assessments took place at baseline and at 1- and 2-year follow-up. Each assessment consisted of a full history of both patient and informal caregiver for clinical, demographic and medical information; physical and neurological examination; and psychiatric and neuropsychological examination. Neuroimaging by MRI was performed only at baseline.

Clinical information included MMSE [16], CDR [17,18], Geriatric Depression Scale-15 (GDS-15) [19], Neuropsychiatric Inventory (NPI) [20], and the Disability Assessment Dementia (DAD) [21]. The neuropsychological examination consisted of a battery of cognitive tests performed by a (neuro)psychologist, and was harmonized across centres. Tests included Auditory Verbal Learning Test [22], Visual Association Test [23], and Digit-Span [24] to assess memory; Letter Digit Substitution Test [25] to assess mental processing speed; and Stroop Color-Word Test [26] and Trail Making Test [27,28] to assess attention and concentration. Raw scores were converted to z-scores, adjusting for age, education and gender. Neuroimaging markers were subtracted from 3 T MRI scan images and included medial temporal lobe atrophy measurements (scored as 0=normal to 4=end stage atrophy) [29], white matter lesions scored by the Fazekas scale [30], number of infarcts, lacunes and micro bleeds.

Cerebrospinal fluid (csf)

At each centre CSF was collected by lumbar puncture between the L3/L4 or L4/L5 intervertebral space, centrifuged, and stored at -80°C in polypropylene tubes. CSF Beta Amyloid (1–42) and total Tau were measured by commercially available sandwich ELISAs (Innotest Beta-Amyloid (1–42); Innotest hTAU-Ag; Innogenetics, Ghent, Belgium) at the laboratory in Amsterdam.

CSF was dichotomized as abnormal (CSF+) when the ratio (Beta Amyloid / total Tau) was lower than 1 or normal (CSF−) when the ratio was higher than or equal to 1. The ratio was predefined by BetaAmyloid/(240 + (1.18 * TotalTau)) [31].

Diagnostic pathways

A diagnosis that represented the current clinical practice diagnostic pathway was reached by multiple trios of clinical experts from 3 different disciplines (geriatric medicine, psychiatry, neurology) with at least 5 years of clinical experience. All relevant outcomes from the baseline measurement (i.e. physical, clinical and neuropsychological examination, a summary of the patient history, and the MRI measurement scores) were summarized in a vignette for each patient individually. The trios of clinical experts were asked to decide on the syndrome, the most likely aetiology and the expected course of cognitive or daily functioning within 2 years in a consensus procedure as explained in chapter 3. Expectations whether or not decline was likely were classified for each patient as A) decline expected due to AD, B) decline expected due to another cause, or C) no decline expected. This classification constituted the 'clinic without CSF' pathway (see figure 1a).

This pathway was compared to a pathway in which a CSF test was positioned after the clinical judgment (as reflected in the 'clinic without CSF' pathway), by means of a decision rule, as an add-on test to evaluate the CSF's added value. The CSF ratio was dichotomized as AD-profile (CSF+) or no AD profile (CSF−). The decision rule reclassified subjects with an AD CSF-profile in group C (no

decline expected) to group A (decline expected due to AD). Subjects with a non-AD CSF profile in group A (decline expected due to AD) were reclassified to group C (no decline expected). Subjects in group B (decline expected due to another cause or AD mixed with another cause) were never reclassified if CSF presented a normal profile. The CSF-driven pathway constituted the 'clinic including CSF' pathway (see figure 1b).

Diagnostic reference standard

Both the results based on the 'clinic without CSF' pathway as well as the results based on the 'clinic including CSF' pathway were verified against the same diagnostic reference standard. This diagnostic reference standard, the outcome measure of this study, was decline on follow-up, determined by a consensus panel that followed an adjusted Delphi method. For that, the vignette with baseline information was updated with the same information on 1- and 2-year follow up (except for MRI scale scores at follow up because they were not available). The same trios of clinical experts were asked to individually judge observed decline by answering the question "In your opinion, what was the course of cognitive or daily functioning during the 2-year follow-up?". If discrepant judgments occurred the trio of clinical experts was invited to express their arguments and reach consensus during a panel meeting group discussion led by a panel facilitator. Panel experts were blinded for any CSF result in this study to prevent incorporation bias. A detailed description of the expert panel protocol can be found in chapter 3.

Statistical analyses

Sensitivity and specificity, including Wilson 95% confidence interval, was calculated by comparing both the 'clinical without CSF' and 'clinic including CSF' diagnoses to the reference diagnosis. Also, positive predictive value and negative predictive value were calculated for both pathways. Furthermore, the change in diagnoses due to CSF was calculated as the percentage of subjects that were reclassified by the 'clinic including CSF' pathway compared to the 'clinic without CSF' pathway (i.e. subjects classified as expected decline in the 'clinic without CSF' pathway and as no decline expected in the 'clinic including CSF' pathway or vice versa). Comparing the changed diagnoses with the reference standard enabled us to split this percentage into correctly reclassified or incorrectly reclassified. Outcomes were presented for each subgroup of subjects with SMC, MCI and dementia. The Wilson 95% confidence intervals and statistical difference in proportions ($p < 0.05$) were analysed using STATA12.

RESULTS

Subjects who, before the first follow-up measurement, died (n=3), were institutionalized (n=4) or were unwilling to further participate (n=33) were excluded for analysis. Reasons for not performing a lumbar puncture (n=100) were unwillingness to receive it or a physician considered CSF too burdensome. Subjects with and without CSF differed significantly regarding female gender (27 vs. 41%; $p < 0.05$) and MMSE (mean 26.3 vs. 25.4; $p < 0.05$). Part of the subjects had only 1 of the 2 follow-up assessments available (21%).

Table 1: Baseline sample characteristics

	Subsample SMC		Subsample MCI		Subsample dementia	
Number of cases	30	(27%)	41	(37%)	40	(36%)
Females, no. (%)	12	(40)	8	(20)	10	(25)
Age, years (SD)	63.4	(7.9)	67.8	(8.8)	69.3	(8.4)
Years of formal education, mean (SD)	12.2	(3.5)	11.5	(3.3)	11.1	(3.4)
MMSE, mean (SD)	28.3	(1.4)	27.2	(1.8)	23.9	(2.9)
NPI, mean (SD)	12.3	(8.8)	14.4	(17.5)	19.4	(15.3)
DAD, mean percentage (SD)	93	(9.8)	89.5	(14.1)	76.7	(18.5)
GDS, mean percentage (SD)	3.4	(2.4)	3.3	(2.5)	3.3	(3)
CDR, mean percentage (SD)	0.33	(0.24)	0.5	(0.23)	0.84	(0.31)
CSF Beta Amyloid (1-42) (SD)	787	(268)	684	(251)	687	(296)
CSF total Tau (SD)	332	(229)	472	(286)	550	(321)
CSF ratio ABeta/(240+(1.18*total-tau)) (SD)	1.45	(0.65)	1.02	(0.57)	0.93	(0.6)
Observed cognitive or functional decline after 2 years, no. (%)	1	(3%)	21	(51%)	36	(90%)
Conversion to dementia (in MCI) and conversion to MCI or dementia (in SMC) after 2 years, no. (%)	4	(13%)	14	(34%)	Not applicable	

Abbreviations: SD, standard deviation; MMSE, Mini-mental state examination; NPI, neuropsychiatric inventory; DAD, disability assessment for dementia; GDS, geriatric depression scale; CDR, clinical dementia rating; CSF cerebrospinal fluid; MCI, mild cognitive impairment; SMC, subjective memory complaints

Hundred eleven patients, aged 39-89 years, could be included in the analyses, of whom 30 had SMC, 41 MCI and 40 dementia. See table 1 for an overview of the baseline sample characteristics by syndrome. According to the diagnostic reference standard a clinically relevant cognitive or functional decline was observed in 58 subjects (52%) in the total sample. The patient flow of the 'clinic without CSF' pathway is graphically presented in figure 1a. In 20 MCI and 35 demented subjects decline was correctly predicted only relying on clinical judgment. In 25 SMC and 9 MCI subjects 'no decline' was correctly predicted. It was falsely predicted in 5 SMC, 12 MCI and 5 demented subjects. The experts could not reach consensus in 3 cases and therefore a majority diagnosis was adopted.

When applying CSF as an add-on test to clinical judgment the patient flow changed due to several reclassifications (figure 1b, representing the clinic including CSF pathway). Of all subjects who were expected to decline, 1 SMC, 3 MCI and 2 demented subjects were correctly reclassified due to CSF (i.e. subjects who were expected to decline by the 'clinic without CSF' and reclassified due to a non-AD profile (CSF–) in whom no decline was observed after 2 years) but 3 MCI and 5 demented subjects were incorrectly reclassified. Ten SMC and 1 MCI subjects were incorrectly reclassified to expected decline (CSF+ in after group C in de subjects who eventually declined) but no subjects were correctly reclassified to expected decline (CSF+ in after group C in de subjects who eventually did not decline).

Figure 1a: Flowchart representing the clinical practice diagnostic pathway to determine cognitive or functional decline
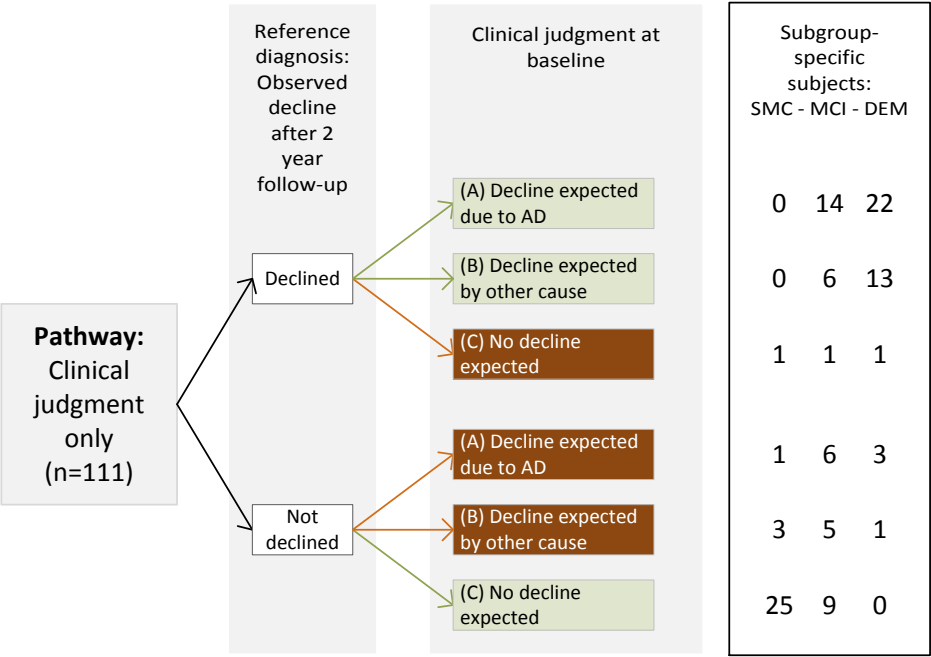
Figure 1b: Flowchart representing a diagnostic pathway in which clinician's expected decline due to AD was changed according to the CSF ratio
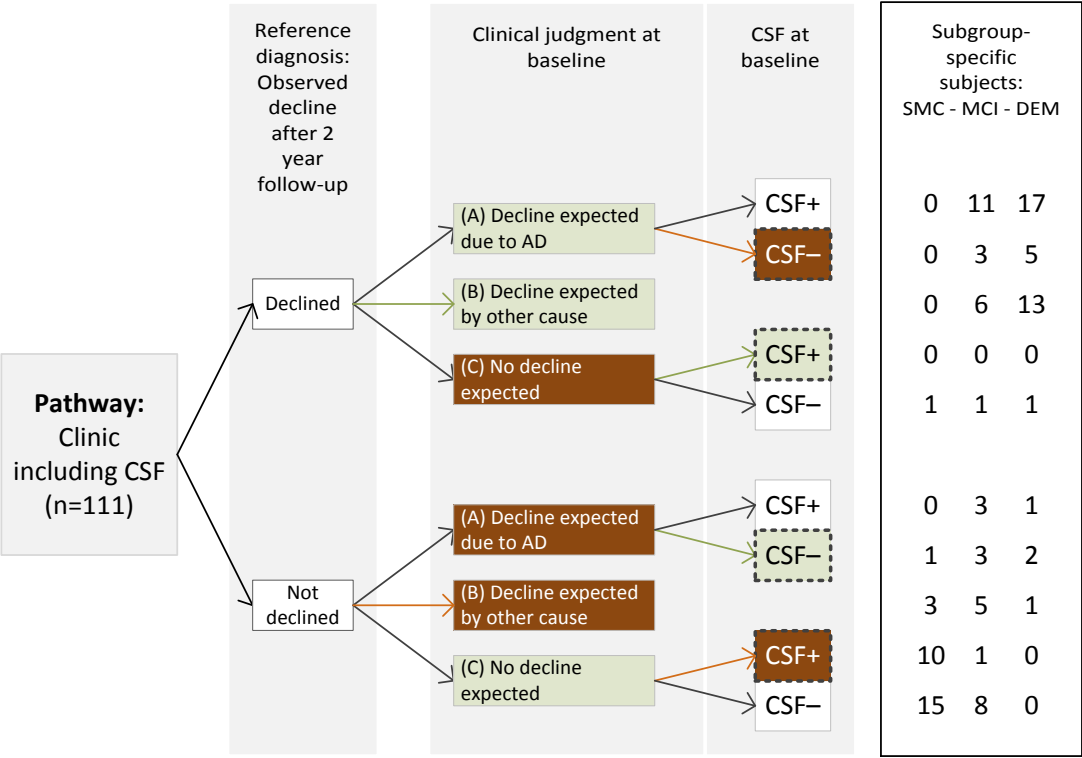
SMC, subjective memory complaints; MCI, mild cognitive impairment; DEM, dementia; green (light shaded) = correctly classified; red (dark shaded) = incorrectly classified; dashed line = reclassified due to CSF

Specificity increased in the MCI (+10%) and dementia (+50%) subgroups at the cost of sensitivity (-14% and -14% respectively) as a result of adding CSF. In SMC the sensitivity did not change and specificity decreased (-31%). The positive and negative predicted values worsened or remained stable in SMC and MCI and improved in dementia. In all subgroups the proportion of subjects in whom a reclassification was correct was lower than 50% (Table 2 and visually presented in Figure 2).

Table 2: Diagnostic accuracy of the 'clinic without CSF' pathway and the 'Clinic including CSF' pathway. Reclassification measures represent the percentage of subjects that were correctly or incorrectly reclassified by the CSF pathway compared to the clinical only pathway.

Subgroup	Pathway	Decline observed (n) exp+[1]	Decline observed (n) exp−[2]	No decline observed (n) exp+[1]	No decline observed (n) exp−[2]	Sensitivity (95%CI)	Specificity (95%CI)	Positive predicted value	Negative predicted value	Reclassified subjects (%)	Of which correctly reclassified (%)
SMC (n=30)	Clinical judgment only	0	1	4	25	0 (0-79)	86 (69-95)	0	0.96	na	na
SMC (n=30)	Clinic including CSF	0	1	13	16	0 (0-79)	55 (38-72)	0	0.94	37%	9%
MCI (n=41)	Clinical judgment only	20	1	11	9	95 (77-99)	45 (26-66)	0.65	0.90	na	na
MCI (n=41)	Clinic including CSF	17	4	9	11	81 (60-92)	55 (34-74)	0.65	0.73	17%	43%
Dementia (n=40)	Clinical judgment only	35	1	4	0	97 (86-100)	0 (0-49)	0.90	0	na	na
Dementia (n=40)	Clinic including CSF	30	6	2	2	83 (68-92)	50 (15-85)	0.94	0.25	18%	29%

1 decline was expected; 2 decline was not expected
Abbreviations: MCI, mild cognitive impairment; SMC, subjective memory complaints; CSF, cerebrospinal fluid; na, not applicable

DISCUSSION

This study determined the prognostic added value of CSF to clinical judgment (consisting of patient history, physical and neurological examination, neuropsychological tests, and MRI scale scores) in predicting cognitive or functional decline within 2 years in memory clinic subjects. CSF was positioned as an add-on test to confirm an AD-related expected decline or to reject a non-expected decline by means of a decision rule. Although adding CSF resulted in several reclassifications, these were mostly incorrect. The data of this study do not show an added value of CSF to clinical judgment in predicting 2-year cognitive or functional decline.

Figure 2: Percentages (y-axis) classified correctly (light shaded above 0 (green)) and incorrectly (light shaded below 0 (red)) by clinic without CSF, and percentages correctly (dark shaded above 0 (green)) and incorrectly (dark shaded below 0 (red)) reclassified due adding CSF to clinical judgement in the total sample and the subsamples.
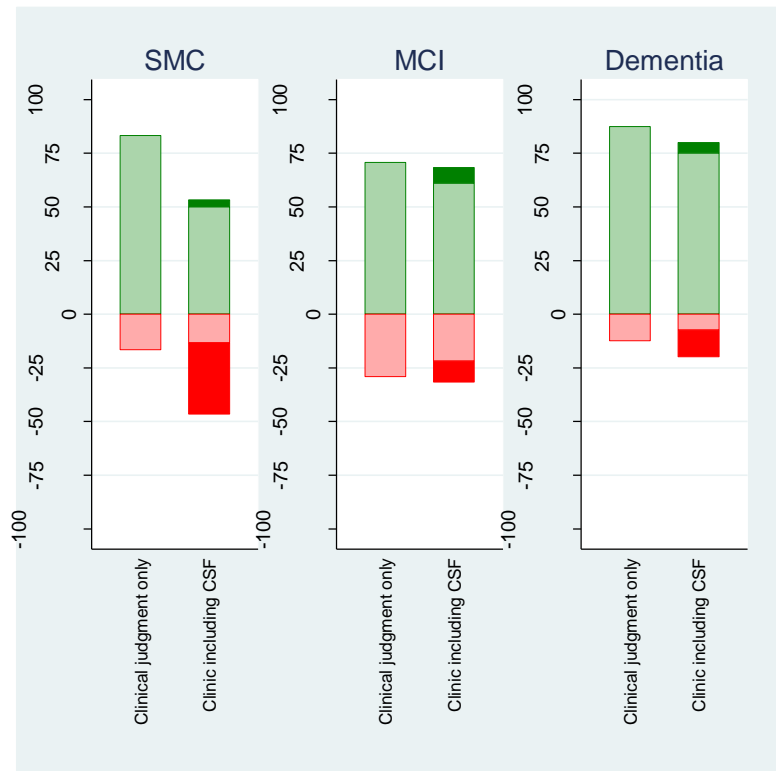
SMC, subjective memory complaints; MCI, mild cognitive impairment

Eleven subjects who remained stable within a 2 year follow-up period had an abnormal CSF value of whom 1 suffered from MCI. The other 10 suffered from SMC in whom it is likely to take more time before AD symptoms come to expression because they seem to be at a more early stage of the disease than MCI subjects [32]. This is reflected in the low prior probability of decline in this subgroup mirrored by the ceiling level of the negative predictive value. In 8 subjects who declined within 2 years clinical judgment expected an AD-related decline while these subjects had a normal CSF value. Possibly the experts incorrectly indicated AD as the cause of the decline and therewith the decision rule wrongfully appealed to CSF to confirm this. If the actual cause was in fact another neurodegenerative disease, a normal CSF would not have rejected the expected decline since other neurodegenerative diseases are not reflected in the AD CSF profile of Beta Amyloid and total Tau ratio. Possible incorrect clinical judgment of aetiology was reflected by 3 cases in which a majority diagnosis had to be adopted because the experts could not reach consensus.

Other studies evaluated the added value of CSF to the MMSE [9,10] and to neuropsychological tests [9,12–14] by logistic regression and found a significantly increased prediction of MCI conversion to AD-type dementia. Post-hoc we performed comparative analyses in our MCI subsample and found similar results of significant increased prediction of decline when adding CSF Beta Amyloid(1-42) total Tau ratio in a logistic regression model with only MMSE (p=0.022) or only standardized 15-WLT delayed recall (p=0.044) (all variables continuously measured). When

correcting for age and gender the results were not significant which could be explained by the small sample size.

Two other clinically relevant outcome measures were reported by these studies. Vos et al. [9] reported a 0.20 probability increase of conversion to dementia when CSF was abnormal and 0.20 probability decrease when CSF was normal in MCI. In a post-hoc analysis we found a respectively 0.17 and 0.19 probability change. The difference could be explained by the data-driven cut-off point applied by Vos et al. [9] which possibly overestimated the biomarker performance [33]. Richard et al. [12] found a negative net reclassification index (-2.2) of CSF when added to a cognitive test (the immediate recall of Rey's Auditory Verbal Learning Test) indicating that CSF decreased diagnostic accuracy as a result of wrong reclassifications of cognitive decline. In a post-hoc analysis we revealed a net reclassification index of -0.10 which has the same direction of effect. The use of the reclassification index, which was negative, already hints that the benefits shown in a logistic regression model do not necessarily reflect improved reclassification (e.g. abnormal CSF in an already predicted conversion by a single cognitive test could improve the model for statistical significance, without improving correct reclassification).

This study relied on the clinical interpretation of all information available in standard clinical practice (history, physical, neurological, and neuropsychological examination, and MRI) which explains why no incremental value of CSF was found. The results of this study, however, are more generalizable to clinical practice than currently available evidence [9,10,12–14].

The study was subject to several limitations. Sample size was limited in the subgroups. The experts' assessments of observed decline (reference diagnosis) were not blinded for the concluded clinical judgement of expected decline at baseline (clinic without CSF pathway) since it was scored by the same experts. This could have resulted in diagnostic review bias that brings the reference diagnosis close to the 'clinic without CSF' leaving less room for added value due to CSF. However, this is also the case in clinical practice.

Furthermore, the method to establish the 'clinic without CSF' diagnoses was based upon expert interpretation of clinical information. This differed from the method to establish the 'clinic including CSF' diagnoses which was based upon a predefined decision rule combined with the aggregated Beta Amyloid and total Tau information into a single ratio. In practice clinicians interpret CSF information in a more refined way by taking into account the different disease concepts that are reflected by Beta Amyloid and total Tau and relate it to each patient's individual context. Ideally, the CSF information is therefore added to the baseline standard clinical information in a second step and interpreted by experts similarly to the 'clinic without CSF' procedure as explained in chapter 3.

Standardized CSF cut-offs are currently lacking and therefore the cut-offs used to define CSF markers as abnormal may not have been optimal to identify all subjects who will show clinical progression at follow-up. Other cut-offs [9,34] were applied in a sensitivity analysis though did not generate an increased proportion correctly reclassified subjects within any of the subgroups.

The reclassification measure assumed similar impact of a false positive and false negative. This, however, may not be the case and therefore preventing a false positive may be worth incorrect labelling several subjects as false negative or vice versa. Possibly the benefits of correctly reclassifying the 3 MCI cases, in whom clinicians expected AD-related decline but remained stable, offset the disutility of the 4 incorrectly reclassified MCI cases. In addition, cognitive or functional decline is subject to individual experience and not necessarily reflected by a clinician's judgment

of it. An increase of prognostic accuracy does not automatically improve patient outcome. Studies on the clinical utility of an early diagnosis are therefore urgently needed to establish the impact of these biomarkers on patient health beyond treatment [35], especially in the highly debated issue of diagnosing pre-dementia [8].

The use of 'course of cognition or daily functioning within 2 years' as the study outcome reflects a change in demand of care, which is assumed relevant to a patient and his/her informal caregiver, opposite to pathology. The results are therefore not intended for research practice nor can they be used for implications for pathological research or for other clinical purposes than prognosis. The study's major strength was the use of the expertise of 3 clinicians with different backgrounds combined with the clinically relevant measure of decline over e.g. the use of a single value such as a cognitive test. This is generalizable to clinical practice.

At last, CSF was performed when an AD-related decline was expected or when no decline was expected. As an alternative, CSF could be performed only in one of these cases. This could result in a change of sensitivity and specificity and possibly also impact the costs per diagnosis. Studying the costs and consequences of various CSF diagnostic test strategies fell outside the scope of this article. It is a relevant topic for further research.

Conclusion

The results of this study showed that adding the CSF Beta Amyloid total tau ratio information to the clinical judgment by a decision rule did not improve the prediction of 2-year cognitive or functional decline based on clinical evaluation in memory clinic subjects with subjective memory complaints, mild cognitive impairment or dementia. Clinical utility studies should reveal the impact of medical decision and consequences beyond treatment decision on patient outcomes when applying CSF in the diagnostic workup, before adoption in clinical practice can be decided.

REFERENCES

[1]   Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013;9:63–75.e2.

[2]   Wimo A, Jönsson L, Bond J, Prince M, Winblad B. The worldwide economic impact of dementia 2010. Alzheimers Dement 2013;9:1–11.e3.

[3]   Prince M, Bryce R, Ferri CP. World Alzheimer Report 2011. London: 2011.

[4]   Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia-- meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand 2009;119:252–65.

[5] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:335–7.

[6] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment 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 2011;7:270–9.

[7] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280–92.

[8] Le Couteur DG, Doust J, Creasey H, Brayne C. Political drive to screen for pre-dementia: not evidence based and ignores the harms of diagnosis. BMJ 2013;347:f5125.

[9] Vos S, Van Rossum I, Burns L, Knol D, Scheltens P, Soininen H, et al. Test sequence of CSF and MRI biomarkers for prediction of AD in subjects with MCI. Neurobiol Aging 2012.

[10] Buerger K, Teipel SJ, Zinkowski R, Blennow K, Arai H, Engel R, et al. CSF tau protein phosphorylated at threonine 231 correlates with cognitive decline in MCI subjects. Neurology 2002;59:627–9.

[11] Palmqvist S, Hertze J, Minthon L, Wattmo C, Zetterberg H, Blennow K, et al. Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up study. PLoS One 2012;7:e38639.

[12] Richard E, Schmand BA, Eikelenboom P, Van Gool WA. MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer's disease in patients with mild cognitive impairment: a diagnostic accuracy study. BMJ Open 2013;3.

[13] Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR, et al. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol Aging 2012;33:1203–14.

[14] Heister D, Brewer JB, Magda S, Blennow K, McEvoy LK. Predicting MCI outcome with clinically available MRI and CSF biomarkers. Neurology 2011;77:1619–28.

[15] Handels RL, Aalten P, Wolfs CA, Olde Rikkert M, Scheltens P, Visser PJ, et al. Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease: the research protocol of a prospective cohort study. BMC Neurol 2012;12:72.

[16] Cushman L, Scherer MJ. Psychological assessment in medical rehabilitation. 1st ed. Washington, DC: American Psychological Association; 1995.

[17] Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993;43:2412–4.

[18] Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982;140:566–72.

[19] Almeida OP, Almeida SA. Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD-10 and DSM-IV. Int J Geriatr Psychiatry 1999;14:858–65.

[20] Cummings JL. The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. Neurology 1997;48:S10–S16.

[21] Gélinas I, Gauthier L, McIntyre M, Gauthier S. Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. Am J Occup Ther 1999;53:471–81.

[22] Brand N, Jolles J. Learning and retrieval rate of words presented auditorily and visually. J Gen Psychol 1985;112:201–10.

[23] Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C. Visual association test to detect early dementia of the Alzheimer type. J Neurol Neurosurg Psychiatry 2002;73:126–33.

[24] Wechsler D. WMS-III: Wechsler Memory Scale Administration and Scoring Manual. Psychological Corporation; 1997.

[25] Natu M V, Agarwal AK. Digit Letter Substitution Test (DLST) as an alternative to Digit Symbol Substitution Test (DSST). Hum Psychopharmacol Clin Exp 1995;10:339–43.

[26] Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935;18:643–62.

[27] Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Ski 1958;8:271–6.

[28] Reitan R. Trail Making Test: Manual for Administration, Scoring and Interpretation. Bloomington: Section of Neuropsychology, Department of Neurology, Indiana University Medical Centre; 1956.

[29] Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J Neurol Neurosurg Psychiatry 1992;55:967–72.

[30] Fazekas F, Kleinert R, Offenbacher H, Schmidt R, Kleinert G, Payer F, et al. Pathologic correlates of incidental MRI white matter signal hyperintensities. Neurology 1993;43:1683–9.

[31] Hulstaert F, Blennow K, Ivanoiu A, Schoonderwaldt HC, Riemenschneider M, De Deyn PP, et al. Improved discrimination of AD patients using beta-amyloid(1-42) and tau levels in CSF. Neurology 1999;52:1555–62.

[32] Visser PJ, Verhey F, Knol DL, Scheltens P, Wahlund L-O, Freund-Levi Y, et al. Prevalence and prognostic value of CSF markers of Alzheimer's disease pathology in patients with subjective cognitive impairment or mild cognitive impairment in the DESCRIPA study: a prospective cohort study. Lancet Neurol 2009;8:619–27.

[33] Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem 2008;54:729–37.

[34] Duits F, Teunissen C, Bouwman F, Visser PJ, Mattsson N, Zetterberg H, et al. The CSF Alzheimer's profile: Easily said, but what do we mean? Alzheimers Dement 2013;9:147.

[35] Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PMM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ 2012;344:e686.

PART II

COST-UTILITY OF CSF

BIOMARKERS

CHAPTER 5

Diagnosing Alzheimer's disease: A systematic review of economic evaluations

Handels RL, Wolfs CA, Aalten P, Joore MA, Verhey FR, Severens JL.

Abstract

Background: The objective of this study is to systematically review the literature on economic evaluations of interventions for the early diagnosis of Alzheimer's disease (AD) and related disorders and to describe their general and methodological characteristics. We focused on the diagnostic aspects of the decision models to assess the applicability of existing decision models for the evaluation of the recently revised diagnostic research criteria for AD.

Methods: PubMed and the National Institute for Health Research Economic Evaluation database were searched for English-language publications related to economic evaluations on diagnostic technologies. Trial-based economic evaluations were assessed using the Consensus on Health Economic Criteria list. Modeling studies were assessed using the framework for quality assessment of decision-analytic models.

Results: The search retrieved 2109 items, from which eight decision-analytic modeling studies and one trial-based economic evaluation met all eligibility criteria.

Conclusions: Diversity among the study objective and characteristics was considerable and, despite considerable methodological quality, several flaws were indicated. Recommendations were focused on diagnostic aspects and the applicability of existing models for the evaluation of recently revised diagnostic research criteria for AD.

INTRODUCTION

Alzheimer's disease (AD) and other dementing disorders are common in the elderly, with a worldwide prevalence estimated in 2010 at 35.6 million, which will increase to 115.4 million in 2050. AD has a substantial impact on the person who suffers from the disease, his or her family, and society [1] and [2]. The total worldwide cost of AD and other dementing disorders was estimated at $604 billion in 2010 [3].

Earlier diagnosis and early intervention are considered important mechanisms to manage the worldwide impact of the disease. Early diagnosis can be described as a "timely" recognition of mild dementia in response to a patient's complaints to ensure that disabled individuals receive the necessary support and care or as the "symptomatic predementia" diagnosis when cognition is impaired but functioning not yet affected (typically referred to as mild cognitive impairment) [4].

Until recently, the diagnosis of AD was largely based on clinical judgment using the NINCDS-ADRDA criteria [5]. These criteria were recently revised [6], [7] and [8] to enhance diagnostic accuracy and enable an early diagnosis even when only very mild clinical symptoms are present. Biomarkers in cerebrospinal fluid (CSF), positron emission tomography (PET), magnetic resonance imaging (MRI), and photon emission tomography (SPECT) are attributed a more prominent role in the new diagnostic research criteria. However, validation of these research criteria is needed before the role of new biomarkers can be adopted in clinical practice [9].

The ultimate goal of diagnostic testing is to guide disease management to improve patient outcomes and patient well-being. Tests that lack this potential are considered obsolete [10] and [11]. Furthermore, because health-care resources are scarce and must be allocated efficiently, decision-makers require evidence of the cost-effectiveness of diagnostic tests before adoption in clinical practice. Such evidence can be generated by decision-analytic models which are defined as a set of mathematical relationships that form a structure reflecting the natural progression of a disease. By simulating patient cohorts, these models enable the estimation of the likelihood of each consequence and its corresponding costs and effects [12] and [13]. Trial-based economic evaluations, in which costs and health-care outcomes are measured during clinical trials, can also provide evidence of cost-effectiveness.

Decision-analytic models of AD have been reviewed extensively by Cohen et al [14]. However, this review only included models that project disease progression, excluding possible relevant evidence on the evaluation of diagnostic techniques. Furthermore, the applicability of existing decision models to evaluate the recently revised research criteria has not been elaborated. This raises the urgent need for a review of economic evaluations of diagnostic interventions for AD.

The objective of this study was to systematically review the literature on economic evaluations of interventions for the early diagnosis of AD and related disorders and to describe their general and methodological characteristics. Using these results, recommendations for future studies were focused on the diagnostic aspects of the decision models to assess the applicability of existing decision models for the evaluation of the recently revised diagnostic research criteria for AD.

METHODS

Search strategy

A systematic literature review was performed to identify economic evaluations of diagnostic interventions for AD or related dementias. The following eligibility criteria were applied:

1. The study should focus on a population, either an empirical (primary data) or a theoretical (model), which is suffering from or suspected of suffering from AD or related disorders (vascular dementia, dementia with Lewy bodies, and frontotemporal dementia). The population should consist of previously undiagnosed individuals. Studies of neurodegenerative disorders were excluded (e.g., Parkinson's, Huntington's disease, or depression).
2. The population reflects humans, 55 years of age or older.
3. The intervention is a diagnostic technology, tool, questionnaire, process, procedure, or protocol used for a timely or symptomatic predementia diagnosis of AD or related dementias in a clinical setting. Screening tools and risk, severity, or progression analyses were excluded (screening tools such as the Mini-Mental State Examination [MMSE] or DNA risk assessment).
4. The study reports primary patient and/or cost data or uses a mathematical model that is based on such data. Reviews, case studies, and publications that merely describe methodological issues were excluded.
5. The study is an economic evaluation: either a cost-consequence [15] analysis or a full economic evaluation (cost-effectiveness, cost-utility, or cost-benefit analyses). Partial economic evaluations, which do not include a comparison between different strategies or do not analyze both costs and consequences [16], were excluded.

PubMed and the National Institute for Health Research Economic Evaluation Database databases [17] and [18] were searched through March 2011. Medical subject headings and free text words on dementia, diagnosis, and economic evaluation were used to identify relevant English-language articles with an available abstract (see Appendix 1 for the full search query).

Two reviewers (R.H. and C.W.) independently assessed titles. A title was excluded if both reviewers agreed that it explicitly met one of the exclusion criteria. The same reviewers independently assessed abstracts of the remaining titles. An abstract was excluded if either reviewer considered that it did not meet all five inclusion and exclusion criteria. Dissimilarities in the reviewers' assessments were resolved by discussion. The full article was assessed if the remaining abstracts had dissimilarities that could not be otherwise resolved. If an article was not accessible, the author was contacted to request a copy of the original publication. A third reviewer (J.S.) resolved the remaining differences in the reviewer's assessments; this third reviewer made the final decision as to whether the article would be included. See Appendix 2 for an overview of the study selection process.

Analyses

General study characteristics of all of the included articles are described in Table 1. For modeling studies, the model type was scored separately for the diagnostic and treatment part of the model. A Markov model is characterized by mutually exclusive disease states that represent the possible consequences of the options under evaluation. Disease progression is reflected by the transition of a patient's disease states over discrete time periods [12]. We used the term "state-independent" for mathematical models that apply a fixed time cycle similar to a Markov model but reflect the disease progression on a continuous scale and not by mutually exclusive disease states, such as the MMSE categories mild AD (MMSE 21–26), moderate AD (MMSE 14–20), moderately severe AD (MMSE 10–14), and severe AD (MMSE <10). If the treatment was only summarized in one or more fixed outcome values and was not represented by a mathematical model, it was scored as "static". Buxton et al [19] described five different roles of modeling in an economic evaluation that we used to classify all of the modeling studies: extrapolating beyond data observed in trial, linking the intermediate (clinical) endpoints to final outcomes, generalizing to other settings, synthesizing head-to-head comparisons where relevant trials do not exist, and informing decisions in the absence of hard data.

Methodological characteristics of the trial-based economic evaluations were assessed using the Consensus on Health Economic Criteria (CHEC) list [20] (see Table 2). Methodological characteristics of the modeling studies were assessed using the framework for quality assessment of decision-analytic models [21] and [22] (see Table 3), which includes three key themes: structure, data, and consistency. Questions S3.2, S3.3, and S8 required expert knowledge in the field of AD and were clarified by a psychiatrist (F.V.). Furthermore, the framework is most likely intended for treatment interventions, although diagnostic interventions are the focus of this review and required adjustment of the items D2b1 and D2b2 that are described in Table 3. Two reviewers (C.W. and R.H.) independently performed analyses. In one case, one of the reviewers was the author of the paper being assessed; therefore, this reviewer was replaced by another reviewer (P.A.) to ensure objectivity. Dissimilarities in scoring were resolved by discussion. If discussion was not successful, a third reviewer (J.S.) determined the final score. Both of the checklist questions were answered with "yes" if both reviewers agreed that the study paid sufficient attention to an item, "no" if the item was not fulfilled or if insufficient information was available to assess the item, or "N/A" (not applicable) if the question was either not applicable to the study or if it referred to a previous question scored as "no". A summary score for each article was not applied because this quantitative scoring system was not considered to be sufficiently reliable and valid as a means of quality assessment. Therefore, a qualitative summary is provided to better distinguish major from minor methodologic flaws.

RESULTS

The search retrieved 2109 titles, of which 90 did not have an abstract available, 84 were duplicates, and an additional 180 were not in English. From the remaining 1755 titles, 936 were excluded based on the exclusion criteria. Of the remaining 819 titles, 735 were excluded after reading the abstracts, leaving 84 articles for assessment by R.H. and C.W. After reading the full articles, they excluded 75 articles on the basis of the exclusion criteria, of which 2 were assessed by the third reviewer (J.S.) because a decision could not be made through discussion between

R.H. and C.W. Thus, a total of nine studies [23], [24], [25], [26], [27], [28], [29], [30] and [31] were included in the review. Fig. 1 displays a flowchart of the inclusion and exclusion process. Three articles [24], [25] and [26] analyzed a combination of two models of which the treatment part was published in a different article [32] and [33]. To fully evaluate the model, this publication was also included for assessment. Furthermore, the economic evaluation in two articles [25] and [26] was based on the same model and a similar situation applies to references 27 and 28. If the articles referred to one another the information from the referenced article was also used to assess the article for this review.

Figure 1: Flowchart of the inclusion process of the search results.
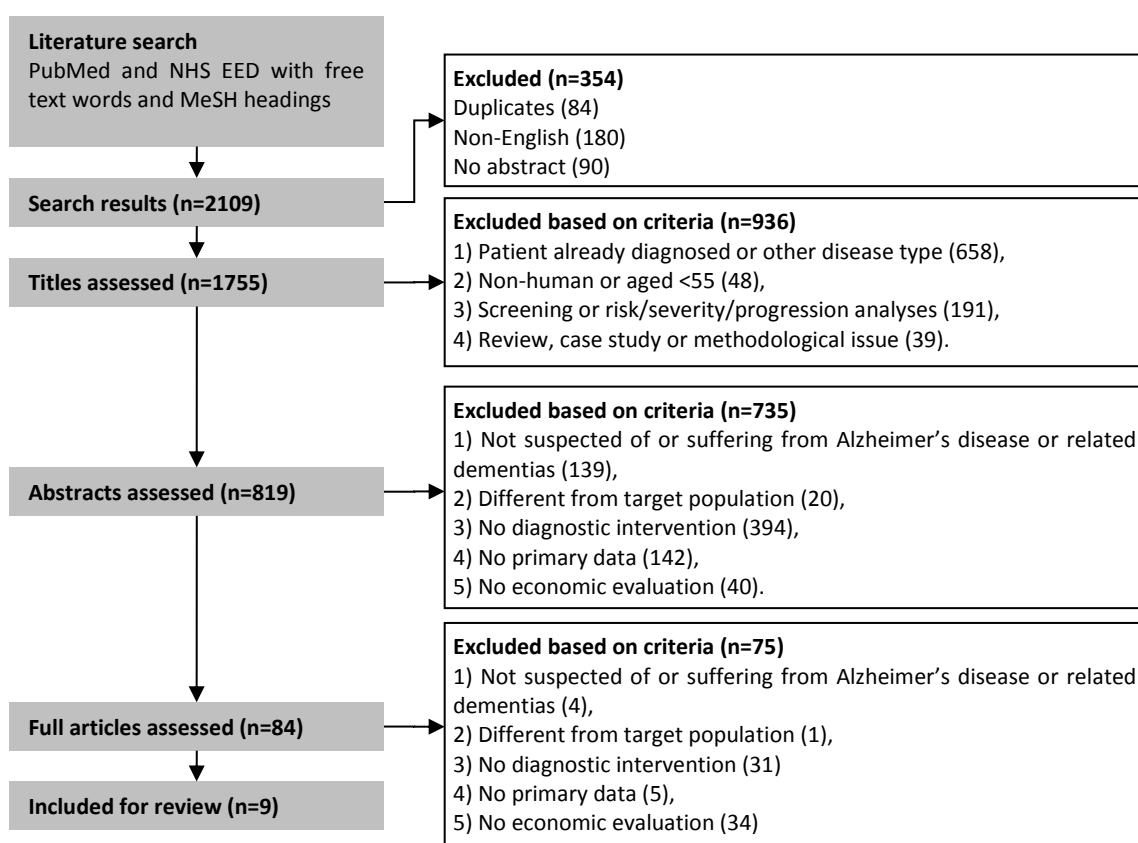
Table 1 provides an overview of the study characteristics. The studies can be divided into two categories on the basis of their objective. The first consists of three studies evaluating early diagnoses, which are strategies aimed at preventing underdiagnosis (i.e., timely recognition of patients suffering from dementia) to bring forward treatment. Although gained treatment effects in patients who would otherwise be undiagnosed were evaluated, the impact of an incorrect diagnosis (i.e., a false positive or false negative test outcome) was not included [23], [24] and [30]. The second category consists of six studies evaluating a specific diagnostic test or imaging strategy. Except for the trial-based study, a disease progression model was applied to describe costs and effects of true- and false-positive and -negative diagnostic pathways [25], [26], [27], [28], [29] and [31].

Diversity among the study characteristics was considerable on several items. Strategies under evaluation varied from specific imaging techniques to multidisciplinary assessments. Population ranged from non-demented patients with early symptoms to moderate dementia evaluated in various care settings. The time horizon ranged from 1 year to lifetime. At last, study outcomes and conclusions on the adoption of the intervention under evaluation in clinical practice differed between the studies.

Table 1: General study characteristics

Author	Banerjee et al. [23]	Getsios et al.[24]*	McMahon et al.[25]†	McMahon et al.[26]†
Publication year	2009	2012	2000	2003
Country	UK	UK	US	US
Objective	Analyze the costs and benefits of commissioning memory services for early diagnosis and intervention for dementia	Evaluate the cost-effectiveness of early assessment and treatment with AD with Donepezil	Evaluate the cost-effectiveness of functional neuroimaging	Evaluate the cost-effectiveness of PET in the diagnosis of AD
Study type	Modeling	Modeling	Modeling	Modeling
Diagnostic model‡	Static	discrete event simulation	Decision-tree	Decision-tree (based upon model McMahon et.al.)
Treatment model‡	Mathematical model	discrete event simulation	Markov model	Markov model
Role of modeling§	Informing decisions in the absence of hard data	Informing decisions in the absence of hard data	Synthesizing head-to-head comparisons where relevant trials do not exist	Synthesizing head-to-head comparisons where relevant trials do not exist
Base case situation	Current practice	Early assessment for AD and treatment with Donepezil	Standard examination	Standard examination
Comparative strategies¶	A multi-disciplinary and interagency team to generate early diagnosis	(1) Treatment without early assessment; (2) No early assessment and no treatment	(1) MRI plus DSC MRI; (2) Visual SPECT; (3) Computed SPECT (all added to standard examination)	(1) DSC MRI; (2) FDG PET; (3) Computed SPECT (all added to standard examination)
Impact of diagnostic test included in the model	Bring forward treatment (incorrect diagnoses not modeled)	Bring forward treatment (incorrect diagnoses not modeled)	A disease progression model is applied on true and false positive and negative diagnoses	A disease progression model is applied on true and false positive and negative diagnoses
Diagnostic care setting	Existing primary and secondary care services in England	GP and specialist centers in UK	A specialized Alzheimer disease center (tertiary Alzheimer disease clinic)	A specialized Alzheimer disease center
Study population	Patients in an early state of dementia	Patients with undiagnosed AD	Patients with mild or moderate dementia (at presentation to an Alzheimer disease center)	patients with mild or moderate dementia who present to specialized AD centers
Primary patient level data	N/A	N/A	N/A	N/A
Perspective	Societal	Health care payer and societal	Societal	Societal
Time horizon	10 years	10 years	18 months	18 months
Outcome measure	Net benefit	ICER	ICER	ICER
Main results	NPV over 10 years: -£950 million. A gain of between 0.01 and 0.02 QALYs per person year would render the service cost-effective	Early assessment and treatment dominates both other options under evaluation	ICER DSC MRI: $479,500/QALY compared with the usual diagnostic work-up. Visual or quantitative SPECT was dominated	FDG PET and sPECT were dominated. ICER DSC MRI: $598,800/QALY compared to standard examination.
Conclusion	This presents for debate support for developing nationwide services for early identification and treatment of dementia	The analysis suggest substantial benefits in terms of both patient and economic outcomes for early assessment and early treatment of AD	Adding functional neuroimaging to usual diagnostic regimen is not cost-effective given the effectiveness of currently available therapies	PET may have high diagnostic accuracy, but adding it to the standard diagnostic regimen at AD clinics would yield limited, if any, benefits at very high costs

Abbreviations: AD, Alzheimer's disease; CT, computerized tomography; DSC, dynamic susceptibility-weighted contrast-enhanced; FDG, fluorodeoxyglucose; ICER, incremental cost-effectiveness ratio; MRI, magnetic resonance imaging; N/A, not applicable; NICE, National Institute for Health and Clinical Excellence; NPV, net present value; NS, not stated; PET, positron emission tomography; QALE, quality adjusted life expectancy; QALY, quality adjusted life year; SPECT, single photon emission computed tomography; UK, United Kingdom; US, United States; AU, Australia;

Table 1: continued

Moulin-Romsee et al.[27]	Silverman et al.[28]	Simon et al.[29]	Weimer et al.[30]	Wolfs et al.[31]
2005	2002	1985	2009	2009
Belgium	US	US	US	NL
Estimate the economic effects of incorporating FDG PET in the diagnostic work-up of AD	Compare the relative value of a conventional and proposed approach using PET to assess early AD	Estimating the health benefits and economic costs of using alternative scanning strategies for diagnosing dementia	Evaluates the costs and benefits of the early identification and treatment of AD patients	Evaluate the cost-effectiveness of integrated multidisciplinary diagnostic facility for dementia
Modeling	Modeling	Modeling	Modeling	Trial-based economic evaluation
Decision-tree (based upon model Silverman et al.)	Decision-tree	Decision-tree	Static	N/A
Static	Static	Static	state independent mathematical model	N/A
Synthesizing head-to-head comparisons where relevant trials do not exist	Synthesizing head-to-head comparisons where relevant trials do not exist	Synthesizing head-to-head comparisons where relevant trials do not exist	Informing decisions in the absence of hard data	N/A
Routine practice for the diagnosis of AD	Current practice for expert evaluation of dementia	CT scanning when historical or physical findings suggest a treatable illness (S-CT)	Later diagnosis and drug treatment	Diagnosis made by GP or existing regional services
FDG PET incorporated in routine practice	FDG PET after ruling out other conditions	(1) Routine evaluation with CT scanning (R-CT); (2) Routine MRI to replace routine CT (R-MRI)	Immediate diagnosis and intervention	Multidisciplinary assessment combining hospital setting and community mental health setting
A summary effect of overtreatment and under treatment is on false diagnoses	A summary effect of overtreatment and under treatment is on false diagnoses	(quality adjusted) life expectancy of patients with under-diagnosed treatable conditions	Bring forward treatment (incorrect diagnoses not modeled)	N/A
NS	Current practice standards for expert evaluation of dementia	NS	NS	GP or home visit and Geriatric Medicine and Geriatric Psychiatry hospital departments
NS	geriatric patients with early symptoms of cognitive decline	individuals aged 60, 70, or 80 and presenting with a dementing illness	AD patients suffering from cognitive decline	patients suspected of having dementia or a cognitive disorder (age 55+)
N/A	N/A	N/A	N/A	Quality of life and resource consumption data on 219 patients
NS	Health care payer	NS	Societal	Societal
NS	12 months	Lifetime	Lifetime	12 months
Cost savings per accurate diagnosis	Costs per accurate diagnoses	ICER	Net benefit	ICER
Cost-savings per accurate diagnosis ranged from 623–6110 Euro in favor of the proposed algorithm with PET	Costs per correct diagnosis by the conventional algorithm: $5,185 and proposed algorithm: $4,047	Cost per additional year of QALE from S-CT to R-CT is below $50,000. Comparing R-MRI to R-CT incremental cost ranges from $46,000 to $144,000	Early identification and treatment potentially result in large, positive net social benefits as well as positive net savings	ICER: €1267/QALY. This estimate is within an acceptable range of uncertainty
Incorporating FDG PET in the clinical work up can result in substantial benefit in terms of health-care cost savings and benefit for the patient.	Improved care can be economically achieved through incorporation of PET into the diagnostic work-up	Given current treatment limitations in dementia it appears that MRI will have little immediate health impact on this problem	Early diagnosis and treatment of AD are socially desirable in terms of increasing economic efficiency, and fiscally attractive	Integrated multidisciplinary diagnostic facility is cost-effective for the diagnosis and management of dementia in community patients

* Part of the model was described in Getsios et.al. [32]

† Part of the model was described in Neuman et.al. [33]

‡ Model type was valued as "static" if not represented by a mathematical model (e.g., when the effect was summarized in one or more fixed outcome values)

§ Model types were based on Buxton et.al. [19]

¶ Additional strategies evaluated in a sensitivity analysis are not stated

Table 2 displays the application of the CHEC list [20] on the trial-based economic evaluation. A time horizon of 1 year was considered too short for the inclusion of all relevant costs and consequences and generalizability was insufficiently addressed.

Table 2: Scores on the CHEC list of trial-based economic evaluation studies

CHEC-list Item	Wolfs et al.[31]
1. Is the study population clearly described?	Y
2. Are competing alternatives clearly described?	Y
3. Is a well-defined research question posed in answerable form?	Y
4. Is the economic study design appropriate to the stated objective?	Y
5. Is the chosen time horizon appropriate in order to include relevant costs and consequences?	N
6. Is the actual perspective chosen appropriate?	Y
7. Are all important and relevant costs for each alternative identified?	Y
8. Are all costs measured appropriately in physical units?	Y
9. Are costs valued appropriately?	Y
10. Are all important and relevant outcomes for each alternative identified?	Y
11. Are all outcomes measured appropriately?	Y
12. Are outcomes valued appropriately?	Y
13. Is an incremental analysis of costs and outcomes of alternatives performed?	Y
14. Are all future costs and outcomes discounted appropriately?	Y
15. Are all important variables, whose values are uncertain, appropriately subjected to sensitivity analysis?	Y
16. Do the conclusions follow from the data reported?	Y
17. Does the study discuss the generalizability of the results to other settings and patient/client groups?	N
18. Does the article indicate that there is no potential conflict of interest of study researcher(s) and funder(s)?	Y
19. Are ethical and distributional issues discussed appropriately?	N

Abbreviations: N, the item was not fulfilled or insufficient information was available in the article to assess the item; Y, both reviewers agreed that the study paid sufficient attention to an item

Table 3 displays the results of the quality assessment using the framework for decision-analytic models [21]. Twenty framework items required scoring by the third reviewer (J.S.); two were due to dissimilarities between the two reviewers that could not be resolved by discussion. Items in several topics were not fulfilled or were insufficiently addressed by most included studies. The following four were also considered of importance regarding this review's focus on the diagnostic aspects of the decision models. Model scope was considered insufficiently addressed because a lifetime horizon was disregarded and the population and setting were mostly not addressed. The options under evaluation were considered not fulfilled because many studies did not evaluate all feasible options or did not justify their exclusion, especially regarding combinations of diagnostic tests. Further, several items of the data identification were not addressed sufficiently, especially regarding the identification of key parameters. Lastly, according to the checklist, all types of uncertainty analyses were considered not described or not performed by almost all included studies.

Table 3: Scores on the framework for quality assessment of decision-analytic models

Study number, author, source			1: Banerjee et al.[23]	2. Getsios et al.[24]	3: McMahon et al. (2000) [25]	4: McMahon et al. (2003) [26]	5. Moulin-Romsee et al.[27]	6. Silverman et al.[28]	7: Simon et al.[29]	8: Weimer et al.[30]
Structure (S)										
S1: Statement of decision problem/objective	1	Is there a clear statement of the decision problem?	Y	Y	N	Y	Y	Y	Y	Y
	2	Is the objective of the evaluation and model specified and consistent with the stated decision problem?	Y	Y	N/A	Y	Y	Y	Y	Y
	3	Is the primary decision maker specified?	Y	N	N	N	N	Y	N	Y
S2: Statement of scope/perspective	1	Is the perspective of the model stated clearly?	Y	Y	Y	Y	N	Y	N	Y
	2	Are the model inputs consistent with the stated perspective?	N	Y	Y	Y	N/A	Y	N/A	Y
	3*	Has the scope of the model been stated and justified?	N	N	N	N	N	N	N	N
	4	Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
S3: Rationale for structure	1	Has the evidence regarding the model structure been described?	N	Y	Y	Y	Y	Y	N	Y
	2†	Is the structure of the model consistent with a coherent theory of the health condition under evaluation?	N	Y	Y	Y	Y	Y	Y	Y
	3‡	Have any competing theories regarding model structure been considered?	N	N	N	N	N	N	N	N
	4	Are the sources of data used to develop the structure of the model specified?	N	Y	Y	Y	Y	Y	N	Y
	5	Are the causal relationships described by the model structure justified appropriately?	Y	Y	Y	Y	Y	Y	Y	Y
S4: Structural assumptions	1	Are the structural assumptions transparent and justified?	Y	Y	Y	Y	Y	Y	Y	Y
	2	Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
S5: Strategies/comparators	1	Is there a clear definition of the options under evaluation?	N	Y	Y	Y	Y	Y	Y	N
	2	Have all feasible and practical options been evaluated?	N	N	Y	Y	N	N	N	Y
	3	Is there justification for the exclusion of feasible options?	N	N	Y	Y	N	N	N	N/A
S6: Model type	1	Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?	N	Y	Y	Y	Y	Y	Y	N

			1: Banerjee et al.[23]	2. Getsios et al.[24]	3: McMahon et al. (2000) [25]	4: McMahon et al. (2003) [26]	5. Moulin-Romsee et al.[27]	6. Silverman et al.[28]	7: Simon et al.[29]	8: Weimer et al.[30]
S7: Time horizon	1	Is the time horizon of the model sufficient to reflect all important differences between options?	Y	Y	Y	Y	N	Y	Y	Y
	2	Is the time horizon of the model, and the duration of treatment and treatment effect described and justified?	N	Y	Y	Y	N/A	Y	Y	N
	3	Has a lifetime horizon been used?	N	N	N	N	N	N	Y	Y
	4	If not, has a shorter time horizon been justified?	N	N	Y	Y	N	N	N/A	N/A
S8: Disease states/pathways	1§	Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?	N	Y	Y	Y	Y	Y	Y	N
S9: Cycle length	1	Is the cycle length defined and justified in terms of the natural history of disease?	N/A	N/A	Y**	Y**	N/A	N/A	N/A	Y
Data (D)										
D1: Data identification	1	Are the data identification methods transparent and appropriate given the objectives of the model?	N	Y**	N	N	Y	Y	N	N
	2	Where choices have been made between data sources, are these justified appropriately?	N	Y**	N	Y	Y	Y	N	N
	3	Has particular attention been paid to identifying data for the important parameters in the model?	N	Y	Y	Y	Y	Y	Y	Y
	4	Has the process of selecting key parameters been justified and systematic methods used to identify the most appropriate data?	N/A	N	N	N	N	N	N	N
	5	Has the quality of the data been assessed appropriately?	N	Y**	N	N	Y	Y	N	N
	6	Where expert opinion has been used, are the methods described and justified?	N/A	N/A	Y	Y	N/A	N/A	N	N/A
D2: Pre-model data analysis	1	Are the pre-model data analysis methodology based on justifiable statistical and epidemiological techniques?	N/A	Y**	Y**	Y**	N/A	N/A	N/A	Y
D2a: baseline data	1	Is the choice of baseline data described and justified?	N	Y**	Y	Y	Y	Y	Y	Y
	2	Are transition probabilities calculated appropriately?	N/A	N/A	Y**	Y**	Y	Y	Y	Y
	3	Has a half cycle correction been applied to both cost and outcome?	N/A	N/A	N**	N**	N/A	N/A	N/A	N/A
	4	If not, has this omission been justified?	N/A	N/A	N**	N**	N/A	N/A	N/A	N/A

			1: Banerjee et al.[23]	2. Getsios et al.[24]	3: McMahon et al. (2000) [25]	4: McMahon et al. (2003) [26]	5: Moulin-Romsee et al.[27]	6: Silverman et al.[28]	7: Simon et al.[29]	8: Weimer et al.[30]
D2b: treatment effects and diagnostic accuracy	1¶	If relative diagnostic accuracy have been derived from trial data, have they been synthesised using appropriate techniques?	N/A	N/A	Y	Y	Y	Y	Y	N/A
	2#	Have the methods and assumptions used to extrapolate diagnostic accuracy to final outcomes been documented and justified?	N/A	N/A	Y	Y	Y	Y	Y	N/A
	3	Have alternative assumptions been explored through sensitivity analysis?	N/A	Y	Y	Y	Y	Y	N	N/A
	4	Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?	N/A	Y	N/A	N/A	N/A	N/A	N/A	N/A
	5	Have alternative assumptions been explored through sensitivity analysis?	N/A	Y	N/A	N/A	N/A	N/A	N/A	N/A
D2c: quality-of-life weights (utilities)	1	Are the utilities incorporated into the model appropriate?	N/A	Y	Y	Y	N/A	N/A	Y	Y
	2	Is the source for the utility weights referenced?	N/A	Y	Y	Y	N/A	N/A	N	Y
	3	Are the methods of derivation for the utility weights justified?	N/A	Y	Y	Y	N/A	N/A	N	N/A
D3: Data incorporation	1	Have all data incorporated into the model been described and referenced in sufficient detail?	Y	Y	Y	Y	N	Y	Y	Y
	2	Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	3	Is the process of data incorporation transparent?	Y	Y	Y	Y	Y	Y	Y	Y
	4	If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?	N/A	Y**	N/A	N/A	N/A	N/A	N/A	Y
	5	If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?	N/A	Y	N/A	N/A	N/A	N/A	N/A	N
D4: Assessment of uncertainty	1	Have the four principal types of uncertainty been addressed?	N	N	N	N	N	N	N	N
	2	If not, has the omission of particular forms of uncertainty been justified?	N	N	N	N	N	N	N	N
D4a: methodological	1	have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	N	N	N	N	N	N	N	N

			1: Banerjee et al.[23]	2. Getsios et al.[24]	3: McMahon et al. (2000) [25]	4: McMahon et al. (2003) [26]	5. Moulin-Romsee et al.[27]	6. Silverman et al.[28]	7: Simon et al.[29]	8: Weimer et al.[30]
D4b: structural	1	is there evidence that structural uncertainties have been addressed via sensitivity analysis?	N	N	N	Y	N	Y	N	Y
D4c: heterogeneity	1	has heterogeneity been dealt with by running the model separately for different subgroups?	N	N	N	N	N	N	Y	Y
D4d: parameter	1	are the methods of assessment of parameter uncertainty appropriate?	N	Y	N	N	N	N	N	Y
	2	has probabilistic sensitivity analysis been done, if not, has this been justified?	N	Y	N	N	N	N	N	N
	3	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	N	N/A	Y	Y	Y	Y	Y	N/A
Consistency (C)										
C1: Internal consistency	1	Is there evidence that the mathematical logic of the model has been tested thoroughly before use?	N	N	N	N	N	N	N	N
C2: External consistency	1	Are the conclusions valid given the data presented?	Y	Y	Y	Y	Y	Y	Y	Y
	2	Are any counterintuitive results from the model explained and justified?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	3	If the model has been calibrated against independent data, have any differences been explained and justified?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	4	Have the results of the model been compared with those of previous models and any differences in results explained?	N	Y	N	Y	N	Y	N	N

Abbreviations: N, item was not fulfilled or insufficient information was available; N/A, not applicable to the study or referred to a previous question scored with "N"; Y, both reviewers agreed sufficient attention was paid to an item.

* According to the framework, the model scope should include the perspective, involved technologies, population, setting, and time horizon at the outset of the study.

† This item required expert knowledge in the field of AD. The structure of a model was therefore considered consistent with a coherent theory of AD if it included the following basic elements in the diagnostic process: history taking, examination of cognition, and structural imaging [43, 44]. Additionally, if treatment was included, it had to reflect cognition and a progressive decline according to a clinical measurement scale [14, 45] (e.g., the Mini-Mental State Examination (MMSE) or Clinical Dementia Rating).

‡ This item required expert knowledge in the field of AD. As stated in the introduction, new research criteria assigning to biomarkers a fundamental role in the diagnostic process have recently been developed. Applying these research criteria to the diagnostic process is considered a competing theory regarding model structure.

§ This item required expert knowledge in the field of AD. The underlying biological process was considered appropriate if disease states or pathways reflected cognition.

¶ This item was changed to derivation of diagnostic accuracy.

This item was changed to extrapolation of diagnostic accuracy, according to its relevance within the hierarchical model of diagnostic efficacy [46].

** Information to score the item was only retrieved from the additional publication that described part of the model

DISCUSSION

The objective of the present study was to systematically review the literature on economic evaluations of interventions for the early diagnosis of AD and related disorders and to describe their methodological characteristics. Eight decision-analytic modeling studies and one trial-based economic evaluation were retrieved. Study quality was assessed using the framework for decision-analytic models [21] and the CHEC list [20], and overall the results reflect considerable methodological quality. However, population and diagnostic setting (scope), interventions under evaluation, selection and quality assessment of key diagnostic input parameters, and uncertainty analyses were not fulfilled or were insufficiently addressed by most included studies. Although the study conclusions were considered valid given the data presented, the diversity in study characteristics and methodology could have a major impact on the cost-effectiveness estimate and do not allow for a valid comparison of the cost-effectiveness outcomes.

We discuss the applicability of the diagnostic part of existing decision models for the economic evaluation of the recently revised diagnostic research criteria for AD. Two aspects were considered of major importance as indicated by the framework for diagnostic test evaluation by Van den Bruel et al [11].

The first aspect regards diagnostic test accuracy, which is highly dependent on patient population characteristics, clinical setting and (patient selection by) results of previously performed tests [10]. Most studies left the above aspects insufficiently described (i.e.: the scope was minimally addressed). Because the newly developed research criteria are targeted on various disease severity in specialized centers, it is important to specify the corresponding targeted population and setting in decision analyses to enable their evaluation. Furthermore, most studies limited the diagnostic options under evaluation and minimally addressed sensitivity analyses to evaluate alternative diagnostic scenarios. Therefore, possible diagnostic strategies such as using biomarkers as add-on or triage were insufficiently evaluated. Furthermore, the newly developed research criteria divide the AD biomarkers into two categories (biomarkers of β-Amyloid deposition and of neuronal injury) and emphasized the importance of evaluating different combinations. This was insufficiently explored by the options under evaluation or by the sensitivity analyses. Questions on optimal decision-making when biomarker results contradict or when to stop testing were not answered.

The second aspect refers to the impact on patient outcome. Most studies in this review included only the impact of pharmacological treatment, which, in our opinion, underexposes the possible nonmedical effects in symptomatic predementia patients. Because current treatment guidelines do not advise medication in this disease phase and nonpharmacological treatments are mainly aimed at dementia-related symptoms [34], it leaves the diagnostic test impact mainly to nonmedical aspects [35], such as anxiety, depression, psychological well-being, (financial) decision-making, and future care planning [4], [36] and [37]).

Recommendations

Following these findings, three recommendations for future decision models to enable the evaluation of the new diagnostic research criteria can be made. First, we advise to describe in detail the model scope, including the intended goal as timely diagnosis or symptomatic predementia diagnosis, the population and clinical setting characteristics regarding disease state

at the moment of diagnosis, and what diagnostic tests already have been performed in the population. We advise to describe these items separately as targeted by the decision model and as applied in the retrieved (literature) evidence on key diagnostic parameters such as test sensitivity and specificity.

Second, we advise to include a variety of possible test combinations as options under evaluation or to address these by structural sensitivity analyses (e.g., the study of McMahon et al [26]). At last, we recommend representing the diagnostic strategy by a decision tree, including probabilities for each test outcome and their corresponding non-medical consequences, followed by a Markov model or discrete event simulation describing long-term disease progression [38] (for example as applied by two of the included studies [25] and [26]).

Limitations

Both the CHEC-list for trial-based economic evaluations and the framework for quality assessment of decision-analytic models support the assessment of studies for a review. Judgments, however, remain subjective and therefore the use of three reviewers was adopted for this evaluation. We recommend that such a procedure be used for future studies. Also, several items in the framework for decision-analytic models were interpreted in different ways and therefore required resolution by a third reviewer (JS).

Furthermore, items including harm of testing, willingness to undergo a test, consequences of (false) diagnosis, purpose of testing regarding screening, diagnosis and prognosis, and place in the clinical pathway regarding other tests that have been performed are important in diagnostic test evaluation. However, major flaws on these items were not differentiated to minor flaws by using the quality assessment checklists. We applied a qualitative summary to enable such distinction though ideally a quantitative score would be applied that combines the different constructs of a model and assessed their relative importance to rate the quality of the studies.

Non-medical consequences of testing in the absence of direct clinical benefits are considered important. Quality of life scales are potentially useful though rarely capture all possible test effects and therefore specific instruments are advised to capture the nonmedical effects [35]. Retrieving quality of life estimates will be a challenge for future research. Meanwhile, willingness to pay estimates can, despite of challenges in obtaining reliable estimates, help to isolate the non-medical value [39]. Neumann et al [40] found that on average people are willing to pay about $450 for an AD test in absence of treatment and that a positive test result influences future planning in part of the subjects. If non-medical consequences could not be quantified they should be assessed qualitatively.

By excluding screening or risk assessment instruments we have not elaborated on the applicability of current decision models for the evaluation of the recently revised criteria for the pre-symptomatic pre-dementia phase. These criteria are however indicated as "strictly for research purposes only" and at this moment less relevant to evaluate for applicability in clinical practice.

Overall, the reviewers found that insufficient information was available for the assessment of all items of the framework for decision-analytic models, thereby increasing uncertainty about the interpretation and generalizability of the results. However, the quantity of publishable information is restricted by journal word count limits. The space available to describe all relevant aspects in detail is limited, especially for economic evaluations of diagnostic interventions, which

most likely include both a diagnostic intervention and treatment options. A practical solution currently is to include an extensive appendix, as demonstrated by the model of Getsios et al [32] and, in a different field, by Van Gestel et al [41].

Issues of implementation were not included in the assessment instruments, though they may be of importance in assessing economic evaluations of diagnostic interventions for AD [16]. MRI and PET scans (and the production of PET tracers), for example, are not available in all hospitals. As a result, additional logistic effort or the reorganization of services may be necessary.

All included studies have been performed before the publication of the revised criteria and obviously therefore have not necessarily adopted the preferred model scope or methodology to compare a variety of test combinations. An example of an ongoing study adopting such scope and taking account the recommendations from this review is the LeARN study [42] (clinicaltrials.gov/ct2/show/NCT01450891) from which the results will be available in 2013.

Conclusion

A systematic review of economic evaluations of interventions for the early diagnosis of AD and related disorders was conducted with the aim of describing their general and methodological characteristics. The search retrieved eight decision-analytic modeling studies and one trial-based economic evaluation. The overall results reflect considerable methodological quality. However, diversity among the study objective and characteristics was considerable and the topics scope, options under evaluation, data identification and uncertainty analyses were not fulfilled or insufficiently addressed. To apply current decision models for the assessment of the recently revised diagnostic research criteria for AD we recommend to describe in detail the model scope, to include a variety of possible test combinations as options under evaluation, and to apply an explicit quality of life estimate to reflect the impact of non-medical aspects.

REFERENCES

[1] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, C.P. Ferri The global prevalence of dementia: A systematic review and metaanalysis Alzheimers Dement, 9 (2013), pp. 63–75.e2

[2] Mathers C, Boerma T, Ma Fat D. The global burden of disease: 2004 update. World Health Organization. Available at: http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html. Retrieved at July 2011.

[3] A. Wimo, L. Jonsson, J. Bond, M. Prince, B. Winblad The worldwide economic impact of dementia 2010 Alzheimers Dement, 9 (2013), pp. 1–11.e13

[4] Prince M, Bryce R, Ferri C. World Alzheimer Report 2011. Alzheimer's Disease International. Available at: http://www.alz.co.uk/research/worldreport/. Retrieved at May 2012.

[5] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E.M. Stadlan 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 (1984), pp. 939–944

[6]   G.M. McKhann, D.S. Knopman, H. Chertkow, B.T. Hyman, C.R. Jack Jr., C.H. Kawas et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease Alzheimers Dement, 7 (2011), pp. 263–269

[7]   M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox et al. The diagnosis of mild cognitive impairment 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 (2011), pp. 270–279

[8]   R.A. Sperling, P.S. Aisen, L.A. Beckett, D.A. Bennett, S. Craft, A.M. Fagan et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease Alzheimers Dement, 7 (2011), pp. 280–292

[9]   C.R. Jack Jr., M.S. Albert, D.S. Knopman, G.M. McKhann, R.A. Sperling, M.C. Carrillo et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease Alzheimers Dement, 7 (2011), pp. 257–262

[10] J.A. Knottnerus, F. Buntinx The evidence base of clinical diagnosis: theory and methods of diagnostic research (2nd ed.)Wiley-Blackwell, Hoboken, NJ (2009)

[11] A. Van den Bruel, I. Cleemput, B. Aertgeerts, D. Ramaekers, F. Buntinx The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed J Clin Epidemiol, 60 (2007), pp. 1116–1122

[12] A.H. Briggs, K. Claxton, M.J. Sculpher Decision modelling for health economic evaluation Oxford University Press, Oxford, United Kingdom (2006)

[13] M.C. Weinstein Recent developments in decision-analytic modelling for economic evaluation Pharmacoeconomics, 24 (2006), pp. 1043–1053

[14] Cohen JT, Neumann PJ. Decision analytic models for Alzheimer's disease: state of the art and future directions. 2008;4:212–22.

[15] Mauskopf JA, Paul JE, Grant DM, Stergachis A. The role of cost-consequence analysis in healthcare decision-making. 1998;13:277–88.

[16] MF Drummond, MJ Sculpher, GW Torrance, BJ O'Brien, GL Stoddart Methods for the economic evaluation of health care programmes (3rd ed.)Oxford University Press, New York (2005)

[17] V. Alton, I. Eckerlund, A. Norlund Health economic evaluations: how to find them Int J Technol Assess Health Care, 22 (2006), pp. 512–517

[18] F. Sassi, L. Archard, D. McDaid Searching literature databases for health care economic evaluations: how systematic can we afford to be? Med Care, 40 (2002), pp. 387–394

[19] M.J. Buxton, M.F. Drummond, B.A. Van Hout, R.L. Prince, T.A. Sheldon, T. Szucs et al. Modelling in economic evaluation: an unavoidable fact of life Health Econ, 6 (1997), pp. 217–227

[20] S. Evers, M. Goossens, H. de Vet, M. van Tulder, A. Ament Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria Int J Technol Assess Health Care, 21 (2005), pp. 240–245

[21] Z. Philips, L. Bojke, M. Sculpher, K. Claxton, S. Golder Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment Pharmacoeconomics, 24 (2006), pp. 355–371

[22] Z. Philips, L. Ginnelly, M. Sculpher, K. Claxton, S. Golder, R. Riemsma et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment Health Technol Assess, 8 (2004) iii-iv, ix-xi, 1–158

[23] S. Banerjee, R. Wittenberg Clinical and cost effectiveness of services for early diagnosis and intervention in dementia Int J Geriatr Psychiatry, 24 (2009), pp. 748–754

[24] D Getsios, S Blume, KJ Ishak, G Maclaine, L Hernandez An economic evaluation of early assessment for Alzheimer's disease in the United Kingdom Alzheimers Dement,8(2012)pp.22–30

[25] P.M. McMahon, S.S. Araki, P.J. Neumann, G.J. Harris, G.S. Gazelle Cost-effectiveness of functional imaging tests in the diagnosis of Alzheimer disease Radiology, 217 (2000), pp. 58–68

[26] P.M. McMahon, S.S. Araki, E.A. Sandberg, P.J. Neumann, G.S. Gazelle Cost-effectiveness of PET in the diagnosis of Alzheimer disease Radiology, 228 (2003), pp. 515–522

[27] G. Moulin-Romsee, A. Maes, D. Silverman, L. Mortelmans, K. Van Laere Cost-effectiveness of 18F-fluorodeoxyglucose positron emission tomography in the assessment of early dementia from a Belgian and European perspective Eur J Neurol, 12 (2005), pp. 254–263

[28] D.H. Silverman, J.L. Cummings, G.W. Small, S.S. Gambhir, W. Chen, J. Czernin et al. Added clinical benefit of incorporating 2-deoxy-2-[18F]fluoro-D-glucose with positron emission tomography into the clinical evaluation of patients with cognitive impairment Mol Imaging Biol, 4 (2002), pp. 283–293

[29] D.G. Simon, M.F. Lubin Cost-effectiveness of computerized tomography and magnetic resonance imaging in dementia Med Decis Making, 5 (1985), pp. 335–354

[30] D.L. Weimer, M.A. Sager Early identification and treatment of Alzheimer's disease: social and fiscal outcomes Alzheimers Dement, 5 (2009), pp. 215–226

[31] C.A. Wolfs, C.D. Dirksen, A. Kessels, J.L. Severens, F.R. Verhey Economic evaluation of an integrated diagnostic approach for psychogeriatric patients: results of a randomized controlled trial Arch Gen Psychiatry, 66 (2009), pp. 313–323

[32] D. Getsios, S. Blume, K.J. Ishak, G.D. Maclaine Cost effectiveness of Donepezil in the treatment of mild to moderate Alzheimer's disease: a UK evaluation using discrete-event simulation Pharmacoeconomics, 28 (2010), pp. 411–427

[33] P.J. Neumann, R.C. Hermann, K.M. Kuntz, S.S. Araki, S.B. Duff, J. Leon et al. Cost-effectiveness of Donepezil in the treatment of mild or moderate Alzheimer's disease Neurology, 52 (1999), pp. 1138–1145

[34] J. Olazaran, B. Reisberg, L. Clare, I. Cruz, J. Pena-Casanova, T. Del Ser et al. Nonpharmacological therapies in Alzheimer's disease: a systematic review of efficacy Dement Geriatr Cogn Disord, 30 (2010), pp. 161–178

[35] P.M. Bossuyt, K. McCaffery Additional patient outcomes and pathways in evaluations of testing Med Decis Making, 29 (2009), pp. E30–E38

[36] S. Iliffe, J. Manthorpe, A. Eden Sooner or later? Issues in the early diagnosis of dementia in general practice: a qualitative study Fam Pract, 20 (2003), pp. 376–381

[37] S. Lliffe, J. Manthorpe The hazards of early recognition of dementia: a risk assessment Aging Ment Health, 8 (2004), pp. 99–105

[38] J.D. Schaafsma, Y. van der Graaf, G.J. Rinkel, E. Buskens Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness J Clin Epidemiol, 62 (2009), pp. 1248–1252

[39] D.W. Lee, P.J. Neumann, J.A. Rizzo Understanding the medical and nonmedical value of diagnostic testing Value Health, 13 (2010), pp. 310–314

[40] P.J. Neumann, J.T. Cohen, J.K. Hammitt, T.W. Concannon, H.R. Auerbach, C. Fang et al. Willingness-to-pay for predictive tests with no immediate treatment implications: a survey of US residents Health Econ, 21 (2012), pp. 238–251

[41] A. van Gestel, J.L. Severens, C.A. Webers, H.J. Beckers, N.M. Jansonius, J.S. Schouten Modeling complex treatment strategies: construction and validation of a discrete event simulation model for glaucoma Value Health, 13 (2010), pp. 358–367

[42] R.L. Handels, P. Aalten, C.A. Wolfs, M. Olderikkert, P. Scheltens, P.J. Visser et al. Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease: the research protocol of a prospective cohort study BMC Neurol, 12 (2012), p. 72

[43] J. Hort, J.T. O'Brien, G. Gainotti, T. Pirttila, B.O. Popescu, I. Rektorova et al. EFNS guidelines for the diagnosis and management of Alzheimer's disease Eur J Neurol, 17 (2010), pp. 1236–1248

[44] D.S. Knopman, S.T. DeKosky, J.L. Cummings, H. Chui, J. Corey-Bloom, N. Relkin et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology Neurology, 56 (2001), pp. 1143–1153

[45] C. Green Modelling disease progression in Alzheimer's disease: a review of modelling methods used for cost-effectiveness analysis Pharmacoeconomics, 25 (2007), pp. 735–750

[46] D.G. Fryback, J.R. Thornbury The efficacy of diagnostic imaging Med Decis Making, 11 (1991), pp. 88–94

APPENDIX 1. SEARCH STRATEGY PUBMED UP TO MARCH 2011

#	Search term
#46	Search #8 and #24 and #45
#45	Search #27 or #44
#44	Search #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43
#43	Search pharmaco-economic∗[Title/Abstract]
#42	Search pharmacoeconomic∗[Title/Abstract]
#41	Search cost$identificat∗[Title/Abstract]
#40	Search cost$comparison∗[Title/Abstract]
#39	Search cost$consequence∗[Title/Abstract]
#38	Search cost$minimi∗[Title/Abstract]
#37	Search cost$utili∗[Title/Abstract]
#36	Search cost$stud∗[Title/Abstract]
#35	Search cost$analy∗[Title/Abstract]
#34	Search cost$evaluation∗[Title/Abstract]
#33	Search cost$benefit∗[Title/Abstract]
#32	Search cost$effective∗[Title/Abstract]
#31	Search costing[Title/Abstract]
#30	Search costly[Title/Abstract]
#29	Search costs[Title/Abstract]
#28	Search cost[Title/Abstract]
#27	Search #25 or #26
#26	Search models, economic/
#25	Search cost and cost analysis/
#24	Search #12 or #23
#23	Search #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22
#22	Search cognitively$impair∗[Title/Abstract]
#21	Search memory$disorder[Title/Abstract]
#20	Search MCI[Title/Abstract]
#19	Search cognitive$impair∗[Title/Abstract]
#18	Search cognitive$disorder[Title/Abstract]
#17	Search psychogeriatr∗[Title/Abstract]
#16	Search neurodegen∗[Title/Abstract]
#15	Search cognitive$dysfunct∗[Title/Abstract]
#14	Search dement∗[Title/Abstract]
#13	Search alzheimer∗[Title/Abstract]
#12	Search #9 or #10 or #11
#11	Search neurodegenerative diseases/
#10	Search dementia/
#9	Search alzheimer's disease/
#8	Search #6 or #7
#7	Search diagnos∗[Title/Abstract]
#6	Search #1 or #2 or #3 or #4 or #5
#5	Search discriminant analysis/
#4	Search predictive value of tests/
#3	Search diagnosis, differential/
#2	Search sensitivity and specificity/
#1	Search diagnosis/

APPENDIX 2. SEARCH STRATEGY NATIONAL INSTITUTE FOR HEALTH RESEARCH ECONOMIC EVALUATION DATABASE UP TO MARCH 2011

#	Search term
#1	MeSH Diagnosis EXPLODE 1
#2	MeSH Sensitivity and Specificity EXPLODE 1 2 3 4 5 6 7
#3	MeSH Diagnosis, Differential EXPLODE 1
#4	MeSH Predictive Value of Tests EXPLODE 1 2 3
#5	MeSH Discriminant Analysis EXPLODE 1 2 3
#6	Diagnos*
#7	#1 or #2 or #3 or #4 or #5 or #6
#8	MeSH Alzheimer Disease EXPLODE 1 2 3
#9	MeSH Dementia EXPLODE 1 2
#10	MeSH Neurodegenerative Diseases EXPLODE 1
#11	Alzheimer* OR dement* OR cognitive$dysfunct* OR neurodegen* OR psychogeriatr* OR cognitive$disorder OR cognitive$impair* OR MCI OR memory$disorder OR cognitively$impair*
#12	#8 or #9 or #10 or #11
#13	#7 AND #12

Abbreviation: MeSH, Medical Subject Headings.

* the wildcard character

CHAPTER 6

NATURAL PROGRESSION MODEL OF COGNITION AND PHYSICAL FUNCTIONING AMONG PEOPLE WITH MILD COGNITIVE IMPAIRMENT AND ALZHEIMER'S DISEASE

Ron Handels, Weili Xu, Debora Rizzuto, Barbara Caracciolo, Rui Wang, Bengt Winblad, Frans Verhey, Johan Severens, Laura Fratiglioni, Manuela Joore, and Anders Wimo

Abstract

Background: Empirical models of the natural history of Alzheimer's disease (AD) may help to evaluate new interventions for AD.

Objective: We aimed to estimate AD-free survival time in people with mild cognitive impairment (MCI) and decline of cognitive and physical function in AD cases.

Methods: Within the Kungsholmen project, 153 incident MCI and 323 incident AD cases (international criteria) were identified during 9 years of follow-up in a cognitively healthy cohort of elderly people aged ≥75 at baseline (n = 1,082). Global cognitive function was assessed with the Mini-Mental State Examination (MMSE), and daily life function was evaluated with the Katz index of activities of daily living (ADL) at each follow-up examination. Data were analyzed using parametric survival analysis and mixed effect models.

Results: Median AD-free survival time of 153 participants with incident MCI was 3.5 years. Among 323 incident AD cases, the cognitive decline was 1.84 MMSE points per year, which was significantly associated with age. Physical functioning declined by 0.38 ADL points per year and was significantly associated with age, education, and MMSE, but not with gender.

Conclusion: Elderly people with MCI may develop AD in approximately 3.5 years. Both cognitive and physical function may decline gradually after AD onset. The empirical models can be used to evaluate long-term disease progression of new interventions for AD.

INTRODUCTION

Alzheimer's disease (AD), the most common cause of dementia, is characterized by a gradual onset and a decline of cognition and functional ability to the stages of complete dependence on informal or formal care. When studying the disease, it is therefore important to adopt a long time horizon to capture all disease-related events. Disease modifying treatments are being developed to reduce individual and societal burden, but randomized trials to evaluate such interventions require many years of follow-up to capture their long-term consequences, whereas most trials have employed a short follow-up period [1]. Long-term follow-up requires large investments of resources, and if new interventions emerge during the follow-up period, ethical concerns arise regarding withholding possibly successful treatment for a prolonged period. As an alternative, empirical models can be used to predict long-term consequences by integrating trial outcomes with estimations of natural disease progression [2, 3]. In addition, empirical models are a crucial component of economic decision models [4] that generate evidence for care policy making.

Natural progression models in AD have been developed in several studies [5], mostly among clinical samples or prevalent AD dementia cases. However, disease modifying treatments are supposed to be effective in early (pre-dementia) AD, thus long-term data on the natural course are required to evaluate their effectiveness. Such target populations have not been reflected by previous studies, leaving an urgent need for population-based empirical models that describe the long-term natural progression of the dementia and pre-dementia phases of AD. In the present study, we aimed to build empirical models that estimate (1) the time from incident mild cognitive impairment (MCI) to AD-type dementia and (2) the changes of cognition and function in incident AD dementia cased from a population-based cohort.

METHODS

Study sample

The study sample was derived from the Kungsholmen Project, a population-based cohort study on aging and dementia, which has been fully described elsewhere [6, 7]. Briefly, all registered inhabitants of the Kungsholmen district of Stockholm, Sweden, who were aged ≥75 years in October 1987, were initially invited to participate in the project. At baseline, 225 of the 1,810 participants were diagnosed with dementia according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition (DSM-III-R) [8], based on a 2-phase survey, and 110 participants refused the extensive evaluations. Of the remaining 1,475 dementia-free persons, 355 with MCI (130 with amnestic MCI (aMCI) and 225 with other cognitive impairment not demented (OCIND)) at baseline and 38 with very low global cognitive status in the absence of a dementia diagnosis (Mini-Mental State Examination (MMSE) [9]<20) were excluded, leaving 1,082 cognitively healthy subjects at baseline.

The participants of the present study were persons with incident MCI and AD-type dementia (either AD or mixed AD & vascular dementia). A 6-year instead of 9-year follow-up for incident MCI was applied to preserve a 3-year exposure term (from 6–9 years) for the progression of MCI to AD dementia as explained below.

During the 9-year follow-up, three sets of clinical examinations were carried out, with average intervals of 3 years. Informed consent was obtained for all participants, with informants providing

consent for cognitively impaired persons. The ethics committee at the Karolinska Institutet, Stockholm, approved all phases of the Kungsholmen Project.

Data collection

Data on demographic features (i.e., age, gender, and education) was collected at baseline using standardized protocols [6, 7].

Global cognitive functioning was assessed with the MMSE, and dependency was assessed using the Katz index of activities of daily living (ADL) [10] with scores ranging from 0 (not dependent for ADL) to 6 (fully dependent for ADL).

Diagnosis of dementia

During the follow-up period, a diagnosis of dementia (including both questionable and definite diagnoses) was established by the examining physicians, based on a comprehensive clinical examination and cognitive tests according to the DSM-III-R criteria [11]. The diagnostic criteria applied were equivalent to probable AD according to the criteria of the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association [12], and according to those of the National Institute of Neurological Disorders and Stroke-Association Internationale pour la Recherche et l'Enseignement en Neurosciences [13].

Definition of MCI

aMCI was defined according to the original Mayo clinic criteria, and operationalized according to previous research [14, 15], as follows: (1) presence of a memory complaint reported by the participant or by a close informant during the nurse interview; (2) preserved general cognitive functioning, defined as scoring above the minus 1 SD cut-off on age and education adjusted MMSE means; (3) absence of dementia, verified by clinical examination; (4) preserved functional independence defined as no impairment on the Katz ADL scale; (5) presence of objective memory impairment defined as scoring ≥1.5 SD below age- and education-specific means on a verbal memory task of free recall of slowly and rapidly presented words [16]. All cases with global cognitive impairment that did not fulfill criteria for dementia were classified as OCIND and operationalized according to previous research [17] as follows: (1) impaired general cognitive function, defined as scoring 1 SD or more below age and education adjusted means on the MMSE derived from the dementia free population at baseline; and (2) absence of dementia, verified by clinical examination. aMCI and OCIND were mutually exclusive in the present study, therefore a broader category of MCI was created which included cases classified as aMCI or OCIND. The analyses were based on the incident cases of AD detected at the 3-, 6-, or 9-year follow-up measurements and incident cases of MCI detected at the 3- or 6-year follow-up measurement.

Statistical analysis

Survival analysis was applied to estimate the time from incident MCI to AD-type dementia. The effect of age at diagnosis of MCI, gender, education, MMSE, and ADL at diagnosis, and all 2-way

interactions was assessed. A stepwise procedure was used (removing interactions with highest p-values first until p < 0.05, followed by the predictors).

A mixed model with random subject effects was applied to determine the decline in cognition and ADL over time in incident AD participants. A stepwise procedure was used and predictors were included if the goodness-of-fit statistics–2 log likelihood change and Wald z of the predictor were significant. The following steps were used to determine the final MMSE prediction model: (1) include time, as years after being diagnosed with AD; (2) include a random intercept; (3) determine if time is non-linear by stepwise adding a higher-order polynomial of time ($time^2$, $time^2$ + $time^{2+n}$, etc.); (4) include a random time factor; (5) include gender, age, and education and all 2-way interactions and remove interactions with highest p-values first until p < 0.05, followed by predictors.

A similar procedure was used to analyze decline in ADL, and the effect of MMSE was also determined.

The onset of MCI as well as that of AD was assumed to have taken place in the middle of each follow-up interval (each lasting an average of 3 years). This was operationalized by adding a time correction of 1.5 years to all diagnoses. Survival analysis was performed using Stata-12, mixed effect models using SPSS-20.

RESULTS

Out of 1,082 cognitively healthy participants at baseline, 153 developed MCI (40 aMCI and 113 OCIND) and 323 developed AD during the 6 and 9 years of follow-up, respectively. The mean age at diagnosis of MCI was 83 years, while the mean age at AD diagnosis was 87 years (Table 1). Figure 1 provides an overview of the observed MMSE scores over time in the pre-dementia phase (captured by the survival analysis) and dementia phase (captured by the regression model).

Table 1: Characteristics of the participants with incident MCI and AD dementia.

Characteristics	Incident MCI (n=153)	Incident AD (n=323)
Age in years; mean (SD)	83.4 (4.0)	86.7 (4.1)
Female; %	75%	83%
Years of education; mean (SD)	8.5 (3.0)	8.2 (2.9)
MMSE score; mean (SD)	24.4 (2.1)	19.7 (5.0)
Katz ADL score; mean (SD)	0.4 (0.7)	1.2 (1.7)

AD-free survival

Among the 153 participants with incident MCI, 48 (31%) developed AD dementia, after a median time of 3.03 years (658 person-years). The incidence rate was 0.073 (95% CI: 0.055 to 0.097). Twenty-nine percent of the participants died during follow-up.

Univariate analysis only showed a significant effect of gender. In the multivariate stepwise analyses, women had a significantly shorter time from MCI to AD than the men (4.2 and 4.6 years, respectively; hazard ratio = 0.38) and none of the 2 way interactions were significant (Table 2). The observed times until 90% and 75% of the MCI cohort were still AD-free were 2.9 and 3.2 years, respectively. The 50% AD-free survival was not reached within the 6-year observation

period. Using the fitted model, the estimated times until 90%, 75%, and 50% of the MCI cohort had survived without developing AD dementia were 2.8, 4.6, and 7.1 years, respectively (see Box 1).

Table 2: Hazard ratio (95% CI) of AD in the MCI cohort using a parametric survival model with Weibull distribution.

Factors	AD Hazard Ratio (95% CI) univariate analyses	AD Hazard Ratio (95% CI) multivariate analysis
Gender (male)	0.38 (0.16 to 0.90)*	0.38 (0.16 to 0.90)*
Age at MCI diagnosis	1.06 (0.99 to 1.14)	
Education	0.94 (0.84 to 1.05)	
MMSE at MCI diagnosis	0.92 (0.80 to 1.07)	
Katz at MCI diagnosis	1.17 (0.72 to 1.88)	

* p<0.05

BOX 1

The survivor function is described by equation (1), where *S(t)* is the proportion of AD-free survival, *t* is time in years, *a* is the exponent of the survival analysis coefficient estimates and *p* is the Weibull shape parameter. This function can be rewritten to estimate the time until a specific proportion (S) of an MCI cohort has progressed to AD dementia, where gender 0 = female and gender 1 = male (equation 2).

$$S(t) = e^{-at^p} \tag{1}$$

$$t = \sqrt[p]{\frac{\ln(S)}{-a}} \tag{2}$$

$$t = \sqrt[2.01]{\frac{\ln(S)}{-e^{-4.06-0.96 gender_i}}} \tag{3}$$

The course of the MMSE can be summarized by regression formula (4) and the course of ADL by regression formula (5), where *r* is a random number from a normal distribution to reflect the variance in the random effects, and *i* is the individual participant.

$$MMSE_i = \left(26.87 \pm \sqrt{2.0}r_i\right) - \left(3.26 \pm \sqrt{1.86}r_i\right)time_i - 0.35(age_i - 75) + 0.10 time_i(age_i - 75) + e_i \tag{4}$$

$$Katz\ index_i = \left(-0.82 \pm \sqrt{0.71}r_i\right) + 0.26 time_i + 0.26(age_i - 75) + 0.06 education - 0.02 MMSE - 0.01 MMSE_i(age_i - 75) - 0.01 MMSE_i time_i + e_i \tag{5}$$

Decline in cognitive and physical function

For the 323 participants who developed AD during follow-up, 313 MMSE scores were available at the moment of AD diagnosis, 109 at 3 years after diagnosis, and 28 at 6 years after diagnosis. Forty-nine percent of the participants died during the follow-up.

Figure 2a presents the observed average MMSE scores over time, with 0 representing the moment of AD diagnosis. The univariate analyses showed that age and time (as years after being diagnosed with AD) significantly predicted the MMSE score, with an average rate of decline of −1.84 MMSE points per year. The multivariate model showed that time and age significantly predicted a decrease in MMSE score (Table 3). The interaction between time and age indicates a decreasing rate of decline over time.

For the 323 participants who developed AD during follow-up, 318 Katz ADL scores were available at the moment of AD diagnosis, 109 at 3 years after diagnosis, and 28 at 6 years after diagnosis.

Figure 2b presents the observed average Katz ADL scores over time. The univariate analyses showed that age, education, MMSE, and time significantly predicted the Katz ADL score. The multivariate model showed that time significantly predicted an increase in the Katz ADL score. The interaction between age (measured at each assessment) and MMSE score, as well as that between MMSE score and time after being diagnosed, were significant. Higher education predicted an increase in dependency (Table 3). Box 1 provides an overview of the regression formulas that describe the course of MMSE and ADL.

Table 3: Regression parameter estimates (95% CI) of univariate and multivariate mixed effects regression model to predict MMSE (to reflect cognition) and Katz score (to reflect ADL), n=323.

	Cognition (MMSE)		ADL (Katz score)	
Factors	Univariate	Multivariate	Univariate	Multivariate
Intercept	–	26.87 (24.45 to 29.29	–	-0.82 (-2.30 to 0.65)
Time as years after being diagnosed with dementia	-1.84 (-2.10 to -1.57)*	-3.26 (-4.56 to -1.97)**	0.38 (0.29 to 0.46)**	0.26 (0.08 to 0.44)**
Gender (male)	-1.14 (-2.89 to 0.60)	–	0.33 (-0.15 to 0.81)	–
Age at each assessment moment normalized at 75	-0.41 (-0.57 to -0.26)*	-0.35 (-0.53 to -0.16)**	0.15 (0.11 to 0.20)**	0.26 (0.16 to 0.36)**
Education	-0.05 (-0.29 to 0.19)	–	0.08 (0.02 to 0.14)*	0.06 (0.00 to 0.11)**
MMSE at each assessment moment	–	–	-0.16 (-0.18 to -0.14)**	0.02 (-0.05 to 0.08)
Time squared (2nd order polynomial)	–	–	–	–
Interaction time*age	–	0.10 (0.01 to 0.19)*	–	–
Interaction MMSE*age	–	–	–	-0.01 (-0.01 to -0.00)**
Interaction MMSE*time	–	–	–	-0.01 (-0.02 to -0.00)*
Variance random intercept	–	2.00	–	0.71**
Variance random time effect	–	1.86	–	–
Covariance random intercept and random time effect	–	1.73*	–	–

* p<0.05, ** p<0.01

Figure 1: Course of observed MMSE over time. Each line represents an individual; time=0 represents the moment when dementia is diagnosed (using a 1.5 year time correction), time=-6 represents 6 years before the diagnosis of dementia.
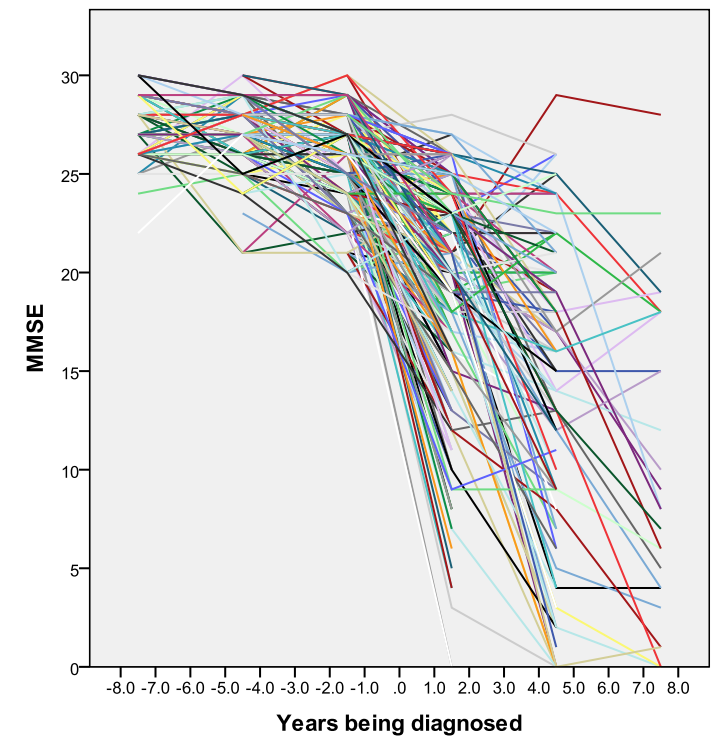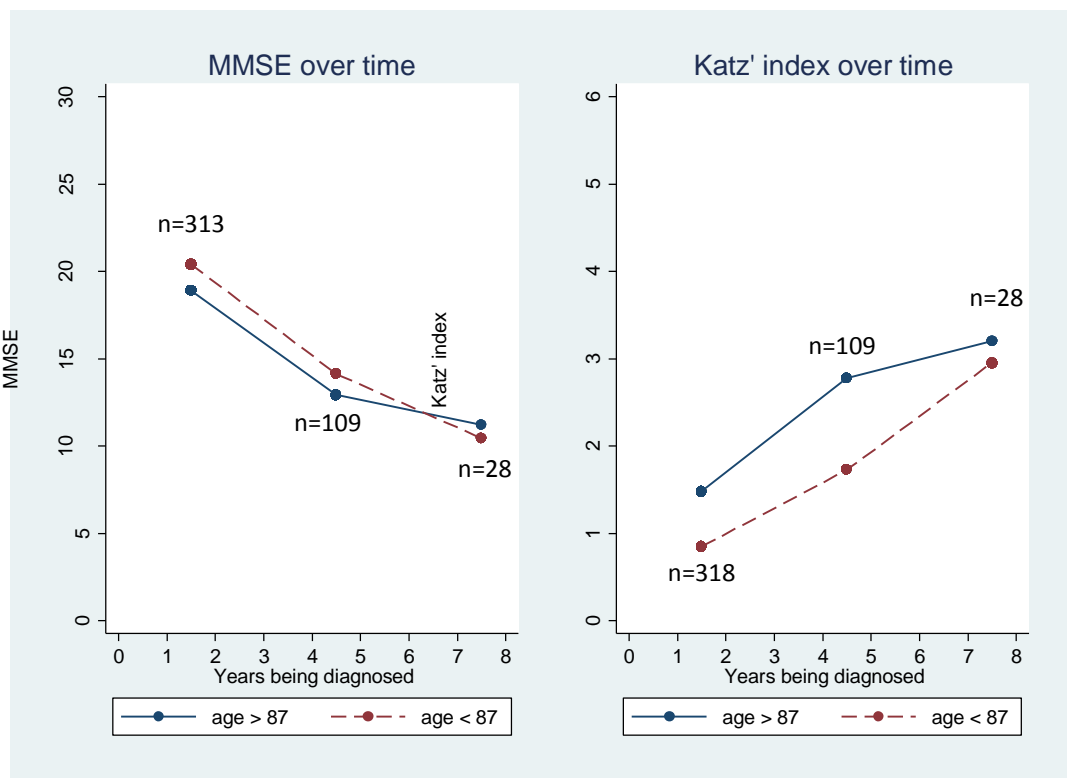

Figure 2: Course of average observed MMSE over time and the average Katz scores observed over time among incident dementia cases (time=0 represents the moment dementia is diagnosed; a time correction of 1.5 years was applied; n=xx represent data points available).

DISCUSSION

In this long-term population-based prospective study, we found that (1) the median AD dementia free survival time was 3.5 years from the onset of MCI; (2) after the onset of dementia, cognition declined at a mean rate of 1.84 MMSE points per year; (3) Katz ADL dependency score increased at a mean rate of 0.38 points per year; and (4) the above results yielded the mathematical expressions presented in box 1 to describe the natural decline of AD in relation to age, gender, and education.

Our estimates (conversion rate: 7.3%, CI: 5.5–9.7%) are within the confidence intervals of a pooled estimate of the conversion of MCI to AD obtained by averaging several population-based studies of MCI (conversion rate 6.8, CI: 1.9–14.5) [18]. Our punctual estimate of 7.3% for MCI conversion to AD is slightly higher that the pooled punctual estimate of 6.8% calculated by Mitchell et al. [18] and this may be due to the fact that our definition of MCI included cases with global cognitive deterioration (OCIND) who progress faster to AD (aMCI conversion rate was 6.7 and OCIND conversion rate was 7.5). The cumulative conversion rate of 31% suggests that some of the MCI subjects improve and some die before developing dementia [19]. Our decision to use parametric survival analysis requires stricter distributional assumptions, though assuming a specific baseline hazard shape allows the survival function to be used to simulate time to event data for health economic modeling.

A population-based study including 95 incident dementia participants [20, 21] found an average rate of cognitive decline of 1.71 MMSE points per 6 months, whereas we found a lower average rate of decline (1.84 / 2 = 0.92 points per 6 months). The difference could be explained by the inclusion of a higher proportion of moderately severe dementia participants in the Kungsholmen Project, who decline less quickly due to the floor effect of the MMSE. According to the multivariate model using average age, subjects decline by 1.2 MMSE points in the first 6 months after being diagnosed.

Mendiondo et al. [22] and Mohs et al. [23] parameterized the annual rate of cognitive decline and found a U-shaped pattern with low decline rates in mild and severe dementia and a higher decline rate in between. We explored this model, but the results were not significant and could be attributed to the use of a population-based sample instead of a clinical sample, as the latter probably includes persons with a poorer prognosis because consulting a medical professional is probably initiated by the person's memory complaints. Han et al. [24] reviewed studies largely based on clinical samples of prevalent cases with an average of 2 years of follow-up, and found a mean annual rate of decline of 3.3 MMSE points per year. Our estimates are at the lower bound of their confidence interval. Besides the use of incident community participants, this difference could be explained by the long follow-up time, in which some participants reach the floor level of the MMSE.

Our model for dependency estimated that more years of education increased dependency. This can be explained by higher educated persons having a cognitive reserve which delays receiving a diagnosis of dementia [25]. By the time the diagnosis is established, the disease is probably more severe than in less educated persons, which could mean that higher educated persons have a poorer prognosis. Gender was not found to be significant in the analyses on the demented subjects, which might be explained by limited statistical power, since the sample included about 17% males. In both the MMSE and ADL models, the effect of age indicates a more rapid decline among younger persons. The interaction effects between time and age in the MMSE model runs

counter to the effect of time. This suggests a decreasing rate of decline in the later stages of the disease. However, it can also be explained by the bottom level of the MMSE that might be less sensitive in severely demented persons.

The strength of our study was the use of a 6-year follow-up period and the prospective study design. Nevertheless, the study was subject to several limitations. The Kungsholmen project included persons aged 75 and older, which resulted in attrition due to death and refusal. However, this reflects reality, since most demented people are older than 75 [26], and the mixed model with random effects and the survival analysis take missing or censored data into account. Nonetheless, generalization to a younger population should be done with caution, although our finding of a positive relation between age and cognition was also found in clinical samples with a younger age [27]. A second limitation is that the Kungsholmen project started in 1987, when the current cholinesterase inhibitors and Memantine treatments that affect cognitive decline were not available. In addition, advances in diagnostics, especially for MCI, might limit the generalization of our findings to the current care standards. Thirdly, the 3-year interval between our measurements may have resulted in a biased estimate of AD dementia free survival in MCI subjects, because a participant may progress to dementia via MCI within an interval without being assessed. Fourthly, the empirical models were not adjusted for comorbidities, as this information was not available to the researchers. Finally, visual comparison between the observed and estimated hazard rates from the survival analysis indicated a difference that can be attributed to the use of 3-year average intervals between assessments. This discrepancy most likely also explains the difference between the observed time until 75% AD free survival (3.2 years) and the estimated survival time derived from the fitted survival model (4.6 years). Furthermore, the 1.5 year correction might limit the precision of the time-to dementia conversion.

The empirical models in Box 1 could be used to simulate the natural disease progression in a cohort and compare this with a scenario where a hypothetical future treatment is available. Such predictions can be integrated with evidence on health care resource usage and quality of life, and enable policy makers to address questions about the potential of new diagnostic or treatment interventions from a cost-effectiveness point of view [28]. Such analyses could provide added value to randomized controlled trials which are limited in terms of follow-up time or the number of scenarios to compare [3]. This should, however, be done with caution, for several reasons. The regression and survival models have not been validated by external datasets, or by predicting the progress of similar patients in current clinical practice. The data available at follow-up was limited, resulting in uncertain predictions. If these results are to be integrated with those from other sources, the populations must be similar. This might represent a difficulty for the evaluation of diagnostic and treatment scenarios, since their evidence is often collected in clinical settings and differs from that derived from a population-based study. This stresses the importance of using sensitivity analysis in decision models to address these issues. Finally, generalizability to other countries is limited because differences in life expectancy might lead to differences in average disease progression rates or the effect of age.

Caution should be used when combining evidence from sources that reflect different populations. Samples recruited within a clinical setting will most likely show more progressive decline, on which age, gender, and education might have different influences. Furthermore, different criteria and subdivisions of MCI have been proposed and modified over time, with different characterizations of cognitive decline (e.g., aMCI specifically reflecting an AD cause and OCIND

including a broader range of potential causes [29]). In addition, extrapolation of the results outside the 6-year time frame should also be done with caution.

In conclusion, our results reflect the natural history of AD in the pre-dementia and dementia phases in terms of cognition and dependency. Since the study was based on community incident cases of MCI and AD dementia, its results can be applied for effectiveness and cost-effectiveness evaluations of interventions in early AD.

ACKNOWLEDGMENTS

REFERENCES

[1] Andrieu S, Coley N, Aisen P, Carrillo MC, DeKosky S, Durga J, Fillit H, Frisoni GB, Froelich L, Gauthier S, Jones R, J¨onsson L, Khachaturian Z, Morris JC, Orgogozo J-M, Ousset P-J, Robert P, Salmon E, Sampaio C, Verhey F, Wilcock G, Vellas B (2009) Methodological issues in primary prevention trials for neurodegenerative dementia. J Alzheimers Dis 16, 235-270.

[2] Cohen JT, Neumann PJ (2008) Decision analytic models for Alzheimer's disease: State of the art and future directions. Alzheimers Dement 4, 212-222.

[3] Schaafsma JD, Van Der Graaf Y, Rinkel GJE, Buskens E (2009) Decision analysis to complete diagnostic research by closing the gap between test characteristics and costeffectiveness. Clin Epidemiol 62, 1248-1252.

[4] Brennan A, Chick SE, Davies R (2006) A taxonomy of model structures for economic evaluation of health technologies. Health Econ 15, 1295-1310.

[5] Green C, Shearer J, Ritchie CW, Zajicek JP (2011) Modelbased economic evaluation in Alzheimer's disease: A review of the methods available to model Alzheimer's disease progression. Value Health 14, 621-630.

[6] Fratiglioni L, Viitanen M, B¨ackman L, Sandman PO, Winblad B (1992) Occurrence of dementia in advanced age: The study design of theKungsholmen Project. Neuroepidemiology 11(Suppl 1), 29-36.

[7] Fratiglioni L, Viitanen M, Von Strauss E, Tontodonati V, Herlitz A, Winblad B (1997) Very old women at highest risk of dementia and Alzheimer's disease: Incidence data from the Kungsholmen Project, Stockholm. Neurology 48, 132-138.

[8] American Psychiatric Association, ed. (1987) Diagnostic and Statistical Manual of Mental Disorders: DSM-III-R, American Psychiatric Association, Washington DC.

[9] Folstein MF, Folstein SE, McHugh PR (1975) Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12, 189-198.

[10] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW (1963) Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. JAMA 185, 914-919.

[11] Fratiglioni L, Grut M, Forsell Y, Viitanen M, Grafstr¨om M, Holmen K, Ericsson K, B¨ackman L, Ahlbom A, Winblad B (1991) Prevalence of Alzheimer's disease and other dementias in an elderly urban population: Relationship with age, sex, and education. Neurology 41, 1886-1892.

[12] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (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 77, 939-944.

[13] Rom´an GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, Amaducci L, Orgogozo JM, Brun A, Hofman A (1993) Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology 43, 250-260.

[14] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: Clinical characterization and outcome. Arch Neurol 56, 303-308.

[15] Petersen RC, Morris JC (2005) Mild cognitive impairment as a clinical entity and treatment target. Arch Neurol 62, 1160- 1163; discussion 1167.

[16] Wahlin A, B¨ackman L, Winblad B (1995) Free recall and recognition of slowly and rapidly presented words in very old age: A community-based study. Exp Aging Res 21, 251-271.

[17] Caracciolo B, Palmer K, Monastero R, Winblad B, Backman L, Fratiglioni L (2008) Occurrence of cognitive impairment and dementia in the community: A 9-year-long prospective study. Neurology 70, 1778-1785.

[18] Mitchell AJ, Shiri-Feshki M (2009) Rate of progression of mild cognitive impairment to dementia–meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand 119, 252-265.

[19] Palmer K, Wang H-X, B¨ackman L, Winblad B, Fratiglioni L (2002) Differential evolution of cognitive impairment in nondemented older persons: Results from the Kungsholmen Project. Am J Psychiatry 159, 436-442.

[20] Ruitenberg A, Kalmijn S, De Ridder MA, Redekop WK, Van Harskamp F, Hofman A, Launer LJ, Breteler MM (2001) Prognosis of Alzheimer's disease: The Rotterdam Study. Neuroepidemiology 20, 188-195.

[21] McDonnell J, Redekop WK,Van Der Roer N, Goes E, Ruitenberg A, Busschbach JJ, Breteler MM, Rutten FF (2001) The cost of treatment of Alzheimer's disease in The Netherlands: A regression-based simulation model. Pharmacoeconomics 19, 543-550.

[22] Mendiondo MS, Ashford JW, Kryscio RJ, Schmitt FA (2000) Modelling mini mental state examination changes in Alzheimer's disease. Stat Med 19, 1607-1616.

[23] Mohs RC, Schmeidler J, AryanM(2000) Longitudinal studies of cognitive, functional and behavioural change in patients with Alzheimer's disease. Stat Med 19, 1401-1409.

[24] Han L, Cole M, Bellavance F, McCusker J, Primeau F (2000) Tracking cognitive decline in Alzheimer's disease using the mini-mental state examination: A meta-analysis. Int Psychogeriatr 12, 231-247.

[25] Qiu C, Backman L,Winblad B, Ag¨uero-Torres H, Fratiglioni L (2001) The influence of education on clinically diagnosed dementia incidence and mortality data from theKungsholmen Project. Arch Neurol 58, 2034-2039.

[26] Lobo A, Launer LJ, Fratiglioni L, Andersen K, Di Carlo A, Breteler MM, Copeland JR, Dartigues JF, Jagger C, Martinez-Lage J, Soininen H, HofmanA(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, S10–S15.

[27] Lopez OL, Schwam E, Cummings J, Gauthier S, Jones R, Wilkinson D, Waldemar G, Zhang R, Schindler R (2010) Predicting cognitive decline in Alzheimer's disease: An integrated analysis. Alzheimers Dement 6, 431-439.

[28] Handels RL, Aalten P, Wolfs CA, Olde Rikkert M, Scheltens P, Visser PJ, Joore MA, Severens JL, Verhey FR (2012) Diagnostic and economic evaluation of new biomarkersfor Alzheimer's disease: The research protocol of a prospective cohort study. BMC Neurol 12, 72.

[29] Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL, De Leon M, Feldman H, Ganguli M, Hampel H, Scheltens P, TierneyMC, Whitehouse P,Winblad B. (2006) Mild cognitive impairment. Lancet 367, 1262-1270.

CHAPTER 7

DETERMINANTS OF CARE COSTS OF PATIENTS WITH COGNITIVE DISORDERS

Handels RL, Wolfs CA, Aalten P, Verhey FR, Severens JL.

Abstract

Introduction: Dementia causes a high burden on patients, caregivers, and societies. Decision analytic models to support allocation of resources are often developed making use of cost-of-illness (COI) studies. However, current COI study estimates are highly variable due to care setting and methodological issues. We aim to explore variables explaining the variation of (formal and informal) health care costs of cognitive disorders, using a broad spectrum of variables, including patient, caregiver, and social context variables.

Methods: A bottom-up COI study design was used in which a societal viewpoint and a validated method to measure and value informal care was applied. Data were analyzed using univariate, multivariate, and forward regression analyses.

Results: The average 1-year health care sector costs were €26,140 ($34,505 or £17,775) and €11,931 ($15,749 or £8113) for patient and family. The analyses indicated that cognitive functioning, caregiver burden, patient sex, and instrumental activities of daily living were significantly associated with care costs independently.

Conclusions: Cognitive functioning and instrumental activities of daily living are important variables to include in health care decision models. We recommend also including caregiver burden and patient sex in decision models for health policy decision makers to fully reflect the heterogeneity of the disease progression of cognitive disorders.

INTRODUCTION

Alzheimer disease (AD) and related dementias affect cognitive function, functional ability, and behavior and could result in a major impact on the quality of a patient's life. Worldwide, 35.6 million persons were estimated to suffer from AD and other dementias in 2010, and this number is predicted to double within the next 20 years. The costs of dementia are estimated at US$604 billion.[1] Because health care resources are scarce, informed decision making on health care management and efficiently allocating resources are important to minimize loss of opportunities. To inform decision makers, several methods to evaluate care interventions are available, among which are trial-based economic evaluations and decision analytic modeling. The first method measures the costs and health care outcomes alongside a clinical trial; the latter provides a framework combining available evidence from different resources. A decision analytic model can be defined as a set of mathematical relationships to form a structure that reflects the natural history of a disease with which the effects of an intervention can be estimated. It enables the calculation of the likelihood of each consequence and its corresponding costs and effects by simulating patients or fractions of a population.[2] In order to build a decision model, information is needed on the relative effect of disease determinants on care costs. Such information can be found in cost-of-illness (COI) studies.[3] This implies that the external validity or the generalizability of decision analytic models, and with that the health care decisions made by policy makers, depends on the quality of COI studies used for decision models.

Several reviews on COI studies have been performed in the field of dementia [4–6] on the basis of 39 COI studies. The total annual costs of care per person suffering from AD and other dementias show a considerable amount of variation,[5] ranging from €6614 to €64,426 in northern and western European countries.[4] To improve COI studies for decision making, it was recommended to use a validated method for the assessment of informal care and to provide an adequate patient sample within each country and care setting.[4,6] We retrieved 1 additional study [7] to the reviews and found that none of the studies identified the independent effect of a broad spectrum of patient disease characteristics and both patient and caregiver social context characteristics, and thus it does insufficient justice to the complexity of the disease to use the results for disease modeling. To reflect a coherent theory of the complex concept of dementia and its treatment, a model used for decision making should include the effects on cognitive function, functional ability, and behavioral problems. This is important in building decision models for care management, as was indicated by 2 recent reviews in the field of decision analytic modeling of AD.[8,9]

Therefore, we aim to explore variables for explaining the variation of health care costs in dementia and cognitively impaired patients, using a broad spectrum of variables, including patient disease and patient and caregiver social context variables. This will provide the relative value of the determinants of care costs for cognitive disorders.

METHODS

Dutch patient-level care cost data were used, from a societal viewpoint (including direct and indirect health care costs and costs made on other sectors), and validated methods were used to evaluate informal care. We analyzed the data of 219 patients and their informal caregivers who participated in the Maastricht Evaluation of a Diagnostic Intervention for Cognitively Impaired

Elderly (MEDICIE) study. General practitioners from 70 practices in the southern part of The Netherlands were asked to refer all 55 years or older patients, suspected of having dementia or a cognitive disorder, not living in a nursing home, and not suffering from an acute disorder from July 2002 to August 2004. The MEDICIE study was published previously.[10,11] In short, the clinical effects and cost effectiveness of an integrated multidisciplinary diagnostic facility for patients with cognitive disorders were examined in a randomized-controlled trial in which the control group received mono-disciplinary usual care. All outcome measures, except for the mini-mental state examination (MMSE)[12] and activities of daily living, were collected through interviews with the patient's proxy, who was also the informal caregiver (ie, proxy's perception of the patient's health was measured). The current study used data of both the intervention and the control group from a 1-year follow-up period. A bottom-up COI design was applied. In such a design, care resource consumption or care costs are measured from a sample of patients and generalized to the total population of the disease under evaluation.

Data Collection

Several baseline outcome measures of the MEDICIE study were used. The MMSE was used to reflect the severity of cognitive function, the Neuropsychiatric Inventory (NPI)[13] to reflect neuropsychiatric symptoms, and the Instrumental Activities of Daily Living scale (IADL)[14] to reflect daily life. Health-related quality of life was measured by the EQ-5D,[15] a validated instrument providing a simple descriptive profile and a single index value for health status. It also includes a visual analogue scale, ranging from 0 (worst imaginable health state) to 100 (best imaginable health state). The experienced caregiver burden was measured by the Self-Perceived Pressure from Informal Care questionnaire (SPPIC),[16] a 9-item scale that measures perceived stress caused by informal caregiving to the demands of the caregiving situation.

Patient characteristics included sex, age, living situation (living together with informal caregiver), diagnosis, and number of comorbidities. The diagnosis was classified as dementia or not. A diagnosis of dementia was specified as AD, vascular dementia, mixed AD and vascular dementia, other dementia, or other cognitive impairment according to regular guidelines. The number of comorbidities of the patient was measured in 2 different ways. First, the number of diseases on the third axis of the Diagnostic and Statistical Manual of Mental Disorders IV [17] diagnosis, as determined by the clinician, was counted and used as an objective measurement of the patient's comorbidities. Second, the informal caregiver was asked to score the presence of several diseases during the interview to obtain a subjective measure of comorbidities. Informal caregiver characteristics included sex, age, experienced burden, number of comorbidities (obtained from caregiver interviews), relation with patient, marital status, education, and net family income.

Because of high correlations and to enhance comparability with other studies, the diagnosis of dementia (yes/no) instead of the differential diagnosis was used for analysis. Likewise, living situation instead of relation between informal caregiver and patient was used. For the same reasons, Global Deterioration Scale scores, Cornell Scale for Depression in Dementia scores, and scores of the 4 items of the 36-item Short Form Health Survey, which were measured in the MEDICIE study, were excluded for this analysis.

Cost Estimates

The costs in the MEDICIE study were determined according to Dutch guidelines [18] from a societal perspective in which all health care costs and health-related patient and family costs were included, regardless of their source of payment. Costs were expressed in euros at 2005 values (at that time, €1.00 was equivalent to US$1.32 and British £0.68).

Volumes of resource use during the 1-year follow-up period after the diagnosis were measured and multiplied by the costs per resource unit. The volumes of resource use were determined by the hospital information system, the electronic patient files of the local community mental health team, the registries of local pharmacies, an informal care survey, and cost diaries.

Informal care activities were assessed using a survey that had been developed for the measurement of informal care [19,20] and valued according to the proxy good method. The average time spent on 16 different informal care tasks was assessed at baseline and at the 6- and the 12-month follow-up and multiplied by the hourly wage rate of a housekeeper (€8.54) or a nurse (€32.67)[18] depending on the nature of the activity.

None of the patients had a paid job, and therefore, work loss in all patients was 0. Work loss of the informal caregiver is included as the time spent caring reimbursed at the wage rate of a professional caregiver.

Annual cost estimates were corrected for intervention costs by subtracting these from the total costs for each patient in the intervention group. Costs are often highly skewed to the right [21] because high care consumption is often concentrated to a small part of the cohort. Therefore, bootstrap credibility intervals were calculated on the basis of 1000 replications; the 2.5 and 97.5 percentiles are presented. More information on the cost analysis can be found elsewhere.[10]

Analyses

To solve the problem of missing data, 10 data sets were created using multiple imputation according to the linear and logistic regression algorithm of SPSS version 17.0. For all analyses, a P-value <0.05 was considered statistically significant unless stated otherwise.

The total costs were skewed to the right (skewness=0.96 and the Shapiro-Wilk test was significant in all imputation sets). A log transformation did not resolve the skewed data. After computing the square root of the costs, skewness (=0.19) was not significant in 9 and the Shapiro-Wilk test in 7 imputation sets. Regression coefficients were back transformed to estimate the effect on actual costs. These estimates apply when the average cost of a variable increases 1 unit of change. Generalization to other situations should be carried out with caution. It is beyond the scope of this analysis to determine the actual value of determinants of care costs. Visual assessment of Q-Q plots and histograms indicated an improvement toward normally distributed data.

Descriptive statistics were applied to all variables. Three different regression models were built using baseline characteristics to determine 1-year follow-up total societal care costs. First, all variables were individually tested using univariate regression analyses and only those variables with a P-value <0.10 were included in a multiple regression model (model 1). Second, a forward regression method (model 2) was performed on each data set using the same significant variables as those selected for the multivariate regression analysis. A variable that proved significant in at least 5 of 10 data sets was considered a priori a relevant cost predictor. Third, because a stepwise

method is solely based on statistics,[22] only variables that were considered to be theoretically relevant for economic decision modeling were included for a multivariate regression analysis (model 3). As described in the introduction, these measures are cognitive function (MMSE), functional ability (IADL), and behavioral problems (NPI). Further, caregiver burden (SPPIC) has been indicated to increase caregiving time and informal care costs [23,24] and was therefore included in this model to assess the effect on total care costs. Expert opinion of co-authors confirmed the face validity of these variables.

RESULTS

Descriptives

Data were missing on the variables spouse of informal caregiver (11%), education of the informal caregiver (11%), patient's objective comorbidities (5%), and patient's and caregiver's subjective comorbidities (both 12%). Patients with incomplete data did not differ significantly from patients with complete data.

Table 1 presents the sample characteristics at baseline. The average patient age was 78 years, ranging from 55 to 94 years, and the mean MMSE score was 20.0 (SD=5.8). Most patients suffered from dementia (72%). The informal caregiver was mainly the son or the daughter (53%) and spent on average 8.5 h/wk providing care to the patient.

Table 1: Sample Characteristics at Baseline of 219 Patients and Informal Caregivers

Characteristic	Frequency and percentage		Average and standard deviation	
Female	143	(65%)		
Age *(years)*			78	(6.7)
Diagnosis of dementia	158	(72%)		
EQ-5D VAS score *(0-10)**			5.8	(1.9)
EQ-5D utility score *(0-10)*†			5.2	(3.2)
MMSE score *(0-30)*			20.0	(5.8)
NPI score (0-144)			23.9	(16.6)
IADL score (0-14)			3.8	(2.1)
Number of co-morbidities of the patient, objectively measured			3.1	(1.9)
Number of co-morbidities of the patient, subjectively measured			3.6	(2.4)
Patient and informal caregiver live together	88	(40%)		
Female informal caregiver	145	(66%)		
Age informal caregiver *(years)*			59.5	(14.1)
Informal caregiver is child (in-law)	117	(53%)		
SPPIC score *(0-9)*			4.6	(2.8)
Number of co-morbidities informal caregiver, subjectively measured			2.0	(2.2)
Informal caregiver has spouse	170	(77%)		
High education informal caregiver‡	115	(53%)		
High net family income§	97	(57%)		

*The VAS-score of 0-100 was recalculated to 0-10 to increase comparability
†The utility-score of 0-1 was recalculated to 0-10 to increase comparability
‡Higher General Secondary Education or higher education
§More than €1589 per month; 49 participants who did not state income were excluded from descriptive analysis.
Abbreviations: IADL, Instrumental Activities of Daily Living scale; MMSE, mini-mental state examination, NPI, Neuropsychiatric Inventory; SPPIC, Self-Perceived Pressure from Informal Care questionnaire; VAS, visual analogue scale;

Table 2 provides a detailed cost overview on the basis of the data of the MEDICIE study.[11] The 1-year sample average health care and health-related patient family costs were €38,071 (or $50,254 or £25,888) (95% bootstrap credibility interval=€36,893–€39,435). The major cost components were admission to a nursing home (€6143 or $8109 or £4177) and informal care (€6620 or $8738 or £4502), contributing toward 34% of the total costs. The interval reflects the 2.5th and 97.5th percentile confidence intervals.

Table 2: Health Care and Patient and Family Costs of the Study Sample in Euros

Cost item*	No. (%) of patients†	Mean visits‡	Mean costs§	Credibility interval¶
Health care sector costs				
Hospital (A)	166(76)	NA	€2,875	2,654-3,114
CMHT (B)	148 (68)	8.4	€1,073	1,014-1,135
Medication (C, E)	215 (98)	NA	€1,505	1,444-1,566
Admission (D)				
Nursing home	33 (15)	28.8	€6,143	5,469-6,885
Elderly home	34 (16)	38.9	€3,475	3,056-3,859
Home care (D)				
Domestic help	122 (56)	91.1	€2,038	1,924-2,151
Nursing	112 (51)	98.9	€4,154	3,790-4,547
Day care (D)				
In nursing home	34 (16)	11.6	€1,442	1,268-1,633
In elderly home	64(29)	28.2	€1,448	1,322-1,598
In community centre	10(5)	3.8	€11	9-14
General practitioner (E)	195 (89)	23.9	€512	488-533
Other professionals (E)	150 (68)	52.4	€1,462	1,392-1,530
Patient and family costs				
Informal care (D)	219 (100)	839.7	€6,620	6,151-7,015
Housekeeper (D)	149 (68)	87.3	€657	62-691
Durable goods (E)	100 (46)	NA	€486	454-518
Consumable goods (E)	151 (69)	NA	€160	154-168
Out-of-pocket patient (E)	155 (71)	NA	€1,099	1,057-1,142
Out-of-pocket caregiver (D)	196 (89)	NA	€2,262	2,201-2,325
Travelling costs#				
to GP and other professionals	NA	NA	€36	34-37
to Hospital (A)	NA	NA	€10	9-10
to Day care (D)	NA	NA	€592	551-635
Parking costs (A)	NA	NA	€10	9-11
Summed costs				
Mean total health care sector costs	-	-	€26,140	25,510-26,796
	-	-	$34,505	33,673-35,371
	-	-	£17,775	17,347-18,221
Mean total patient and family costs	-	-	€11,931	11,571-12,368
	-	-	$15,749	15,274-16,326
	-	-	£8,113	7,868-8,410
Mean total societal care costs	-	-	€38,071	36,893-39,435
	-	-	$50,254	48,699-52,054
	-	-	£25,888	25,087-26,816

*Source of volume information: A, hospital information system; B, electronic patient file, CMHT; C, registries of local pharmacies; D, informal care questionnaire; and E, cost diaries. See the "Methods" section for more information.
† Number of patients and the percentage of the total group that made use of the specific item
‡Mean visits per patient
§Mean costs per patient, €
¶2.5th and 97.5th percentile credibility interval of 1000 bootstrap replications.
Abbreviations: CMHT, community mental health team; GP, general practitioner; NA, not applicable.

Regression Analyses

In both the first and the third regression analysis, multi collinearity was evaluated and was not present. Visual assessment of standardized residual plots and normal probability plots did not imply a violation of the assumption of homoskedasticity or the assumption of normally distributed residuals.

The results of the individual univariate regression analyses are shown in Table 3. This table presents the pooled results of the 10 data sets. A positive direction of the [beta]-value indicates an increase in cost. The variables age and the objective number of comorbidities of the patient and age, subjective number of comorbidities, spouse, education, and net family income of the informal caregiver were not significant (P>0.10) and were therefore excluded from further analyses.

Table 3: Pooled Results of the Individual Univariate Regression of 1-year Square Root Total Care Costs

Variable	Beta	P-value
Female *(1=yes, 0=no)*	26.13	.01
Age *(years)*	1.31	.06
Diagnosis *(1=dementia, 0=no dementia)*	30.22	.00
EQ-5D VAS score *(0-10)**	-9.95	.00
EQ-5D utility score *(0-10)*†	-8.08	.00
MMSE score *(0-30)*	-3.79	.00
NPI score *(0-144)*	1.51	.00
IADL score *(0-14)*	-11.23	.00
Number of co-morbidities of the patient, objectively measured	2.80	.29
Number of co-morbidities of the patient, subjectively measured	4.34	.04
Patient and informal caregiver live together *(1=yes, 0=no)*	-24.12	.02
Female informal caregiver *(1=yes, 0=no)*	10.33	.30
Age informal caregiver *(years)*	-0.12	.73
SPPIC score *(0-9)*	9.76	.00
Number of co-morbidities of informal caregiver, subjectively measured	2.49	.29
Informal caregiver has spouse *(1=yes, 0=no)*	-10.27	.43
High education informal caregiver‡ *(1=yes, 0=no)*	9.24	.33
High net family income§ *(1=yes, 0=no)*	7.03	.62

Abbreviations: IADL, Instrumental Activities of Daily Living scale; MMSE, mini-mental state examination, NPI, Neuropsychiatric Inventory; SPPIC, Self-Perceived Pressure from Informal Care questionnaire; VAS, visual analogue scale
*The VAS-score of 0-100 was recalculated to 0-10 to increase comparability
†The utility-score of 0-1 was recalculated to 0-10 to increase comparability
‡Higher General Secondary Education or higher education
§More than €1589 per month

The results of the multivariate regression (model 1) on the basis of the significance of individual univariate regression analyses are presented in Table 4. Patient sex, cognition (MMSE), functioning (IADL), and SPPIC determined the total care costs. Again, a positive direction of the [beta] indicates an increase in cost. This model explained 34% of the variation in the total costs (R^2=0.34).

Table 4: Pooled Results of the Multivariate Regression Model and Forward Regression Results of 10 Imputation Sets of 1-year Square Root Total Care Costs

Variable	Multivariate regression				Forward regression †
	Beta	P-value	Back trans-formed β*	Confidence interval back transformed beta*	
Female *(1=yes, 0=no)*	25.9	.02	10,120	1,936 to 18,304	7
Age	-0.2	.73	-82	-533 to 369	0
Diagnosis *(1=dementia, 0=no dementia)*	-5.0	.64	-1,816	-8,619 to 4,988	0
EQ-5D VAS score *(0-100)*	0.2	.55	60	-134 to 254	0
EQ-5D utility score *(0-1)*	-21.5	.24	-7,396	-19,072 to 4,279	7
MMSE score *(0-30)*	-1.9	.03	-701	-1,276 to -126	9
NPI score *(0-144)*	0.5	.12	173	-34 to 380	4
IADL score *(0-14)*	-6.7	.01	-2,396	-4,170 to -621	7
Number of co-morbidities of the patient, subjectively measured	0.6	.76	233	-1,061 to 1,527	0
Patient and informal caregiver live together *(1=yes, 0=no)*	-13.7	.16	-4,811	-11,015 to 1,393	6
SPPIC score *(0-9)*	6.3	.00	2,357	1,146 to 3,568	10
constant	215.1	.00			

*Back-transformed costs were estimated by calculating the effect on costs if the average variable value increased by 1 using the following equations for back transformation; equation 1 for the cost estimate, equation 2 for the variance to calculate the confidence interval.

†Number of models in which variable is significant.

$$E = \left(\sum_i \left(C + \beta_t x_t + \beta_i x_i \right) \right)^2 - \left(\sum_i \left(C + \beta_t x_{(t+1)} + \beta_i x_i \right) \right)^2 \qquad (1)$$

$$\mathrm{var}_E = 4(C + \beta)^2 \, \mathrm{var}_\beta + 4\beta^2 \, \mathrm{var}_C + 4\beta(C + \beta)\mathrm{cov}(C, \beta) \qquad (2)$$

E, cost estimate of back-transformed beta; *C*, constant; *β_t*, beta of the variable to back transform; *β_i*, beta of variable *i* in the regression formula; *x_t*, average value of the variable to back transform; *x_(t+1)*, average value of the variable to back transform increased with 1; *x_i*, average values of variable *i*; *var_E*, variance of a back-transformed estimate; *var_C*, variance of the constant; *var_β*, variance of the beta; *cov(C, β)*, covariance between constant and beta.

To properly calculate the estimated total costs for a certain set of values for the covariates, a smearing factor must be applied[25]. However, in case of square root transformations, this smearing factor is an additive constant that cancels out in the subsequent subtraction.

IADL indicates Instrumental Activities of Daily Living scale; MMSE, mini-mental state examination, NPI, Neuropsychiatric Inventory; SPPIC, Self-Perceived Pressure from Informal Care questionnaire; VAS, visual analogue scale.

The results of the forward regression (model 2) are shown in the last column of Table 4. For each variable, it is shown in how many of the 10 data sets this variable is significant in the forward regression analyses (eg, MMSE is significant in 9 of 10 forward analyses). Sex of the patient, the EQ-5D utility score, MMSE, IADL, living situation, and SPPIC significantly contributed toward the square root total care costs of patients in at least 5 imputation sets and were therefore considered to be relevant cost predictors.

The multivariate regression (model 3) is presented in Table 5. All variables except IADL significantly predicted costs. This model explained 28% of the variation in square root total care costs (R^2=0.28).

Table 5: Pooled Results of the Multivariate Regression Model Including Theoretical Variables Only to Explain 1-year Square Root Total Care Costs

	Beta	P-value	Back transformed β*	Confidence interval back transformed β*
MMSE score *(0-30)*	-2.53	.001	-918	-1,437 to -399
NPI score *(0-144)*	0.69	.016	252	52 to 452
IADL score *(0-14)*	-4.31	.057	-1,555	-3,049 to -61
SPPIC score *(0-9)*	6.32	.000	2,349	1,155 to 3,543
constant	203.94	.000		

*See the footnote of Table 4 for calculation details.
IADL indicates Instrumental Activities of Daily Living scale; MMSE, mini-mental state examination, NPI, Neuropsychiatric Inventory; SPPIC, Self-Perceived Pressure from Informal Care questionnaire.

The overall results of all 3 analyses indicated that MMSE and SPPIC were significant predictors of the 1-year square root total care costs. Patient sex and IADL were important variables in both the multivariate and the forward analyses for explaining the 1-year square root total care costs.

DISCUSSION

Using the 12-month follow-up data of 219 patients with cognitive disorders, 4 variables were the most relevant predictors of annual care costs of patients suffering from dementia or a cognitive impairment: severity of cognitive function, experienced caregiver burden, patient sex, and functional ability.

These results are robust as 3 different analytic methods and a broad spectrum of disease severity variables and patient and informal caregiver social context variables were used. In addition, although the proxy good method does not include preference for providing informal care, it is a validated method to assess informal care costs. Therefore, this study adds value to the current COI studies and can thus be used to improve decision analytic modeling of AD and related dementias.

We identified patient sex, MMSE, IADL, and caregiver burden as the key variables to explain the variability of the care costs. This is in line with a recent review that identified caregiver burden as an important predictor for institutionalization.[26] In this study sample, this is reflected by the positive correlation with home-based nursing care and care in a nursing home. Cost differences by patient sex might be explained by differences in caregiving experience and approaches to caregiving that were likely to result in women experiencing more depressive symptoms and anxiety, more time spent caregiving, higher burden, and less support from other family members or friends.[27] The strong relation between both MMSE and IADL and care costs from our analyses is in line with the literature.[3,28,29]

In contrast to the literature, behavioral problems and care costs [30,31] were not significantly related in the multivariate and forward analyses. A plausible explanation is the significant correlation with the IADL (Pearson correlation=-0.39) that eliminates part of the variance explained by the NPI. Another explanation is that the NPI may not be sensitive to subtle changes. Additional analyses by replacing the total NPI with each individual item of the NPI only indicated that "irritability" was the only variable that predicted costs without any change in significance in the other variables. In addition, caregiver distress due to patient behavior, but not behavior in itself, is a risk factor for patient institutionalization.[32] The number of comorbidities was not

110

significant. Although the literature might suggest that this item is a cost predictor, it is not conclusive on this item.[28,33] A possible explanation could be that part of the variance is explained by the experienced burden by the informal caregiver.

These differences from other studies may also be explained by sample differences with the complex patient group of this study that was in need of care on multiple domains.

This study is in line with the findings of previous research that informal care represents a considerable part of the care costs in dementia,[28,34,35] which emphasizes the importance of applying a societal viewpoint to include all relevant care costs for patients with dementia. This could, for example, imply that if direct health care resources are diminished, these probably will be compensated by informal care.

Limitations

All outcome measures except the MMSE and activities of daily living were collected through personal interviews with the patient's proxy. It is unknown whether the proxy's response exactly represents the situation of the patient, because it is not known as to which measurement method (ie, patient, caregiver, or physician) is the most representative.[36]

The study sample represents the 1-year care costs of patients after they received a diagnosis. This may limit long-term extrapolation of the cost data. Three possible causes for selection bias are the willingness of participants to participate in research, the general practitioner's inability to recognize all dementia patients in the general population, and the applied inclusion criteria of the MEDICIE study, which was not originally designed for a COI study analysis. These factors limit the generalizability, because institutionalized patients were excluded and costs were measured from the point of diagnosis under representing the cognitive impairment stage preceding dementia. Determinants of care costs may be different in the mild cognitive impairment stage, early dementia stage, and late dementia stage. Therefore, generalization of the results should be done with caution. This is a common problem of the bottom-up design in which not all typical disease events could have occurred in the sample population and the follow-up period of the study. Furthermore, clinical trial data of both the intervention and the control group were used.[10] Although it was controlled for additional costs and the total mean difference between both groups was €23, it might have limited the generalizability to the Dutch population. Finally, several factors, including unit costs, resource use, clinical practice, and patient case mix, were expected to generate some variation in the total cost estimates between countries.[37]

Implications for Modeling

The findings of this study build on the evidence to recommend basing a decision model structure on cognitive function, functional ability, and behavior to model disease progression.[8,9] Moreover, we identified sex and the burden of the informal caregiver as key variables for predicting square root annual total care costs for newly diagnosed patients adjusted for the 3 variables of disease severity. Taking into account the influence of sex on costs in a decision analytic model structure can provide a more precise prognosis of care costs and enables the estimation of demographical differences regarding a population. Including caregiver burden fully reflects the heterogeneity of disease progression. Most AD models describe the economic effects

of a medical intervention,[8] although they do not explicitly include caregiver burden.[38–40] Nonetheless, several models included institutionalization and therefore might have comprised caregiver burden indirectly in the probability for institutionalization.[41] For the evaluation of nonpharmacotherapeutic interventions, such as a diagnostic technique or caregiver support that could have an isolated effect on caregiver burden, it is important to include caregiver burden as this research indicates. The application of all recommended variables in a decision model could have implications for the choice of model type, for example, because of transparency issues. A recommendation for model type is, however, beyond the scope of this article.

CONCLUSIONS

Cognitive function, functional ability, caregiver burden, and patient sex are independently the most relevant explanatory variables of care costs of cognitive disorders. The findings stress the need for multicomponent decision models and correspond to current recommendations for decision analytic modeling to include cognitive function, functional ability, and behavior. In particular, this study adds the recommendation to include patient sex and caregiver burden to fully reflect the heterogeneity of disease progression. The results of this study may improve decision models and help policy makers to allocate scarce health care resources more efficiently.

ACKNOWLEDGMENT

REFERENCES

1. Wimo A, Prince M World Alzheimer Report 2010 The Global Economic Impact of Dementia. 2010. Available at: http://www.alz.co.uk/research/worldreport/. Accessed October 1, 2010
2. Briggs AH, Claxton K, Sculpher MJ Decision Modelling for Health Economic Evaluation. 2006 Oxford Oxford University Press
3. Tarricone R. Cost-of-illness analysis. What room in health economics? Health Policy. 2006;77:51–63
4. Jonsson L, Wimo A. The cost of dementia in Europe: a review of the evidence, and methodological considerations. Pharmacoeconomics. 2009;27:391–403
5. Quentin W, Riedel-Heller SG, Luppa M, et al. Cost-of-illness studies of dementia: a systematic review focusing on stage dependency of costs. Acta Psychiatr Scand. 2009;121:243–259

6. Wimo A, Jonsson L, Gustavsson A, et al. The economic impact of dementia in Europe in 2008-cost estimates from the Eurocode project. Int J Geriatr Psychiatry. 2011;26:825–832

7. Coduras A, Rabasa I, Frank A, et al. Prospective one-year cost-of-illness study in a cohort of patients with dementia of Alzheimer's disease type in Spain: The ECO Study. J Alzheimers Dis. 2009;19:601–615

8. Cohen JT, Neumann PJ. Decision analytic models for Alzheimer's disease: state of the art and future directions. Alzheimers Dement. 2008;4:212–222

9. Green C. Modelling disease progression in Alzheimer's disease: a review of modelling methods used for cost-effectiveness analysis. Pharmacoeconomics. 2007;25:735–750

10. Wolfs CA, Dirksen CD, Kessels A, et al. Economic evaluation of an integrated diagnostic approach for psychogeriatric patients: results of a randomized controlled trial. Arch Gen Psychiatry. 2009;66:313–323

11. Wolfs CA, Kessels A, Dirksen CD, et al. Integrated multidisciplinary diagnostic approach for dementia care: randomised controlled trial. Br J Psychiatry. 2008;192:300–305

12. 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

13. Cummings JL, McPherson S. Neuropsychiatric assessment of Alzheimer's disease and related dementias. Aging (Milano). 2001;13:240–246

14. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist. 1969;9:179–186

15. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy. 1990;16:199–208

16. Pot AM, van Dyck R, Deeg DJ. Perceived stress caused by informal caregiving. Construction of a scale. Tijdschr Gerontol Geriatr. 1995;26:214–219

17. Diagnostic Criteria From DSM-IV-TR. 2000 Washington, DC American Psychiatric Association

18. Oostenbrink JBKM, Rutten FFH Manual for costing: methods and standard costs for economic evaluations in health care. 2004. [in Ducth] Amstelveen, The Netherlands College voor zorgverzekeringen

19. van den Berg B, Al M, Brouwer W, et al. Economic valuation of informal care: the conjoint measurement method applied to informal caregiving. Soc Sci Med. 2005;61:1342–1355

20. van den Berg B, Brouwer W, van Exel J, et al. Economic valuation of informal care: lessons from the application of the opportunity costs and proxy good methods. Soc Sci Med. 2006;62:835–845

21. Briggs AH, Claxton K, Sculpher MJ Decision Modelling for Health Economic Evaluation. 2006 Oxford Oxford University Press

22. Tabachnick BG, Fidell LS Using Multivariate Statistics. 20014th ed Boston, MA Allyn and Bacon

23. Wimo A, von Strauss E, Nordberg G, et al. Time spent on informal and formal care giving for persons with dementia in Sweden. Health Policy. 2002;61:255–268

24. Harrow BS, Mahoney DF, Mendelsohn AB, et al. Variation in cost of informal caregiving and formal-service use for people with Alzheimer's disease. Am J Alzheimers Dis Other Demen. 2004;19:299–308

25. Jones M Models for Health Care. University of York. Available at: http://www.york.ac.uk/res/herc/documents/wp/10_01.pdf. Accessed November 15, 2011

26. Gaugler JE, Yu F, Krichbaum K, et al. Predictors of nursing home admission for persons with dementia. Med Care. 2009;47:191–198

27. Connell CM, Janevic MR, Gallant MP. The costs of caring: impact of dementia on family caregivers. J Geriatr Psychiatry Neurol. 2001;14:179–187

28. Jonsson L, Eriksdotter Jonhagen M, Kilander L, et al. Determinants of costs of care for patients with Alzheimer's disease. Int J Geriatr Psychiatry. 2006;21:449–459

29. Wolstenholme J, Fenn P, Gray A, et al. Estimating the relationship between disease progression and cost of care in dementia. Br J Psychiatry. 2002;181:36–42

30. Herrmann N, Lanctot KL, Sambrook R, et al. The contribution of neuropsychiatric symptoms to the cost of dementia care. Int J Geriatr Psychiatry. 2006;21:972–976

31. Murman DL, Chen Q, Powell MC, et al. The incremental direct costs associated with behavioral symptoms in AD. Neurology. 2002;59:1721–1729

32. de Vugt ME, Stevens F, Aalten P, et al. A prospective study of the effects of behavioral symptoms on the institutionalization of patients with dementia. Int Psychogeriatr. 2005;17:577–589

33. Scuvee-Moreau J, Kurz X, Dresse A. The economic impact of dementia in Belgium: results of the National Dementia Economic Study (NADES). Acta Neurol Belg. 2002;102:104–113

34. Souetre E, Thwaites RM, Yeardley HL. Economic impact of Alzheimer's disease in the United Kingdom. Cost of care and disease severity for non-institutionalised patients with Alzheimer's disease. Br J Psychiatry. 1999;174:51–55

35. Wimo A, Winblad B, Jonsson L. The worldwide societal costs of dementia: estimates for 2009. Alzheimers Dement. 2010;6:98–103

36. Tamim H, McCusker J, Dendukuri N. Proxy reporting of quality of life using the EQ-5D. Med Care. 2002;40:1186–1195

37. Sculpher MJ, Pang FS, Manca A, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technol Assess. 2004;8:iii–iv 1–192

38. Caro JJ, Getsios D, Migliaccio-Walle K, et al. Assessment of health economics in Alzheimer's disease (AHEAD) based on need for full-time care. Neurology. 2001;57:964–971

39. Francois C, Sintonen H, Sulkava R, et al. Cost effectiveness of memantine in moderately severe to severe Alzheimer's disease: a Markov model in Finland. Clin Drug Investig. 2004;24:373–384

40. Ward A, Caro JJ, Getsios D, et al. Assessment of health economics in Alzheimer's disease (AHEAD): treatment with galantamine in the UK. Int J Geriatr Psychiatry. 2003;18:740–747

41. Torti FM Jr, Gwyther LP, Reed SD, et al. A multinational review of recent trends and reports in dementia caregiver burden. Alzheimer Dis Assoc Disord. 2004;18:99–109

CHAPTER 8

EARLY COST-UTILITY ANALYSIS OF GENERAL AND CEREBROSPINAL FLUID-SPECIFIC ALZHEIMER'S DISEASE BIOMARKERS FOR HYPOTHETICAL DISEASE-MODIFYING TREATMENT DECISION IN MILD COGNITIVE IMPAIRMENT

Ron L.H. Handels, Manuela A. Joore, An Tran-Duy, Claire A. G. Wolfs, Frans R. J. Verhey, Johan L. Severens

Abstract

Introduction: Various strategies are possible to apply Alzheimer's disease (AD) biomarkers for revealing the cause of mild cognitive impairment (MCI) and support decision making when a hypothetical disease-modifying treatment (DMT) is available. The study aim was to determine 1) the headroom of a perfect cerebrospinal fluid (CSF) biomarker and 2) the societal incremental net monetary benefit (NMB) of a CSF biomarker either in a strategy to verify an AD diagnosis or a strategy to verify a non-AD diagnosis as set by the current clinical practice diagnostic workup. All analyses were performed in MCI subjects assuming that a hypothetical DMT (50% progression reduction at € 5,853/year assumed) for AD was available.

Methods: A probabilistic patient-level model combining a decision tree for the diagnostic phase and a time to event model for the treatment phase was developed to model MCI subjects from a memory clinic. It was used to evaluate the lifetime potential benefit of a perfect biomarker and to compare current practice to two strategies using a CSF biomarker, positioned either as an add-on test when the current practice diagnostic workup concluded the presence of AD or as an add-on test in case the current practice diagnostic workup concluded the absence of AD. Since the diagnostic accuracy of a CSF biomarker to predict DMT response is unknown, sensitivity and specificity were ranged over all theoretically possible values.

Results: The simulated MCI population had an average age (SD) of 68.3 (8.9), of which 36% was female and 49% suffered from AD pathology. The headroom analysis revealed 0.39 (0.26 to 0.54) QALYs gained and € 33,622 (21,232 to 50,780) savings per subject if a perfect test was simulated in all subjects. The CSF biomarker was cost-effective (at a willingness to pay of € 80,000/QALY) in more possible combinations of sensitivity and specificity when it was positioned after clinical practice concluded the absence of AD rather than after the presence of AD. Different optimal positions of the CSF biomarker were found in the sensitivity analyses.

Discussion: The results indicated more potential benefit from a biomarker positioned to verify subjects who are not expected to have AD (i.e. a treatment selection approach to prevent undertreatment) rather than to verify subjects expected to have AD (a treatment selection approach to prevent overtreatment) given the assumption on the availability of a hypothetical DMT.

INTRODUCTION

With a global prevalence of 35.6 million and a corresponding economic impact of US$604 billion dementia has a substantial burden on societies worldwide [1,2]. There has been a growing interest in identifying Alzheimer's disease (AD) pathology in patients with Mild Cognitive Impairment (MCI), a pre-dementia stage in which a cognitive impairment is objectively measured but activities of daily living are intact. Revealing the cause of MCI is of scientific value for the development of drugs that prevent such patients from conversion to dementia.

The proposed role of biomarkers in cerebrospinal fluid (CSF), Positron Emission Tomography and Magnetic Resonance Imaging has been described in the recently revised diagnostic criteria for MCI due to AD [3]. Despite that these biomarkers are only meant for research purposes, CSF is finding its way in clinical practice [4]. The decision to adopt CSF or any other biomarker in the clinical routine should however depend on the improvement of a patient's health and, in a resource-constrained health care system, on cost-utility. In other words, society must be willing to spend the required resources to perform biomarkers for their gain in quality-adjusted life years (QALY).

In general, a test itself does not improve patient-important outcomes and/or health related quality of life. Therefore a test can only be evaluated in combination with its downstream consequences of treatment [5]. The latter could be a future (pharmacological) intervention, initiated based on a test's result, that influences health as was earlier shown by Sköldunger et al. [6]. Several economic evaluations that evaluated the added value of AD biomarkers in clinical practice have been reviewed [7], though none of them focussed on an early diagnosis in the MCI phase. Disease-modifying drug compounds have not reached effectiveness in phase 4 clinical trials in the MCI phase yet [8] which makes it difficult to estimate the impact of a CSF biomarker in the current clinical practice. The uprising application of AD biomarkers in clinical practice urges the need for an early technology assessment that sketches future scenarios of CSF use in combination with treatment management to support policy making of AD biomarkers [9].

By identifying the headroom, i.e. the benefits when the current practice diagnostic accuracy is maximally improved by a perfect biomarker test, the potential of biomarkers can be revealed. However, such perfect test does not exist and estimating the actual accuracy of CSF for predicting the treatment response of a future hypothetical disease-modifying treatment (DMT) is practically impossible, even by expert opinion. A conventional economic evaluation only compares a few alternative strategies. This would be insufficient to evaluate all the strategies reflected by the many future possible values of diagnostic accuracy that CSF could have. The position of AD biomarkers in the clinical routine will nonetheless likely be either to confirm AD in patients who are suspected of AD (which will cause increased sensitivity at the cost of specificity) or to rule-out AD (which will cause increased specificity at the cost of sensitivity, see figure 1). Exploring the consequences of these 2 strategies over its full range of possible values of test sensitivity and specificity in an early technology assessment can aid biomarker research. Using a range of input values and present each corresponding result differs from conventional evaluations of only a few alternative options, though it is assumed to generate more useful results.

The aim of this early health technology assessment was to determine, from a societal perspective and compared to current clinical practice, 1) the headroom of a perfect CSF biomarker and 2) the incremental net monetary benefit (NMB) of a CSF biomarker either in a strategy to verify an AD

diagnosis or to verify a no-AD diagnosis as set by the current clinical practice diagnostic workup, in MCI subjects under the condition that a hypothetical DMT for AD is available.

Figure 1: Change in diagnostic accuracy. If a biomarker is performed to verify subjects who are expected of Alzheimer's disease (AD) after clinical practice diagnostic workup (category A and B) then some of these diagnoses are reclassified (blue arrows) to 'no AD' (category C and D). This can lead to an increase of specificity and/or decrease of sensitivity of the diagnosis after the biomarker test.

	Actual disease state: AD	Actual disease state: Non AD
Current practice test positive	A	B
Current practice test negative	C	D
	Sensitivity = A / (A+C)	Specificity = D / (B+D)

METHODS

Design

A probabilistic patient-level model was used to synthesize available evidence on various disease components and simulate the difference in lifetime consequences of a group of individuals with MCI [10,11].

Evidence was mainly derived from the LeARN study that included patients suspected of having a primary neurodegenerative disease [12], the Kungsholmen project which is a general population-based cohort from which incident MCI and incident dementia cases were filtered [13], the MEDICIE study on quality of life and resource utilization on a multidisciplinary diagnostic and management approach in psychogeriatric patients [14], and literature and expert opinion.

The headroom of a new technological intervention is the benefit when the most optimistic plausible situation would be realized [15]. For the headroom analysis (the first study aim) the current practice diagnostic workup was compared to a biomarker strategy that was assumed to be 100% sensitive and 100% specific in determining AD as the cause of MCI. For the incremental NMB analysis (the second study aim) the CSF biomarker was compared to current practice. It was positioned in two alternative ways as an add-on test to the current clinical practice diagnostic workup that consisted of a physical, clinical and neuropsychological examination, patient and informal caregiver history, and MRI, in MCI subjects who visited a memory clinic. First, the CSF test was performed only if the current practice workup concluded on the presence of AD (referred to as the 'verify AD' strategy; see figure 2); second, the CSF test was performed only if the current practice workup concluded on the absence of AD (referred to as the 'verify non-AD' strategy; not presented in figure 2). The cause of MCI was evaluated in both the current practice strategy and the CSF intervention strategy and, if AD was concluded, a hypothetical DMT was provided. This was modelled as a one-time only treatment decision at incident MCI. The subject's lifetime costs and QALYs were compared between the current practice and each of the CSF strategies. Current available treatments (Cholinesterhase inhibitors and Memantine) were not modelled because they are not intended for subjects suffering from MCI [16].

Annual discount rates for costs and effects were set at 4% and 1.5% respectively, according to the Dutch guideline for pharmaco-economic research [17].

Model structure

An individual subject simulation model was considered the most flexible to describe the diagnostic workup and the gradual disease progression in cognition and functioning. Therefore, a combination of a decision tree, to reflect the diagnostic workup, and a time to event model, to reflect disease progression, were combined in one model. The model started with a population of 2,000 incident MCI subjects from a memory clinic setting. Each subject's specific baseline attributes (age, gender, education, pathology (AD, other neurodegenerative disease (NDD) or no NDD), Mini-Mental State Examination [18] (MMSE) score and Katz score [19]) were sampled from a normal distribution with mean and standard deviation estimated using an existing cohort of incident MCI memory clinic patients (see the LeARN study's MCI estimates in table 1). Each subject was quadruplicated; one went through the current practice strategy and the other 3 went through the intervention strategies ('perfect test' for the headroom analysis, and the 'verify AD' and 'verify non-AD' for the incremental NMB analysis). If a strategy's diagnostic workup concluded AD as the cause of MCI a hypothetical DMT was applied. Next, the time to conversion from MCI to dementia was estimated for each individual subject based on its attributes (gender, pathology and treatment effect) using a survival function (see Appendix). At the same moment the time to death was estimated. The event with the shortest time was let occur for that subject.

Disease progression in the dementia phase was modelled by the annual change in cognition (MMSE) and activity of daily living (ADL) (Katz score) on a continuous scale based on each subject's attributes including treatment status using regression formulas (see Appendix). Eventually the model stopped when the subject died or had been 30 years in the dementia phase. Figure 2 represents the general model structure for the incremental NMB analysis for the control and the 'verify AD' strategy. The model was built in Microsoft Excel 2010. A detailed description of the model can be found in the Appendix. Table 1 presents patients' baseline characteristics of the data sources. Table 2 provides an overview of all model input parameters and their source.

Face validity of the model was confirmed by Prof. Dr. A. Wimo (affiliated at the Karolinska Institutet) though remarking the limitations of not including the disease progression of behavioural problems, and the use of multiple data sources based on cohorts with different characteristics.

Model assumptions

A subject's cause of MCI (AD, other NDD or no NDD) was assigned at the model start (based on the prevalence in the LeARN study, see Appendix), and never changed because NDDs were considered non-reversible. All subjects with underlying AD or other NDD developed the dementia syndrome at some point in time if they survived long enough. Subjects with no NDD when presenting at the memory clinic were assumed to never develop dementia. The hypothetical DMT only affected subjects with AD and was assumed to have no effect on survival. No stopping rules (e.g. due to side effect) were modelled. Changes in behaviour were excluded because the long-term data that enabled to model its non-gradual nature were not available to the researchers.

Figure 2: General model structure for the incremental NMB analysis reflecting the current practice strategy and the 'verify AD' strategy. The strategy 'verify non-AD' was not reflected. This strategy is identical to 'verify AD', except that CSF is performed after clinical practice result RAD− instead of RAD+. The perfect test strategy for headroom analysis is identical to 'verify AD', except that a perfect test is performed both after clinical practice RAD+ and after the RAD− result.

Abbreviations: AD = Alzheimer's disease; NDD = neurodegenerative disease; RAD+ = result of test workup indicates AD positive; RAD− = result of test workup indicates AD negative; Dem<death = the time to dementia conversion is estimated lower than the time to death
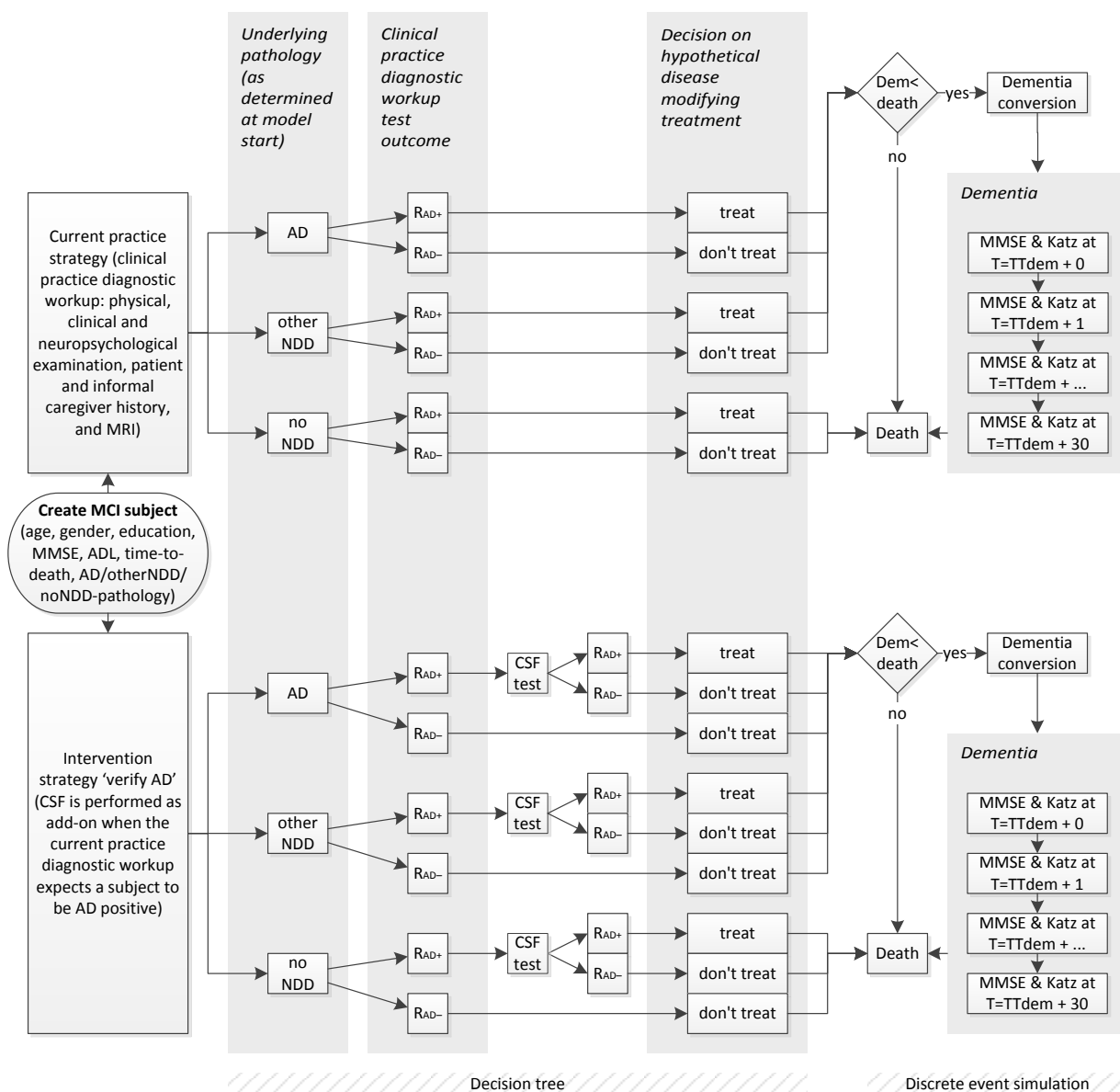
Table 1: Participant characteristics of the different data sources cohort studies and the simulated cohort

Characteristic	MCI				Dementia	
Characteristic	Kungs-holmen	MEDICIE	LEARN	Simulated cohort	Kungs-holmen	MEDICIE
Number of measurements	153	57	58	1000	323	144
Age; mean (SD)	83.4 (4.0) 76.3-97.7	74.9 (6.8) 55-87	68.4 (8.9) 50-89	66.1 (4.1) 54.0-81.1	86.7 (4.1) 78-100	78.7 (6.2) 60-94
Female; %	75%	53%	36%	42%	83%	73%
Years of education; mean (SD)	8.5 (3.0) 3-16	N.A.	11.4 (3.5) 6-17	11.5 (2.8) 6-17	8.2 (2.9) 3-16	N.A.
MMSE; mean (SD)	24.4 (2.1) 20-30	23.9 (4.3) 5-29	27.2 (2.2) 22-30	27.1 (1.9) 20.5-30	19.7 (5.0) 0-28	18.6 (5.5) 4-28
Katz; mean (SD	0.4 (0.7) 0-4	0.9 (0.7)[2] 0-2.7	0.4 (0.8) 0-4	0.5 (0.5) 0-2.5	1.2 (1.7) 0-6	1.3 (0.7) 0-2.8

[1] Katz is a surrogate Katz score based on the DAD (see appendix paragraph 3.5.1)

Predictive accuracy of treatment response

The diagnostic accuracy to predict hypothetical DMT response of both the current practice strategy as well as the 'verify AD' and 'verify non-AD' strategies is currently unknown. For the headroom analysis the sensitivity and specificity of the current practice strategy were assumed to be 77% and 68%, respectively. These estimates were derived from the LeARN study which included 251 patients suspected of having a primary neurodegenerative disease (MMSE ≥ 20, CDR 0-1). Within the LeARN study 41 subjects suffered from MCI and underwent a lumbar puncture. Current practice accuracy was reflected by estimating the accuracy with which the patient's pathology as expected by clinicians was related to the CSF Beta-Amyloid(1-42) total Tau ratio. The clinician's expectations were derived by 3 clinicians who interpreted the physical, clinical and neuropsychological examination, patient and informal caregiver history, and MRI, and judged in a consensus procedure the pathological cause of the patient's complaints (see Appendix for a detailed description).

The benefits as a result of increasing the accuracy from 77% to 100% sensitivity and 68% to 100% specificity represent the headroom. As explained in the introduction in reality 100% accurate treatment response prediction is not plausible because most likely in practice subjects are triaged to undergo a biomarker test (i.e. a biomarker is performed in a subsample of all persons visiting a memory clinic). Such subsample can consist of all patients in whom clinical practice tests indicate AD and a CSF biomarker is thus applied to verify this AD diagnosis. If the CSF result changes the diagnosis from AD to non-AD, and the change is correct, it will increase specificity. In case the change is incorrect it will decrease sensitivity (see figure 1). When subjects are triaged to verify non-AD this is vice versa. A precise estimate of the future CSF's sensitivity and specificity for predicting a not-yet-developed DMT response is hardly possible. Therefore in this early technology assessment all theoretically possible combinations of CSF sensitivity and specificity were used as input values for the incremental NMB analyses. This was done by ranging their values between the current practice value (77% sensitivity and 68% specificity) and the minimum (50%) and maximum (100%) possible accuracy. This resulted in a sensitivity range of 50% to 77% and specificity range of 68% to 100% for the 'verify AD' strategy; and a sensitivity range of 77% to 100% and specificity range of 50% to 68% for the 'verify non-AD' strategy.

Disease progression

Long-term disease progression could not be measured within the 2-year LeARN project and was therefore assessed in the Swedish Kungsholmen project, a general population-based cohort design with 6 year follow-up after incident MCI or incident dementia [13] (see Appendix). In short conversion from MCI to dementia was captured in a parametric survival analysis in 153 incident MCI persons. Age, gender, education, MMSE, and ADL, and all 2-way interactions were tested. Only female gender significantly shortened the time to dementia conversion. Symptomatic disease progression in dementia was reflected by the MMSE and the Katz score and predicted using a mixed-effects model based on the data of 323 persons with incident dementia. Time, age, education and several interactions (i.e. the interactions between time and age, MMSE and age, and MMSE and time) significantly predicted progression of MMSE and Katz.

Data was extrapolated outside the observed time interval of 6 years and truncated at the minimum and maximum scale value of the MMSE (0-30) and Katz (0-6).

The time to death was sampled from the 2012 survival table from the Dutch national central bureau for statistics and truncated at 99 years since the table did not provide estimated for older persons.

Treatment effect

A hypothetical constant treatment effect of 50% progression reduction was applied on the time from MCI to dementia, and on the rate of cognitive (MMSE) and ADL (Katz score) decline which was based on similar assumptions on hypothetical DMT efficacy found in the literature [6]. Once treatment was started it was assumed not to stop.

Costs

Cost data was analysed from 1-year resource use interview data of 201 participants from the MEDICIE study [14,20] because it was not available from the Kungsholmen database and the LeARN study insufficiently covered the disease severity spectrum of patients with dementia. Health care sector and patient and family costs were included, with an hour of informal care valued at € 9.57. Productivity costs were not included. A Katz surrogate score was constructed based on the relation between the Lawton score and the Disability Assessment for Dementia scale (DAD) items that corresponded to the Katz scale (see Appendix). In a generalized linear model, with a square root link function to transform skewed distributed cost data, MMSE and Katz surrogate score significantly predicted costs (see equations 1 and 2).

Equation 1: Formula based on a generalized linear regression model to estimate annual costs in each subject in the MCI phase (i = individual subject; Katzs = Katz surrogate score at model start)

$$Costs_{MCI,i} = (140.16 + 62.63 Katzs_i)^2$$

Equation 2: Formula based on a generalized linear regression model to estimate annual costs in each subject in the dementia phase (i = individual subject; t = model time; MMSE_{it} = Mini-Mental State Examination score of individual i at time t; Katzs_{it} = Katz surrogate score of individual i at time t)

$$Costs_{dementia,it} = (232.81 - 2.54 MMSE_{it} + 38.04 Katzs_{it})^2$$

The costs of the CSF biomarker were estimated € 211 based on expert opinion and included the performance of the lumbar puncture, assay run, analyses of total Tau, phosphorelated Tau and Amyloid-β42 protein, reporting, and quality check in a high turn-over situation. The costs of the diagnostic tests as performed in standard clinical practice (history taking in patient and informal caregiver, physical examination, neuropsychological examination, and structural MRI) were included in the measurement of resource utilization of the MEDICIE study. Treatment costs were estimated to be € 5,853 per year [6]. All costs were expressed in Euros at 2012 values (at that time, €1.00 was equivalent to US \$1.29 and British £0.81).

Health utilities

Similar to the cost analyses, quality of life analyses were performed using the MEDICIE data [14]. The proxy's perception of the patient's quality of life was measured by the EQ-5D [21] and the health utility score was calculated using the Dutch tariff [22]. Backward linear regression was used for variable selection and the final models for health utility are presented in generated equations 3 and 4.

Equation 3: Regression formula to estimate health utility in each subject in the MCI phase (i = individual subject; Katzs_{i} = Katz surrogate score of individual i at the start of the simulation)

$$U_{MCI,i} = 0.831 - 0.210 Katzs_i$$

Equation 4: Regression formula to estimate health utility in each subject in the dementia phase (i = individual subject; t = time; MMSE_{it} = Mini-Mental State Examination score of individual i at time t; Katzs_{it} = Katz surrogate score of individual i at time t)

$$U_{dementia,it} = 0.653 + 0.382 gender_i + 0.007 MMSE_{it} - 0.205 Katz_{it} - 0.018 gender_i MMSE_{it}$$

Disutility due to DMT side effects was estimated by expert opinion (FV) at a constant 0.05 as long as the treatment was provided. The non-medical consequences of an incorrect prognosis of dementia were assumed to be caused by coping with the personal decisions such as incorrect personal planning that were made based on the wrong information. The disutility of an incorrect AD positive diagnosis was estimated to be 0.25 for 1 year and for an incorrect AD negative diagnosis 0.15 for 1 year by expert opinion (FV, RH).

Using the trapezium rule the accumulated disutility was multiplied with the corresponding time period to derive QALY.

Table 2: Overview of input parameters

Category	Name	Value	Standard error	Distribution	Source type
Medication effect	Effect size	0.50	NA	Fixed	Literature[6]
Diagnosis control	Sensitivity MCI	0.77	NA	Fixed	Manuscript
	Specificity MCI	0.68	NA	Fixed	Manuscript
Diagnosis intervention	Sensitivity MCI	0.50 to 0.77[1] 0.77 to 1.00[2]	NA	Uniform	Manuscript
	Specificity MCI	0.68-1.00[1] 0.50-0.68[2]	NA	Uniform	Manuscript
MMSE	Intercept	26.87	1.24	Normal	Literature[13]
	Time (years)	-3.26	0.66	Normal	Literature[13]
	Age (years – 75)	-0.35	0.09	Normal	Literature[13]
	Time*age	0.10	0.04	Normal	Literature[13]
	Variance random intercept	2.00	NA	Fixed	Literature[13]
	Variance random slope time	1.86	NA	Fixed	Literature[13]
Katz	Intercept	-0.82	0.75	Normal	Literature[13]
	Time (years)	0.26	0.09	Normal	Literature[13]
	Age (years – 75)	0.26	0.05	Normal	Literature[13]
	Education (years)	0.06	0.03	Normal	Literature[13]
	MMSE	0.02	0.03	Normal	Literature[13]
	MMSE*age	-0.01	0.00	Normal	Literature[13]
	MMSE*time	-0.01	0.01	Normal	Literature[13]
	Variance random intercept	0.71	NA	Fixed	Literature[13]
Costs MCI	Intercept	140.16	14.43	Normal	Manuscript
	Katz	62.63	10.36	Normal	Manuscript
Costs dementia	Intercept	232.81	24.29	Normal	Manuscript
	MMSE	-2.54	0.96	Normal	Manuscript
	Katz	38.04	8.38	Normal	Manuscript
Costs treatment	Annual treatment costs	5853.25	NA	Fixed	Literature[6]
Costs CSF	Costs CSF	211.20	NA	Fixed	EO
Utility MCI	Intercept	0.83	0.04	Normal	Manuscript
	Katz	-0.21	0.03	Normal	Manuscript
Utility dementia	Intercept	0.65	0.15	Normal	Manuscript
	Gender (female)	0.38	0.16	Normal	Manuscript
	MMSE	0.01	0.01	Normal	Manuscript
	Katz	-0.21	0.03	Normal	Manuscript
	Gender*MMSE	-0.02	0.01	Normal	Manuscript
Disutility diagnosis	False positive MCI	0.25	NA	Fixed	EO
	False negative MCI	0.15	NA	Fixed	EO
Disutility treatment	Treatment side effects	0.05	NA	Fixed	EO
Time to dementia	P	2.01	0.23	Normal	Literature[13]
	Ln(gender=female)	-0.96	0.44	Normal	Literature[13]
	Intercept	-4.06	0.42	Normal	Literature[13]
Discounting	Costs	4%	NA	Fixed	EO
	Effects	1.5%	NA	Fixed	EO

1 These values were used for intervention 1 (verify AD)
2 These values were used for intervention 2 (verify non-AD)
Abbreviations: EO, expert opinion; NA, not applicable; MMSE, Mini-mental state examination; MCI, mild cognitive impairment

Outcome measures

The primary outcome measures were the average lifetime costs and average lifetime QALYs over the 2,000 individually simulated subjects. These were calculated for both the current practice

strategy and the strategy of a perfect biomarker test. QALYs were multiplied by a willingness to pay of € 80,000 per QALY [23] to express them in monetary terms. To assess the benefits achieved from each of the interventions compared to the next best alternative, which is current practice, the incremental effects were calculated. Then, the incremental costs (i.e. the difference between the current practice and perfect test strategy's costs) were subtracted from the 'in monetary terms expressed incremental QALYs (i.e. the difference between the current practice and perfect test strategies' QALYs). This formed the incremental NMB for the headroom analysis (see equation 5). Secondary outcome measures were average age at dementia conversion and average potential beneficial treatment years (calculated as the number of years that a subject suffered from AD and received the potentially beneficial DMT).

Equation 5: Incremental net monetary benefit calculation (head = headroom analysis; incr.NMB = incremental net monetary benefit; cur = current practice strategy; per = perfect test; C = costs)

$$incr.NMB_{head} =$$
$$\left(80{,}000 \frac{\sum_{i=1}^{2000} QALY_{i,per}}{2000} - \frac{\sum_{i=1}^{2000} C_{i,per}}{2000} \right) - \left(80{,}000 \frac{\sum_{i=1}^{2000} QALY_{i,cur}}{2000} - \frac{\sum_{i=1}^{2000} C_{i,cur}}{2000} \right)$$

Output representation

As explained in the section 'PREDICTIVE ACCURACY OF TREATMENT RESPONSE' when using a biomarker to 'verify AD' for the prediction of DMT response the accuracy is unknown and therefore improved specificity was ranged over all theoretically possible values (68% to 100%) at the cost of sensitivity (50% to 77%). This was operationalized by dividing both ranges into 4 equal consecutive intervals (50, 56.75, 63.5, 70.25 and 77 per cents for the sensitivity and at 68, 76, 84, 92 and 100 per cents for the specificity), and taking randomly the cutting points for sensitivity and specificity, independently of each other. This was done alongside with drawing a set of values during the probabilistic sensitivity analysis (as explained below in the 'uncertainty analyses' paragraph) to ensure that the incremental NMB outcome measure was calculated for all 25 possible combinations occurred equally (which was about 10,000/25 = 400 times). The 400 incremental NMB estimates for each corresponding combination of sensitivity and specificity were averaged and plotted in a 3-axis graph (see figure 3a). In this graph the horizontal and depth axis represent a receiver operating characteristic (ROC) plane. The combination of 77% sensitivity and 68% specificity represents the current practice situation and the other values represent the possible combinations of the 'verify AD' strategy. Next, an interpolation line was drawn through the sensitivity and specificity combinations at which the incremental NMB was zero. This represents the minimally required level of sensitivity and specificity to ensure cost-effective use of the CSF intervention when it is employed to 'verify AD'. An identical method was used for the 'verify non-AD' strategy represented by figure 4a.

Because even the accuracy of current practice accuracy in terms of sensitivity and specificity are highly uncertain a more generic graph was built in figure 5. It relies on the relative accuracy change (from current practice sensitivity and specificity to the 'verify AD' intervention or to the 'verify non-AD' sensitivity and specificity) instead of absolute accuracy values (as described in the previous paragraph) as input value in the model. The generated results can be applied in any situation of control versus intervention accuracy estimates for any biomarker in AD under the

condition that DMT is available (see Appendix). This figure 5 represents NMB from a range from the simultaneous changes in sensitivity and specificity, where 0% to 25% increase in specificity was at the cost of -25% to 0% decrease in sensitivity for the 'verify AD' relative to the current practice strategy, and 0% to 25% increase in sensitivity was at the cost of -25% to 0% decrease in specificity for the 'verify non-AD' relative to the current practice strategy. It presents the corresponding incremental NMB by isoquants.

Uncertainty analyses

Cholesky decomposition of the variance-covariance matrixes of the regression analyses (which reflect the parameter uncertainty of and correlation between the beta coefficient estimates) was performed to draw random multivariate normally distributed values of the beta coefficients in regression models. This was done for all beta coefficients simultaneously in the probabilistic sensitivity analysis which resulted in a different outcome each time the simulation of the same 2,000 subjects was run. For each strategy (the headroom, verify AD and verify non-AD) 10,000 sets of random parameter values were drawn to reflect parameter uncertainty in both the headroom and the incremental NMB analyses. The 95% credibility intervals (CI) were based on the 2.5 and 97.5 percentile of the 10,000 simulations.

In univariate sensitivity analyses of the headroom the outcomes were calculated for scenarios different from the base case by running the simulation using 2,500 iterations for each scenario. First, a more efficient treatment (DMT costs = € 1,171; DMT efficacy = 0.75) and an inefficient treatment (DMT costs = € 35,119; DMT efficacy = 0.25) were evaluated using the same estimates as used in the sensitivity analysis of Sköldunger et al. [6]. Next, the outcomes were calculated for the scenario of a high AD prevalence (75%) and a low prevalence (25%) estimate. At last, the scenarios of 10-folded CSF costs (€ 2,112), longer disutility of the non-medical consequences (disutility of a false positive (FP) and false negative (FN) diagnosis lasted during whole MCI phase), and higher disutility of DMT (0.20) were separately evaluated. The estimates were based on expert opinion (FV) since in this early stage of HTA likely estimates were unknown.

The univariate sensitivity analyses of the 'verify AD' and 'verify non-AD' strategies are presented by figure 3b and 4b respectively. By taking other scenarios (the same as described above) the specific combinations of sensitivity and specificity at which the CSF was cost-effective to 'verify AD' or 'verify non-AD' changed. This was expressed by a shift from the dotted black line (that represents the base case conditions) to the coloured lines that represent each of the sensitivity analyses conditions.

RESULTS

The 2,000 simulated MCI subjects had an average (SD) age of 68.3 (8.9), female proportion of 36%, MMSE of 27.1 (2.0), Katz of 0.51 (0.56) and 49% suffered from AD pathology.

In the current practice strategy subjects had on average (95% credibility interval) 8.67 (7.84 to 9.47) QALYs and consumed care resources worth € 545,712 (465,746 to 636,095). The headroom analysis revealed 0.39 (0.26 to 0.54) additional QALYs and € 33,622 (21,232 to 50,780) savings on average per subject if a perfect test existed (100% sensitivity and 100% specificity for predicting a hypothetical DMT response); this resulted in an incremental NMB of € 64,940 (€ 43,995 to €

90,755). Secondary outcome measures revealed 2.0 (1.7 to 2.3) additional potential beneficial treatment years and dementia conversion occurred 1.3 (1.0 to 1.7) years later than the current practice strategy.

Figures 3a and 4a represent the incremental NMB of each combination of sensitivity and specificity for the 'verify AD' and the 'verify non-AD' strategies, respectively, compared to the current practice strategy. The dotted black line indicates the specific combinations of CSF sensitivity and specificity at which the corresponding incremental NMB is, compared to current practice diagnostic workup, 0. These represent the minimal sensitivity and minimal specificity for the CSF test to be cost-effective in DMT decision making (at a willingness to pay of € 80,000).

Figure 3a (TOP): Incremental NMB of strategy 'verify AD'. CSF was performed if the current practice diagnostic workup was AD positive. The incremental NMB was drawn in a Receiver Operator Characteristic plane with sensitivity ranging from 0.50 to 0.77 and specificity from 0.68 to 1 at a willingness to pay of € 80.000.

Figure 3b (BOTTOM): Results of varying the fixed parameters from the base case in a univariate sensitivity analysis. The base case DMT costs=5853 & DMT efficacy=0.5, AD prevalence=49%, CSF costs=€211, disutility FP & FN only first MCI year, DMT side effects=0.05.

The dotted black line in both graphs represents the specific combinations of minimal sensitivity and minimal specificity at which the CSF intervention was cost-effective in the base case and thus had a positive incremental NMB. These lines correspond to each other.
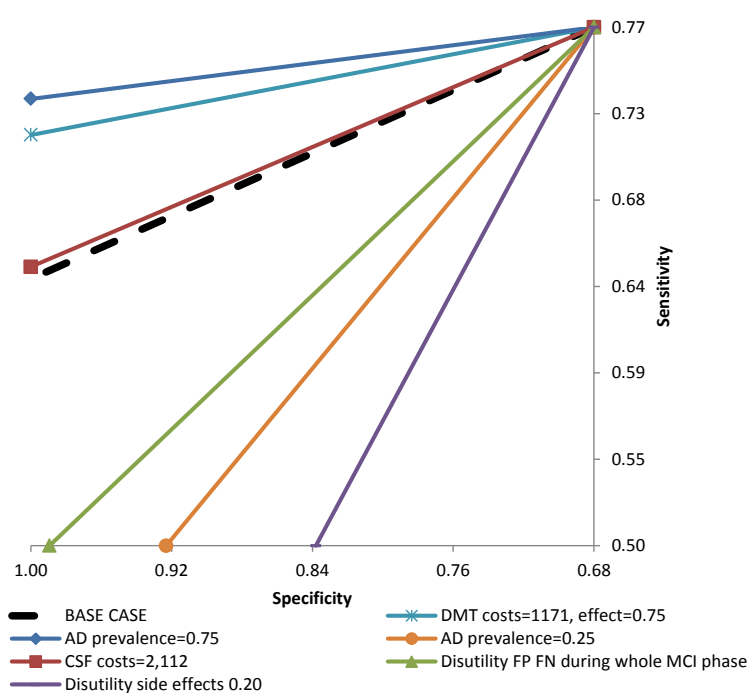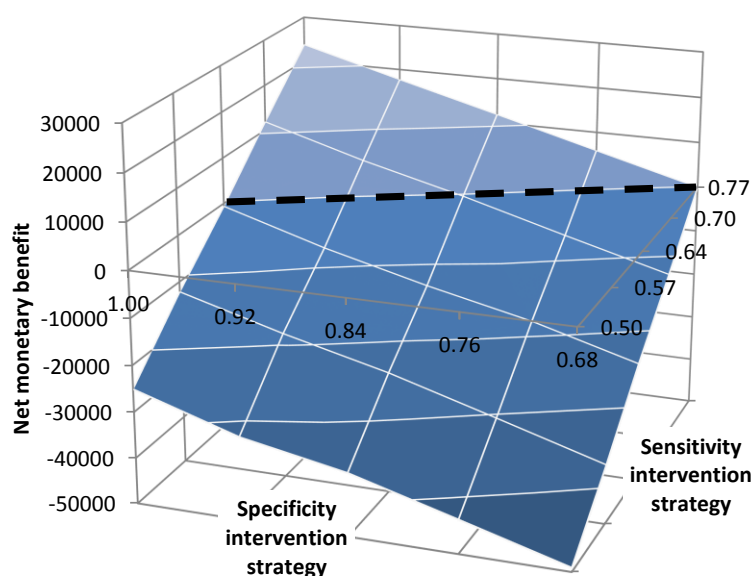
Figure 4a (TOP): Incremental NMB of the strategy 'verify non-AD'. CSF was performed if the current practice diagnosis was AD negative. The incremental NMB was drawn in a Receiver Operator Characteristic plane with sensitivity ranging from 0.77 to 1 and specificity from 0.50 to 0.68 at a willingness to pay of € 80.000.

Figure 4b (BOTTOM): Results of varying the fixed parameters from the base case in a univariate sensitivity analysis. The base case DMT costs=5853 & DMT efficacy=0.5, AD prevalence=49%, CSF costs=€211, disutility FP & FN only first MCI year, DMT side effects=0.05.

The dotted black line in both graphs represents the specific combinations of minimal sensitivity and minimal specificity at which the CSF intervention was cost-effective in the base case and thus had a positive incremental NMB. These lines correspond to each other.
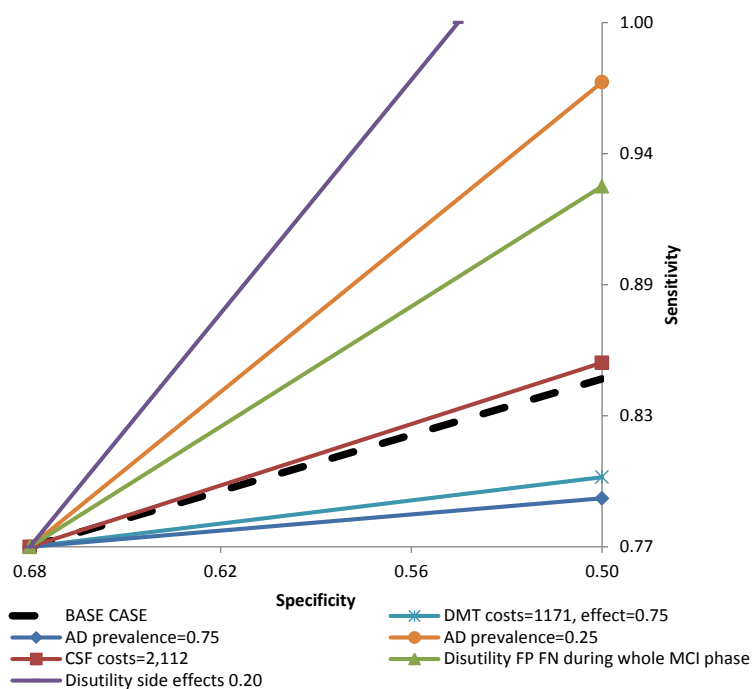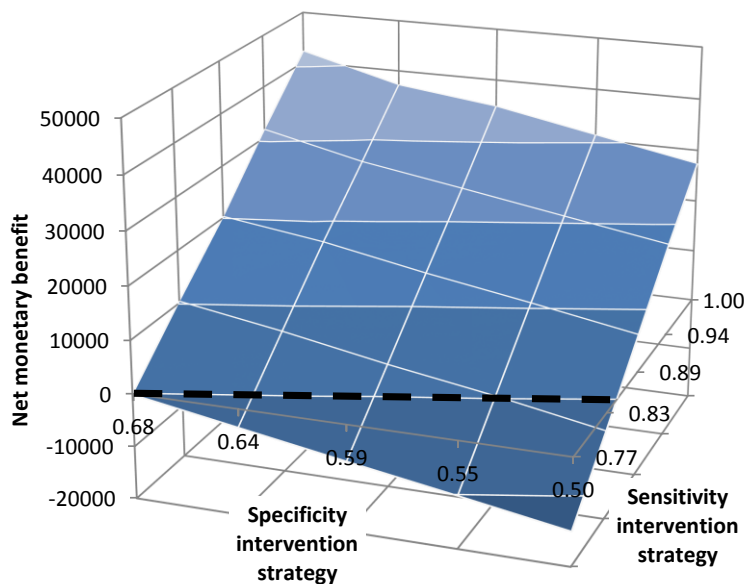
Univariate sensitivity analysis

The results of the univariate sensitivity analyses on the headroom analysis are presented in table 3. The NMB of both the current practice and the perfect test strategy was highest under the condition of a more efficient treatment.

Table 3: Primary and secondary outcome estimates (95% credibility interval based on 2,500 iterations) of the sensitivity analyses of the current practice (cur) and CSF headroom (head) under the condition of hypothetical DMT

	Costs (k€)		QALYs		NMB (k€)		Conversion age		Beneficial treatment years		Incremen -tal NMB (k€)
	Cur	Head	Cur	Head	Cur	Head	Cur	Head	Cur	Head	
Base case[1]	546	512	8.7	9.1	148	213	81.1	82.4	6.6	8.6	64.9
More efficient treatment[2]	491	452	8.9	9.4	223	299	89.8	93.7	6.6	8.6	76.0
Inefficient treatment[3]	777	741	8.2	8.4	-122	-66	78.2	78.6	6.6	8.6	55.5
High AD prevalence (75%)	548	508	8.7	9.1	144	219	82.2	83.8	10.1	13.1	74.7
Low AD prevalence (25%)	540	512	8.8	9.1	161	218	79.7	80.5	3.5	4.6	56.3
High CSF costs (€2,112)	546	514	8.7	9.1	148	211	81.1	82.4	6.6	8.6	63.0
Longer disutility FP/FN[4]	548	514	8.1	9.1	99	211	81.1	82.4	6.6	8.6	112.1
High disutility DMT (0.20)	546	512	7.5	8.0	53	126	81.0	82.3	6.6	8.6	73.9

1 Base case analysis: DMT costs = € 5,855; DMT efficacy = 0.5; AD prevalence = 49%; CSF costs = € 211; Disutility FP = 0.25 only in the first year; Disutility FN = 0.15 only in the first year; Disutility side effects = 0.05.
2 DMT costs = € 1,171; DMT efficacy = 0.75
3 DMT costs = € 35,119; DMT efficacy = 0.25
4 Disutility of the non-medical consequences of a false positive and false negative diagnosis is during the whole MCI phase

Figure 3b and 4b represent the sensitivity analyses of the incremental NMB of the 'verify AD' and 'verify non-AD'. They show that a lower AD prevalence, a longer disutility due to incorrect diagnoses and higher DMT side effect's disutility increased the range of sensitivity and specificity combinations for which CSF was cost-effective in the 'verify AD' strategy. These scenarios decreased the range of combinations for which CSF was cost-effective in the 'verify non-AD' strategy. For the other analyses (DMT costs=€1171 and effect=0.75, and AD prevalence=0.75) this was vice versa. Increased CSF costs (€2112) slightly decreased the number of cost-effective combinations. If DMT was inefficient (costs €35,119 and effect=0.25) the 'verify AD' strategy was cost-effective for all combinations of increased specificity and decreased sensitivity, and the 'verify non-AD' strategy was never cost-effective.

Figure 5 is similar to figure 3 and 4 though presents the situation of a relative change in sensitivity and specificity due to a generic biomarker intervention compared to the current practice situation for both the 'verify AD' in the right bottom quadrant and 'verify non-AD' in the left top quadrant of the graph.

Figure 5a (TOP): Incremental NMB of 1) the strategy where a biomarker was performed if the current practice diagnosis was AD positive (verify AD) in the bottom right quadrant and 2) the strategy where a biomarker was performed if the current practice diagnosis was AD negative (verify non-AD) in the left top quadrant. The black dot represents the current practice diagnostic situation. The axes present the sensitivity increase and specificity decrease of predicting DMT response due to an AD biomarker relative to current practice. The right top represents a situation with both improved sensitivity and improved specificity, the left bottom quadrant represents both decreased sensitivity and decreased specificity. These scenarios fell outside the scope of this article.

Figure 5b (BOTTOM): Results of varying the fixed parameters from the base case (base case DMT costs=5853 & DMT efficacy=0.5, AD prevalence=49%, CSF costs=€211, disutility FP & FN only first MCI year, DMT side effects=0.05) in a univariate sensitivity analysis. The intervention is cost-effective at all combinations of sensitivity and specificity at the right top of each line, and not cost-effective at the left bottom of each line.

The dotted black lines in both graphs represent the specific combinations of minimal sensitivity and minimal specificity at which the biomarker intervention was cost-effective in the base case and thus had at a positive incremental NMB. These lines correspond to each other.

131
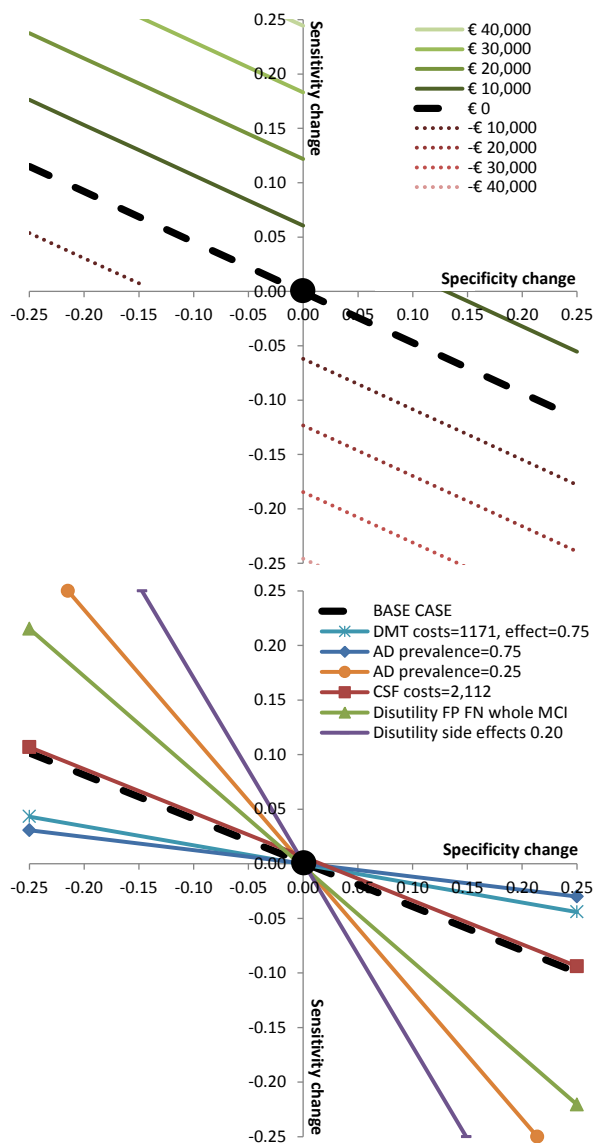
DISCUSSION

The costs and QALYs of a perfect test to reveal the headroom and two strategies in which a CSF biomarker was added to the current clinical practice diagnostic workup (either to verify an AD diagnosis or a no-AD diagnosis) were estimated. This was done by means of a patient-level decision analytic model for MCI under the condition that a hypothetical DMT was available. The analysis indicated a headroom of 0.39 (0.26 to 0.54) QALYs gained and € 33,622 (21,232 to 50,780) savings on average per subject if a perfect test existed compared to current practice, when DMT was assumed available. The strategy that employed CSF to verify non-AD as set by the current clinical practice diagnostic workup indicated more potential benefit than the strategy that employed CSF to verify AD, given the assumptions on a hypothetical DMT effect.

The reason for the larger potential benefit of the 'verify non-AD' strategy is due to the treatment benefits in subjects with AD. In a post-hoc analysis we retrospectively filtered the subjects who had AD though were not diagnosed as such by the current practice strategy but who were correctly diagnosed as AD by the CSF test and treated accordingly (i.e. undertreatment was prevented by CSF). On average these subjects gained 2.04 (0.90 to 3.34) QALYs and € 201,099 (93,555 to 347,341) savings during lifetime. This was much higher than the benefits for preventing overtreatment as revealed by a post-hoc analyses (i.e. subjects with non-AD who were wrongly diagnosed as AD and thus unnecessarily treated with DMT by the current practice and who were diagnosed as non-AD and correctly not treated by the CSF test strategy) which were 0.99 QALYs (0.95 to 1.03) gained and € 67,536 savings (64,409 to 70,594). The potential benefit in untreated AD subjects reflects the high demand on care resources in dementia, which has been estimated between €6,614 and €64,426 per year per person in Western European countries [20,24]. In the decision model DMT diminished the dementia care costs more than the assumed costs to operationalize DMT itself (€5,855 per year). The care demands for subjects with dementia were postponed by the 'verify non-AD' by 0 to 2.1 years. The impact of the required resources to operationalize the CSF test (€ 211) on the incremental NMB was minimal.

The various univariate sensitivity analyses of the 'verify AD/nonAD' changed the optimal position of the CSF test due to a shift in the balance between benefits of preventing over- and undertreatment. For example under the condition of a lower AD prevalence the number of subjects with overtreatment increased which emphasizes the importance of a high test specificity to prevent this and keep DMT provision cost-effective (i.e. the forgone resources and QALY loss due to overtreatment in a large part of the population were not compensated by the benefits of preventing undertreatment of a small part of the population).

The high and low prevalence conditions in the univariate sensitivity analyses on the headroom analysis (table 3) had only a minor effect on the costs and QALYs in the base case situation. This can be explained by the fact that not all subjects with AD survived up to the point of dementia conversion for whom the treatment benefits could not bear fruit. Furthermore, a lower prevalence of AD led to an increase of other NDDs for which no treatment was modelled. This increased the care resources required for dementia care. The potential beneficial treatment years increases with a higher prevalence because it was estimated as an average in the cohort of 2,000 subjects.

Other studies have evaluated the cost-effectiveness of hypothetical DMT combined with an AD biomarker in MCI subjects. Sköldunger et al. [6] found that a population enriched with more treatment responders was more cost-effective. They extensively elaborated on the impact of

mortality on cost-effectiveness but evaluating diagnostic strategies was not their goal. Biasutti et al. [25] showed that MRI with contrastophore-linker-pharmacophore in combination with a hypothetical DMT available at € 1144 per year was cost-effective. A similar result was found by Guo et al. [26] for Florbetaben PET. The latter study included the consequences of a reduced time to confirmed diagnosis for example on a delay of institutional care. The main conclusion of these studies indicates that their AD biomarker or enriched population under evaluation was cost-effective. Our study explicitly positions the CSF biomarker as an add-on clinical practice test to confirm or rule-out AD and elaborate on preventing over- and undertreatment. Furthermore, we estimated the headroom of a CSF biomarker.

Strengths and limitations

The model's major strengths included explicitly positioning the CSF diagnostic marker in the current clinical diagnostic workup and using input evidence that was based on long-term progression observations. The model was subject to several limitations. Lumbar puncture (headache) side effects were not taken into account. DMT was simplified by not taking willingness to receive treatment, stopping rules due to side effects, or the effect of age on treatment efficacy into account. The effect of DMT on survival was not modelled since it has a complex interaction with cost-effectiveness which has been extensively researched by Sköldunger et al. [6]. This resulted in a small proportion (1.6%) surviving dementia for more than 30 years which is a limitation of the lifetime generalizability to the real-world.

The discretization and the truncation of the polynomial disease progression formulas simplified the long-term predictions though ensured generalizability. Furthermore, disease progression in MCI phase was simplified by reflecting it with a single time-to-dementia estimate while a gradual decline of -0.6 MMSE points per year was observed in the Kungsholmen Project. Also, the transformation of the DAD to a Katz surrogate score using the Lawton scale most likely generates uncertainty to how well they reflect the actual estimates. The possible error of these simplifications is identical in both the control and intervention strategy and does not influence the incremental estimates. Therefore the impact on the conclusions was judged small, although the exact impact of these simplifications remains unknown.

Resource use and utility scores were obtained from a clinical sample and disease progression from a general population-based sample. Memory clinic subjects are likely to have more severe problems at diagnosis because the complaint must be severe enough to initiate a visit. This might underestimate the CSF benefits since these subjects miss a window of opportunity for early treatment.

The study relied on assumptions of a hypothetical DMT efficacy and a biomarker's accuracy to predict its response. A similar situation counts for the non-medical consequences of an early diagnosis on quality of life which were unknown, especially the impact of an incorrect diagnosis (i.e. false positive or false negative). Including such consequences, even if the true value is unknown, has been emphasized [27]. The unknown of several parameters is inherent to this early assessment of technologies. For this reason we deviated from the conventional method of comparing a limited number of alternative strategies and evaluated any combination of test sensitivity and specificity. It enabled us to show the most optimal position of a new AD biomarker under various predefined conditions.

Conclusions

CSF and related AD biomarkers have the potential to increase care efficiency when DMT becomes available. The incremental NMB analysis indicated more potential benefit from a biomarker that was positioned to verify subjects who are not expected to have AD (i.e. a treatment selection approach to prevent undertreatment) rather than to verify subjects expected of AD (a treatment selection approach to prevent overtreatment) given the assumptions on a hypothetical DMT.

ACKNOWLEDGMENT

REFERENCES

[1]   Wimo A, Jönsson L, Bond J, Prince M, Winblad B. The worldwide economic impact of dementia 2010. Alzheimers Dement 2013;9:1–11.e3.

[2]   Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013;9:63–75.e2.

[3]   Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment 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 2011;7:270–9.

[4]   Spies PE, Slats D, Ramakers I, Verhey FRJ, Olde Rikkert MGM. Experiences with cerebrospinal fluid analysis in Dutch memory clinics. Eur J Neurol 2011;18:1014–6.

[5]   Bossuyt PMM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Med Decis Making 2009;29:E30–8.

[6]   Sköldunger A, Johnell K, Winblad B, Wimo A. Mortality and treatment costs have a great impact on the cost-effectiveness of disease modifying drugs in Alzheimer's disease. Curr Alzheimer Res 2012:1–10.

[7]   Handels RLH, Wolfs C a G, Aalten P, Joore M a, Verhey FRJ, Severens JL. Diagnosing Alzheimer's disease: A systematic review of economic evaluations. Alzheimers Dement 2013:1–13.

[8]   Mangialasche F, Solomon A, Winblad B, Mecocci P, Kivipelto M. Alzheimer's disease: clinical trials and drug development. Lancet Neurol 2010;9:702–16.

[9]   Hartz S, John J. Public health policy decisions on medical innovations: what role can early economic evaluation play? Health Policy 2009;89:184–92.

[10]   Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. Pharmacoeconomics 2006;24:1043–53.

[11]   Cohen JT, Neumann PJ. Decision analytic models for Alzheimer's disease: state of the art and future directions. Alzheimers Dement 2008;4:212–22.

[12]   Handels RL, Aalten P, Wolfs CA, Olde Rikkert M, Scheltens P, Visser PJ, et al. Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease: the research protocol of a prospective cohort study. BMC Neurol 2012;12:72.

[13]   Handels RLH, Xu W, Rizzuto D, Caracciolo B, Wang R, Winblad B, et al. Natural Progression Model of Cognition and Physical Functioning among People with Mild Cognitive Impairment and Alzheimer's Disease. J Alzheimers Dis 2013;37:357–65.

[14] Wolfs CAG, Dirksen CD, Kessels A, Severens JL, Verhey FRJ. Economic evaluation of an integrated diagnostic approach for psychogeriatric patients: results of a randomized controlled trial. Arch Gen Psychiatry 2009;66:313–23.

[15] Cosh E, Girling A, Lilford R, McAteer H, Young T. Investing in new medical technologies: A decision framework. J Commer Biotechnol 2007;13:263–71.

[16] Russ TC, Morling JR. Cholinesterase inhibitors for mild cognitive impairment. Cochrane Database Syst Rev 2012;9:CD009132.

[17] Hakkaart- van Roijen L, Tan SS, Bouwmans CAM. Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg 2010.

[18] 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;12:189–98.

[19] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of adl: a standardized measure of biological and psychosocial function. JAMA 1963;185:914–9.

[20] Handels RLH, Wolfs CAG, Aalten P, Verhey FRJ, Severens JL. Determinants of care costs of patients with dementia or cognitive impairment. Alzheimer Dis Assoc Disord n.d.;27:30–6.

[21] Brooks R. EuroQol: the current state of play. Health Policy 1996;37:53–72.

[22] Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJ V. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006;15:1121–32.

[23] RVGZ. Zinnige en duurzame zorg. Raad Voor Volksgezond En Zorg Aan Minist van Volksgezond Welz En Sport 2006.

[24] Jönsson L, Wimo A. The cost of dementia in Europe: a review of the evidence, and methodological considerations. Pharmacoeconomics 2009;27:391–403.

[25] Biasutti M, Dufour N, Ferroud C, Dab W, Temime L. Cost-effectiveness of magnetic resonance imaging with a new contrast agent for the early diagnosis of Alzheimer's disease. PLoS One 2012;7:e35559.

[26] Guo S, Getsios D, Hernandez L, Cho K, Lawler E, Altincatal A, et al. Florbetaben PET in the Early Diagnosis of Alzheimer's Disease: A Discrete Event Simulation to Explore Its Potential Value and Key Data Gaps. Int J Alzheimers Dis 2012;2012:548157.

[27] Fang C, Otero HJ, Greenberg D, Neumann PJ. Cost-utility analyses of diagnostic laboratory tests: a systematic review. Value Health 2011;14:1010–8.

APPENDIX: COST-UTILITY MODEL BACKGROUND AND METHODS

This appendix contains additional information to the manuscript "Early cost-utility analysis of general and cerebrospinal fluid-specific Alzheimer's disease biomarkers for hypothetical disease-modifying treatment decision in mild cognitive impairment"

Ron L.H. Handels

Manuela A. Joore

An Tran-Duy

Claire A. G. Wolfs

Frans R. J. Verhey

Johan L. Severens

1   Background

In the manuscript "Early cost-utility analysis of general and cerebrospinal fluid-specific Alzheimer's disease biomarkers for hypothetical disease-modifying treatment decision in mild cognitive impairment" the rationale, scope, basic concepts and outcomes of a decision analytic model to evaluate the use of cerebrospinal fluid (CSF) or related biomarkers for a diagnosis of AD in combination with a hypothetical treatment for Alzheimer's disease (AD) in subjects suffering from mild cognitive impairment (MCI) were reported. This appendix provides a detailed overview of the methods for developing the decision analytic model and model results.

1.1   Disease mechanism

AD is an irreversible, progressive brain disease that disrupts critical metabolic processes that keep neurons healthy. People with AD have an abundance of two abnormal structures in their brains: amyloid plaques (Aβ) and neurofibrillary tangles (Tau). This is caused by deformed proteins and particularly occurs in brain areas responsible for memory. These disruptions cause dysfunction of brain cells, lose connections to other brain cells and finally cell death. Brain cell destruction and cell death cause memory failure, personality changes, and problems in carrying out daily activities. It is believed that the abundance of abnormal structures begins 10 to 20 years before clinically detectable symptoms appear.

Research criteria developed in 2011 formalize three disease stages: preclinical, symptomatic predementia MCI, and dementia [1–4]. The preclinical stage, in which persons do not suffer from disease related symptoms, was outside the scope of this research.

1.2   Dementia

Dementia is defined by multiple cognitive deficits (in memory and one or more of the following: aphasia, apraxia, agnosia, disturbance in executive functioning) that cause significant impairment in social or occupational functioning, represent a significant decline from a previous level of functioning and are not due to systemic or central nervous system conditions or occur in delirium [DSM-IV] [5]. Disease progression is characterised by worsening of cognition, functional ability, and behaviour and mood. Disease severity can be categorized into mild, moderate or severe, in which the ability to live alone without problems or care diminishes with increasing severity. A variety of underlying brain pathologies can cause the dementia syndrome (e.g. Alzheimer's disease, vascular dementia, dementia with Lewy bodies, frontotemporal dementia and other non-frequent causes).

1.3   Symptomatic predementia MCI phase

MCI represents the transitional state between normal cognitive decline due to aging and the earliest clinical features of dementia. The original criteria define MCI by: memory complaint, memory impairment documented according to appropriate reference values, essentially normal performance in non-memory cognitive domains, generally preserved activities of daily living, and not demented [6]. Diagnosing MCI enables to identify mildly impaired subjects possibly in a

transitional stage between normal aging and dementia though not all patients suffering from MCI deteriorate [7].

1.4   Describing disease progression

Disease progression can be expressed by pathological changes reflected by a variety of AD biomarkers [8] and by clinical symptoms. In the MCI phase symptomatic disease progression is typically described by an annual conversion rate to dementia [9]. In the dementia phase it is recommended to be described by cognition, activities of daily living (ADL) and behavioural problems [10,11]. Several scales are available to describe these 3 domains though clinical criteria do not describe pre-defined cut-offs values to set the diagnosis of dementia in a subject.

Disease progression is schematically presented in Figure 1. It represents: subjects with MCI and no underlying AD pathology who do not convert to dementia; conversion from MCI to dementia due to AD where no DMT was available (early conversion); treatment with a disease modifier that prolongs conversion from MCI to dementia and decreases the speed of progression (late conversion).

Figure 1: Theoretical representation of progression pathway of a subject with no AD pathology (straight line), AD pathology under DMT (late conversion and gradual decline) and AD pathology without DMT (early conversion and steep decline); † = death
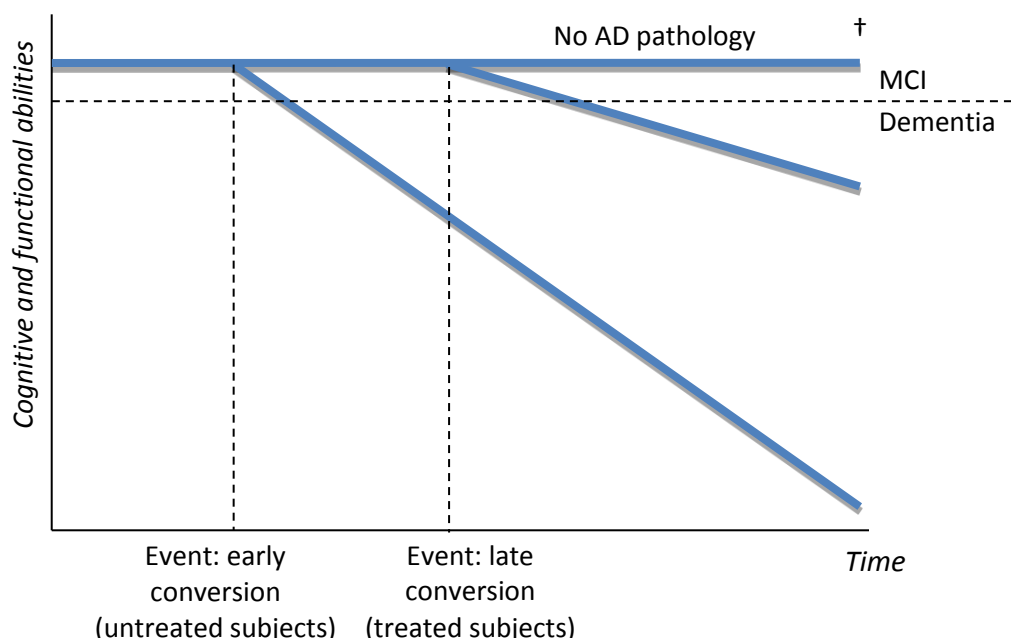
1.5   Diagnostic tests

Practice guidelines [12,13] describe how to set a diagnosis of AD type dementia using diagnostic criteria [14,15]. It is intended for all patients with suspected or diagnosed dementia. Several diagnostic tools are recommended in routine practice:

- History taking in patient and informal caregiver (to assess the cognitive domains affected, ADL, medical history, co-morbidities, family history, educational history, neurological and physical examination, behavioural and psychological symptoms)
- Cognitive assessment and quantitative neuropsychological testing
- Assessment of co-morbidities: Depression, hypothyroidism (blood test), B12 deficiency (blood tests)
- Structural imaging (non-contrast CT or T1, T2 and FLAIR MRI sequences)
- In case of diagnostic doubt or suspected other (neurodegenerative) disorders several other tests can be performed.

Scientific knowledge, advanced imaging techniques and cerebrospinal fluid analyses have evolved since the publication of the NINCDS-ADRDA criteria in 1984. This has led to much debate and the proposition of new clinical and research criteria to enhance diagnostic accuracy, even at the stage of early clinical symptoms [3,4]. These criteria distinguish between the AD pathophysiological process and the clinically observable syndrome to enable determination of AD in the pre-dementia state of mild cognitive impairment (MCI). Eventually the criteria are meant to support therapy decision making (when effective treatments are available) or to determine the likelihood of cognitive and functional progression to a more severe disease state. Emerging biomarkers are attributed a more prominent role in these new diagnostic criteria; amyloid β42, total tau and phosphorylated-tau in the CSF, amyloid tracer uptake and fluorodeoxyglucose in positron emission tomography, hippocampal volume and medial temporal atrophy in structural magnetic resonance imaging (MRI) and single photon emission tomography (SPECT) perfusion imaging. Evaluation of these biomarkers before adopting them in clinical practice was advised.

1.6   Treatment

The use of cholinesterase inhibitors for symptomatic treatment of AD-type dementia is advised in the Dutch guidelines for dementia treatment [16]. Treatment with Memantine is advised in moderate to severe AD-type dementia. No drug treatments are registered for MCI subjects. Despite the fact that no new drug treatments have been approved since 2002, a variety of compounds that is under investigation and developments in unrevealing the disease mechanism suggests that it is likely for new pharmacotherapeutic AD treatments to become available over the next decade [17].

1.7   Principal choices for modelling

For this research an individual subject simulation model was considered the most flexible to describe gradual disease progression in cognition and functioning and build in all relationships between diagnostic test outcomes, treatment and disease progression, and therewith best reflects clinical practice. A combination of a decision tree and a DES model was used. The basic concepts of DES include entities, attributes, events, relationships and outcomes. Entities typically represent a patient that 'walks' through the model. Attributes are the characteristics or features attached to an entity (e.g. age, sex, quality of life, is under treatment, etc.). Events are the things that can happen to an entity during the simulation at a particular moment in time (e.g. start of treatment, conversion from MCI to dementia, adverse treatment reaction, etc.). Events can have an influence on the attributes of an entity. For example when a patient (entity) is positively

diagnosed (event), a treatment can be initiated (event) and cognitive decline (attribute) can slow. At last, relationships are mathematical descriptions that link all model elements (entities, attributes, events, and outcomes) and calculate model outcomes (costs and quality adjusted life years).

2 Model characteristics

2.1 Model structure

The model started with 2,000 individual incident MCI subjects (entities) with each having a specific age, gender, education, pathology (AD, other neurodegenerative disease (NDD) or no NDD), MMSE score, Katz score, and time to death. The subjects were quadruplicated and ran through the control strategy, the perfect test strategy, and the two intervention strategies. AD was diagnosed using clinical practice tests only in the control strategy and using CSF as an add-on test to clinical practice tests in the perfect test and the two intervention strategies. The diagnosis was based on the prevalence of AD pathology and test sensitivity and specificity of either the control or intervention strategies. Subjects who were tested AD positive received a hypothetical DMT. Next, time to dementia conversion was estimated which reflected disease progression from MCI to dementia. If a subject was treated this delayed the conversion time. At the same moment the time to death was estimated. The event with the shortest time was let occur. As recommended by Cohen et al. [10] and Green et al. [11] disease progression in the dementia phase was modelled by the change in cognition and ADL on a continuous scale. Cognition was reflected by the MMSE [18] and ADL by the Katz index score [19]. Changes in behaviour were excluded because data on this parameter was not available to the researchers. In the dementia phase the effect of treatment was modelled as a reduction of the speed of decline. Eventually the model stopped when the subject died or after being 30 years in the dementia phase. Figure 2 of the manuscript represents the general model structure of the control strategy (i.e. current practice) and the 'verify AD' strategy (as explained below). The model was built in Microsoft Excel 2010.

2.2 Assumptions

Several assumptions were made in the model:
- A subject's cause of MCI (AD, other NDD or no NDD) was assigned at the model start (based on the prevalence in the LeARN study, see paragraph 3.3), and never changed because NDDs were considered non-reversible.
- All subjects with underlying AD or other NDD developed the dementia syndrome at some point in time if they survived long enough. Subjects with no NDD never developed dementia.
- The hypothetical DMT was assumed to affect the Aβ and Tau pathology of AD and not the pathological features of other NDD
- The hypothetical DMT only affected subjects with AD, was assumed to have no effect on survival, and no stopping rules (e.g. due to side effect) were modelled.
- A hypothetical DMT was provided only in subjects with an AD positive diagnosis

2.3   Strategies under evaluation

The following strategies were compared: 1) a control strategy reflecting clinical practice diagnostic workup in MCI subjects, 2) a perfect test strategy for the headroom analysis, 3) an intervention strategy with CSF complementary to the clinical practice diagnostic workup in MCI subjects when AD was diagnosed by the current clinical practice, and 4) similar to 3 though when AD was not diagnosed by the current clinical practice. The diagnostic test workup of the control strategy consisted of the following tests: history taking in patient and informal caregiver, physical examination, neuropsychological examination, and structural MRI, but no CSF test. Applying this diagnostic test workup led to an aetiological diagnosis of either AD pathology or no AD pathology (see paragraph 3.3). The diagnostic workup of the intervention strategies consisted of the same tests as the control strategy including the CSF Amyloid β42 total tau ratio as an add-on test. The result of the diagnostic test workup in the control as well as in the intervention strategies 3 and 4 was a dichotomous conclusion of AD pathology (see RAD+ in Figure 2 of the manuscript) or no AD pathology (see RAD– in Figure 2 of the manuscript). In both strategies each subject who was diagnosed with AD was treated with a hypothetical future DMT. The consequences of treatment are explained in paragraph 3.5.

CSF was positioned as an add-on test to either verify the test results of AD pathology based on the clinical practice diagnostic workup (strategy 3, displayed in Figure 2 of the manuscript), or to verify the test results of no AD pathology based on the clinical practice diagnostic workup (strategy 4, not displayed in Figure 2 of the manuscript). To verify no-AD the CSF test was positioned after the AD– instead of after AD+ in Figure 2 of the manuscript. Thus AD was expected if both the standard clinical diagnostic workup test as well as the CSF test was AD positive, an AD negative diagnosis was made if the standard clinical diagnostic workup was AD negative (referred to as intervention strategy 'verify AD'). In strategy 4 AD was expected if the standard clinical diagnostic workup test was AD positive, an AD negative diagnosis was made if both the standard clinical diagnostic workup as well as the CSF test was AD negative (referred to as intervention strategy 'verify non-AD'). Each row of Table 1 represents for an individual the disease state, diagnostic conclusion and it consequences in relation to treatment.

Table 1: Consequences of diagnostic aetiological test outcome in AD, other NDD and no NDD (the point of departure for the described consequences (i.e. the last 2 columns) is a subject having no NDD as represented by the last row)

AD pathology	Specification other pathology	Diagnostic conclusion	Treatment decision	Effect on progression	Diagnostic group regarding AD status	Consequences for costs	Consequences for quality of life
AD	na	R_{AD+}	treat	slowed	TP	+ treatment + AD-disease	+ treatment effect − treatment side effects − AD-disease
AD	na	R_{AD-}	don't treat	fast	FN	+ AD-disease	− false diagnosis − AD-disease
no AD	other NDD	R_{AD+}	treat	fast	FP	+ treatment + other NDD disease	− treatment side effects − false diagnosis − other NDD-disease
no AD	other NDD	R_{AD-}	don't treat	fast	TN	+ other NDD-disease	− other NDD-disease
no AD	no NDD	R_{AD+}	treat	stable	FP	+ treatment	− treatment side effects − false diagnosis
no AD	no NDD	$R_{AD}-$	don't treat	stable	TN	(point of departure)	(point of departure)

Abbreviations: na, not applicable; TP, true positive; FN, false negative; TN, true negative; FP, false positive; AD, Alzheimer's disease; NDD, neurodegenerative disease; R_{AD+}, result of the diagnostic conclusion is AD; R_{AD-}, result of the diagnostic conclusion is non-AD; +, increased; −, decreased.

2.4 Model elements

The model entity was an individual patient. This patient reflected a person who visited a memory clinic due to memory complaints and was concluded to suffer from the MCI syndrome. Such a subject was selected from a pre-defined cohort (see paragraph 3.1 for more details) and entered the model (step 1) by reading in its attributes (step 2):

- Age
- Gender
- years of formal education
- AD pathology present
- MMSE
- Katz

Decision model step	Item	Relationship
1) Entity	MCI patient enters the model	
↓		
2) Read in base case attributes	Gender	Probability = 0.36 for female gender
	Age	Average = 68.4, sd = 8.9
	Years of education	Average = 11.4, sd = 3.5
	Cognition (MMSE)	Average = 27.2, sd = 2.2
	ADL (Katz)	Average = 0.4, sd = 0.8
	Pathology	Probability AD = 0.49, other NDD = 0.24, no NDD = 0.27
	AD diagnosed	No
	Hypothetical future medication	No
	Time passed	0
↓		
3) Quadru-plicate patient		Entity will go through the model's control strategy
		Entity will go through the model's intervention strategy 'perfect test'
		Entity will go through the model's intervention strategy 'verify AD'
		Entity will go through the model's intervention strategy 'verify non-AD'
↓		
4) Time to event	Dementia conversion	Time was estimated from survival function
	Death	Time was estimated from survival function
↓		
5) Update attributes	Gender, education, AD pathology	Did not change over time
	Time passed	= time at previous event + passed time until current event
	Age	= age + time passed since last event
	AD diagnosed	Dependent on sensitivity and specificity test characteristics which were dependent on the intervention or control strategy
	Hypothetical treatment	Provided if AD was positively diagnosed
	MMSE	IF MCI then MMSE did not change. IF dementia then MMSE was estimated from mixed regression function using attributes: time, age, medication. MMSE was updated annually within the dementia phase
	Katz	IF MCI then Katz did not change IF dementia then Katz was estimated from mixed regression function using attributes: time, age, education, MMSE, medication. Katz was updated annually within the dementia phase.
↓		
6) Accumulate outcomes	Cost	= treatment costs + intervention costs + regression formula using attributes MMSE and Katz
	Utility	= disutility of (in)correct test + disutility of treatment + regression formula using attributes gender, MMSE and Katz

The attributes 'AD diagnosed', 'hypothetical future medication provided' and 'time passed since model start' were assumed no, no and 0 respectively. Next the subject and its attributes were

quadruplicated (step 3) and one ran through the control strategy and the others through the intervention strategies. Two events could occur in a subject: conversion from MCI to dementia and death. The time to these events was estimated at model entry based on the subject's attributes and the event took place on the exact estimated time and was not discretized (step 4). A subject could follow 2 different pathways: from model start to death via dementia conversion or from model start to death without dementia conversion. The attributes MMSE, Katz, AD diagnosed, treatment applied, and time passed since model start were updated (step 5) at the event of dementia conversion. Due to the gradual change in both MMSE and Katz a direct estimation of total costs and QALY over the period between the 2 events dementia onset and death was difficult. Capturing a change in MMSE and Katz by events (e.g. event of declining 1 MMSE point) was also difficult because their non-linear nature of decline and the interrelation between them. Such model would result in small intervals and turned out not significant. Due to these limitations and as alternatively proposed [20] progression in the dementia phase was cut in periods of 1 year and attributes were updated annually, independent from the events. Outcomes consisted of cumulative QALY and costs (step 6). The time horizon was lifetime. Table 2 provides a basic overview of the model steps, the items affected with each particular step, and the relationship or logical rule how the item was determined.

3   Input

3.1   Data sources

Input parameters were mainly derived from existing databases available to the researchers. These include the database from the LeARN study [21], the Kungsholmen project [22] and the MEDICIE study [23].

The LeARN study was specifically designed to estimate the diagnostic accuracy of CSF and other biomarkers in PET and MRI. In a cohort-based design 241 subjects were followed in a clinical setting for 2 years. The study's Inclusion criteria were: suspected of having a primary neurodegenerative disease (MMSE ≥ 20, CDR 0-1). Subjects with normal pressure hydrocephalus, Huntington's disease, CVA/TIA in last 2 years, psychiatric history, major depression in last year, alcohol abuse, brain tumour or epilepsy were excluded. Fifty eight subjects suffered from MCI [see chapter 4].

Long-term disease progression could not be measured within the LeARN project due to its 2-year study period. For disease progression the Swedish Kungsholmen Project data was analysed due to its long-term data (6 years after incident MCI or incident dementia cases). All subjects from the Kungsholmen area were approached and general health measures including cognition and ADL were measured. For this model, a subsample of subjects who suffered from MCI or from dementia were selected for analyses to enable the prediction of natural progression both in MCI and dementia phase [22].

At last data from the MEDICIE study was used to estimate the relation between resource use, quality of life and disease severity. In the MEDICIE study a cohort of 219 subjects was followed for 1 year to measure resource use and quality of life in memory clinics, general practitioners and community mental health teams. It covered the MCI, mild and moderate dementia phase and therefore included more severely demented subjects than the LeARN study which was important to reflect resource use and quality of life of the whole spectrum of disease severity. Subjects were

aged 55+, suspected or having dementia or a cognitive disorder and had no other referral. Subjects with acute disorders that required prompt therapeutic intervention and residence in a nursing home were excluded. Only subjects who were recruited from secondary care providers were included (n=201).

All other input parameters estimates were retrieved from the literature or expert opinion. Table 1 of the manuscript presents the baseline characteristics of the data sources. Table 2 of the manuscript provides an overview of all input parameters and their source.

3.2   Simulated cohort

The simulated cohort was based on the MCI subsample from the LeARN study. Average and standard deviation was estimated for age, gender, education, MMSE and Katz in the LeARN study from 58 MCI subjects. A simulation cohort of 2,000 individuals was generated by sampling each subject's characteristics from this average and SD or probability and truncated at minimum and maximum possible MMSE and Katz scale scores. The simulated cohort was sampled once and this same sample was used in all sensitivity analyses.

Part of the input parameters reflected variability: the parameters to determine the cohort characteristics (age, gender, education, MMSE, Katz, AD presence), time to dementia conversion, time to death, diagnostic sensitivity and specificity, and the random effects of the regression functions for MMSE and Katz). This resulted in a different outcome every time a new simulation cohort was sampled (this variety reflected individual patient differences and did not reflect probabilistic parameter uncertainty). Simulation cohort size was increased until the deterministic model result was stable by trial and error. Simulation cohort sizes of 200, 500, 1000 and 2000 were drawn each 10 times and the following model outcomes that were considered of importance were averaged and evaluated: total costs and QALY, age, AD prevalence, start MMSE and Katz, time to dementia, time to death, and age at death. The average, minimum and maximum of these 10 cohort averages were assessed and considered valid if the range indicated less than 5% variation [20] or were clinically irrelevant (see Table 3). A simulation cohort size of 2000 was adopted since only the Katz had more than 5% variation (6.5%) but the absolute difference was considered clinically irrelevant and was mainly caused by truncating the value.

Table 3: Assessment of minimum, average and maximum to determine cohort critical simulation cohort size

	200	500	1000	2000
Start age	68, 69, 70[1]	68, 68, 69[1]	68, 68, 69[1]	68, 68, 69[1]
AD prevalence	0.44, 0.5, 0.55	0.46, 0.49, 0.53	0.47, 0.49, 0.51	0.48, 0.49, 0.5[1]
MMSE	27, 27, 27[1]	27, 27, 27[1]	27, 27, 27[1]	27, 27, 27[1]
Katz	0.46, 0.54, 0.6	0.49, 0.51, 0.56	0.49, 0.52, 0.54	0.51, 0.53, 0.55
TTdementia	7.59, 8.65, 9.57	7.96, 8.3, 8.84	7.91, 8.15, 8.42	8.06, 8.2, 8.42[1]
TTdeath	16, 17, 18	17, 18, 18	17, 17, 18	17, 18, 18[1]
Age at death	85, 86, 87[1]	85, 86, 87[1]	85, 86, 86[1]	86, 86, 86[1]
Total costs intervention (x1000)	430, 470, 506	464, 490, 509	469, 487, 516	481, 490, 499[1]
Total costs control (x1000)	436, 478, 516	471, 497, 518	473, 493, 522	487, 497, 506[1]
Total QALY intervention	8.16, 8.6, 9.11	8.46, 8.8, 9.02	8.15, 8.6, 8.78	8.5, 8.62, 8.8[1]
Total QALY control	8.04, 8.49, 8.99	8.39, 8.71, 8.93	8.08, 8.51, 8.7	8.42, 8.52, 8.71[1]

1At this cohort size the range had less than 5% variation

145

Because the mentioned input parameters reflected variability this might produce a difference between identical comparison strategies. Therefore, the random values generated to reflect variability were used both in the control and intervention strategy [20] (e.g. de random number to determine the random intercept of the mixed model for MMSE was generated once and applied in the regression formula both in the control and all intervention strategies).

3.3 Predictive accuracy of treatment response

Both the diagnostic accuracy of current practice strategy to predict hypothetical DMT response as well as the 'verify AD' and 'verify non-AD' to predict hypothetical DMT response are unknown. Therefore all theoretically possible estimates were applied in this model. For the current practice, the minimally predictive accuracy was reflected by the accuracy of current practice to predict a dichotomized CSF AD outcome. Plausible values of CSF lay between the minimal accuracy as reflected by the current practice (50% for sensitivity or specificity) and 100% accurate prediction of DMT response.

The best estimate for sensitivity and specificity of the current clinical practice was assumed 77% and 68% respectively. These estimates were derived from the LeARN study. The data of 41 subjects with MCI who had CSF measurements at their disposal was obtained. CSF was dichotomized as abnormal (CSF+) or normal (CSF−) using predefined cut-off values for the Aβ1-42/totalTau ratio (Aβ1-42/(240 + (1.18 * Tau))) <1 [24].

Sensitivity and specificity of the current practice strategy was determined by relating the patient's pathology as expected by clinicians to the CSF Beta-Amyloid(1-42) total Tau ratio. In other words, CSF formed the reference standard against to determine the current practice's accuracy as a best estimate of treatment response prediction (making the underlying assumption that CSF reflects AD pathology and that it is equal to treatment responders). Clinical expectations of AD were derived by asking 3 clinicians to interpret the physical, clinical and neuropsychological examination, patient and informal caregiver history, and MRI. If discrepancies occurred between the 3 expectations the experts were asked to arrive at consensus in a discussion meeting [see chapter 4].

As explained in the introduction of the manuscript in reality 100% accurate treatment response prediction is unlikely because most probably subjects are triaged to undergo a biomarker test (i.e. a biomarker is performed in a subsample of all persons visiting a memory clinic). The 'verify AD' strategy positioned CSF as add-on to verify an 'AD pathology' outcome based on the clinical practice diagnostic workup. Therefore, a diagnosis of AD was provided more conservatively than the control strategy which will result in a specificity gain at the cost of sensitivity to predict DMT response. Plausible sensitivity values ranged from the minimum (50%) to the maximum (77%, similar to the control strategy) and specificity from the minimum (68%, similar to the control strategy) to the maximum (100%) (see Table 4). The 'verify non-AD' intervention positioned CSF as add-on to verify a 'no AD pathology' outcome based on the clinical practice diagnostic workup which will result in a sensitivity gain at the cost of specificity. Plausible sensitivity ranged from 77-100% and specificity from 50-68% (see Table 4). Any combination of sensitivity and specificity within the theoretically possible range could represent a future scenario of a test and treat protocol and was evaluated in this model.

Table 4: Diagnostic accuracy of MCI subjects in the control strategy (by relating clinical judgment aetiology to CSF) and both intervention strategies 'verify AD' and 'verify non-AD' (range between minimum and maximum accuracy)

	Sensitivity	Specificity
Control strategy	77%	68%
Intervention with CSF as add-on to verify 'AD pathology' as the test results based on the clinical practice diagnostic workup (range)	50% to 77%	68% to 100%
Intervention with CSF as add-on to verify 'no AD pathology' as the test results based on the clinical practice diagnostic workup (range)	77% to 100%	50% to 68%

In the model each subject was labelled as TP, TN, FP, or FN. If a random number from a uniform distribution was lower than the AD prevalence a subject was labelled as AD. Then, in AD, if another randomly drawn number from a uniform distribution was lower than the test's sensitivity the subject was labelled as test positive and thus a true positive. If this number was higher than the sensitivity the subject was labelled as FN. A similar calculation was performed for the specificity if the subject was determined as non-AD. Because these 3 random numbers differ for each cohort run they were, as explained at the end of the previous paragraph, used both in the control and intervention within each subject (e.g. the random number for AD in MCI is 0.80, sensitivity is 0.77 in the control group and 0.85 in the intervention group. The random number is higher than the sensitivity in the control group and thus determined as non-AD (False negative), the random number is lower than the sensitivity in the intervention group and thus determined as AD (True positive)).

The AD prevalence was derived from the subsample of subjects with CSF available and based on an expert panel diagnosis using the information on the 2-year course of decline [See chapter 3]. The prevalence of AD was measured in the LeARN study and estimated 49%. Subjects with no AD (51%) could suffer from another NDD (48%) or no NDD (52%). Since no data on other NDDs was available to the researchers other NDDs were modelled similar to AD as a best estimate.

## 3.4   Disease progression

3.4.1   Time to dementia

Time to dementia conversion was based on a survival analysis using data from the Kungsholmen project [22] which is shortly summarized. A parametric survival analysis was applied, because it was used for predictive modelling and extrapolated outside the observed data. Cox regression was limited for this purpose since a baseline hazard is not defined and it only estimates the relative effects of covariates. Dementia free survival was predicted by gender, age at MCI diagnosis, education, MMSE at MCI diagnosis, Katz at MCI diagnosis and all 2-way interactions. A stepwise procedure was applied by removing interactions with highest p-values first until $p<0.05$ followed by the predictors. Only age was significant. The survival function was transformed to enable calculating an individual's time to the event of dementia conversion (see Equation 1). Both the onset of MCI was estimated to have taken place in the middle of each follow-up interval that was on average 3 years. This was operationalized by adding a time correction of 1.5 years to all diagnoses.

$$td_i = \sqrt[2.01]{\frac{\ln(S)}{-e^{-4.06-0.96gender_i}}}$$

Figure 2 represents the observed and predicted dementia-free survival. The observed survival (solid red line) is a stepwise shape caused by the 3-year time interval at which MCI is measured. The observed survival starts at 1.5 years due to the time correction. Although the observed and predicted lines not fully overlap the predicted line seems to interpolate well in between the observed 3-year time intervals.

Figure 2: Observed (solid red) and predicted using a parametric survival model with Weibull distribution (dashed blue) dementia-free survival in MCI
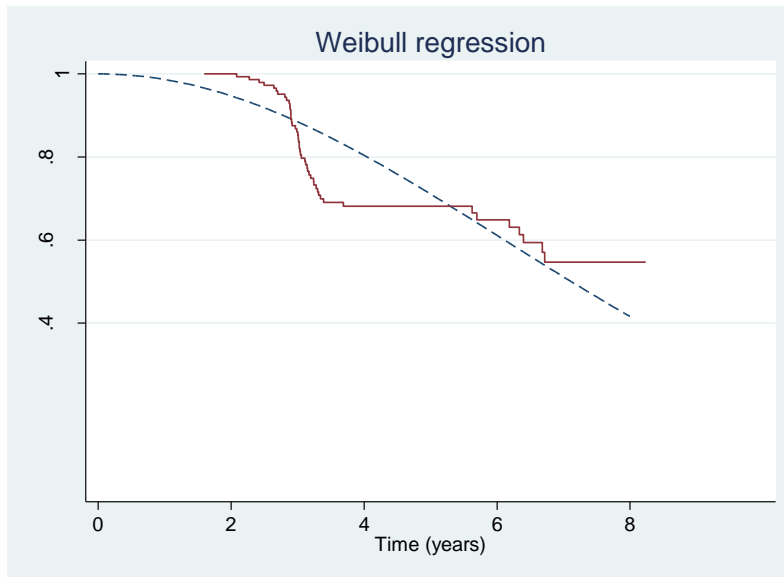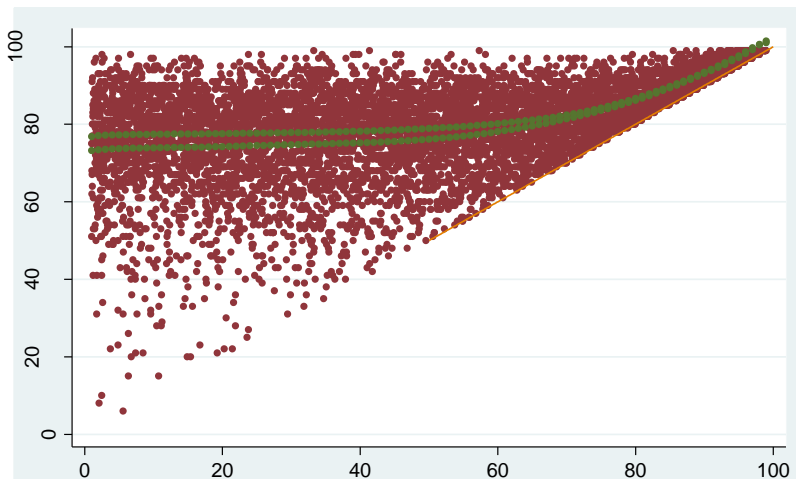

3.4.2   Time to death

Time to death was sampled from a survival table from the Dutch national central bureau for statistics (CSB) (see Table 5). The life table represents the probability to survive up to a specific age from birth using estimates from 2012. Age at death was estimated by taking the probability of surviving up to the sampled age at model entry of an individual with a specific gender. This probability was multiplied with a random value (from a univariate distribution between 0 and 1) and the corresponding age represented the age at death. Age at death was truncated at 99 years since the table does not provide estimated for older subjects. Figure 3 represents the average survival age from CBS data and the survival estimates for 1000 subjects with age sampled from a univariate distribution between 0 to 99 years old.

Table 5: Life table representing the probability to survive up to a specific age from birth using estimates from 2012.

Age	Men	Women	Age	Men	Women	Age	Men	Women
1	0.9965	0.997	34	0.9862	0.9915	67	0.8642	0.8997
2	0.9958	0.9965	35	0.9856	0.9911	68	0.8517	0.8914
3	0.9955	0.9964	36	0.985	0.9908	69	0.838	0.8822
4	0.9954	0.9962	37	0.9844	0.9904	70	0.8235	0.8727
5	0.9953	0.9962	38	0.9837	0.9899	71	0.8071	0.8625
6	0.9951	0.9961	39	0.983	0.9893	72	0.79	0.8507
7	0.995	0.996	40	0.9822	0.9887	73	0.7714	0.8377
8	0.9949	0.9959	41	0.9812	0.9881	74	0.7511	0.8251
9	0.9948	0.9958	42	0.9802	0.9873	75	0.7294	0.8112
10	0.9947	0.9957	43	0.979	0.9864	76	0.7059	0.7949
11	0.9945	0.9957	44	0.9779	0.9855	77	0.6805	0.7767
12	0.9944	0.9956	45	0.9767	0.9845	78	0.6529	0.7575
13	0.9944	0.9956	46	0.9753	0.9834	79	0.6218	0.7358
14	0.9942	0.9954	47	0.9737	0.982	80	0.5883	0.7119
15	0.9941	0.9954	48	0.9719	0.9802	81	0.5531	0.6855
16	0.994	0.9953	49	0.9699	0.9785	82	0.5172	0.6557
17	0.9938	0.9951	50	0.9678	0.9766	83	0.4768	0.6238
18	0.9935	0.995	51	0.9651	0.9746	84	0.4375	0.5896
19	0.9932	0.9948	52	0.9624	0.9723	85	0.3957	0.5517
20	0.9929	0.9947	53	0.9593	0.97	86	0.3527	0.5124
21	0.9926	0.9945	54	0.9558	0.967	87	0.3123	0.4695
22	0.9921	0.9943	55	0.9521	0.964	88	0.2709	0.4244
23	0.9918	0.9941	56	0.9478	0.9608	89	0.2326	0.379
24	0.9913	0.994	57	0.9433	0.9574	90	0.1966	0.3346
25	0.9908	0.9938	58	0.938	0.9534	91	0.1608	0.2884
26	0.9903	0.9936	59	0.9322	0.9488	92	0.1282	0.243
27	0.99	0.9934	60	0.9259	0.9442	93	0.1007	0.1988
28	0.9895	0.9933	61	0.9193	0.939	94	0.077	0.1614
29	0.9891	0.9931	62	0.9121	0.9334	95	0.0563	0.1262
30	0.9886	0.9928	63	0.904	0.928	96	0.0396	0.0969
31	0.9881	0.9926	64	0.8954	0.9216	97	0.0273	0.0718
32	0.9876	0.9922	65	0.8858	0.9152	98	0.019	0.0517
33	0.987	0.9919	66	0.8754	0.9074	99	0.0118	0.0359

Figure 3: Average survival age from CBS data (green) and the survival estimates for 1000 subjects with age sampled from a univariate distribution between 0 to 99 years old (red)

Time to death was not updated at the event of receiving a dementia diagnosis. This would cause lead-time bias [25,26] and therewith generating a longer survival time for subjects who became demented because their updated survival time at conversion takes into account the fact that they survived up to a certain age. Not including a dementia specific survival probability resulted in a small part of the subjects (2%) surviving dementia for more than 30 years which is a limitation in the generalizability to the real-world.

3.4.3 Symptomatic progression in dementia

Symptomatic disease progression in dementia was analysed using data from the Kungsholmen Project [22] and shortly summarized below. Cognition was reflected by the MMSE and ADL by the Katz score. From the literature Mendiondo et al. [27] and Mohs et al. [28] parameterized a model for the annual rate of cognitive decline and found a U-shaped pattern of a low decline rate in mild and severe dementia and a higher decline rate in between. We explored this model to describe the time to a 1-point change in a scale by a piecewise and a quadratic regression model. The results were not significant which could be attributed to the use of the population based sample instead of a clinical sample that probably includes persons with a worse prognosis because the consultation of a medical professional is most likely initiated by the person's memory complaints. For these reasons MMSE and Katz over time were described by 2 separate regression formulas.

A mixed model regression was applied to reflect individual differences and enabled the direct estimation of an individual's course of decline as a random variation to the group average. A stepwise procedure was applied and predictors were included if the goodness of fit statistics -2 Log Likelihood change and Wald z of the predictor were significant. The interaction between time and age (measured at each assessment) significantly predicted the MMSE score. The interaction between age (measured at each assessment) and MMSE score, the interaction between MMSE score and time being diagnosed, and education significantly predicted the Katz.

Equation 2: Regression formula to describe an individual's cognitive decline over time in the dementia phase (i = individual subject; MMSEit = Mini-Mental State Examination of an individual i at time t; t = model time; time = years being demented)

$$MMSE_{it} = \left(26.87 \pm \sqrt{2.0}\, r_i\right) - \left(3.26 \pm \sqrt{1.86}\, r_i\right) time - 0.35(age_{it} - 75) + 0.10 time(age_{it} - 75)$$

Equation 3: Regression formula to describe an individual's functional decline over time in the dementia phase (i = individual subject; Katz indexit = Katz index score of an individual i at time t; t = model time; time = years being demented; education = years of formal education; MMSE = MMSE as a function over time, thus is Equation 2)

$$Katz\ index_{it} = \left(-0.82 \pm \sqrt{0.71}\, r_i\right) + 0.26t + 0.26(age_{it} - 75) + 0.06 education_i - 0.02 MMSE_{it} - 0.01 MMSE_{it}(age_{it} - 75) - 0.01 MMSE_{it} time$$

Using the regression analyses formula average MMSE and Katz were calculated at each time point and compared to observed data in Figure 4. The predicted values (dashed red line) represent the

average of the predicted scores in each individual using the individuals' characteristics of the Kungsholmen Project database.

Subjects in the Kungsholmen project were followed up for a maximum of 6 years from incident dementia diagnosis. Data extrapolated outside this time interval contained values outside the measurement instrument. Furthermore, due to the random slope and the non-linear regression in some cases the progression stopped and cognition or functioning increased. The possibility of improvement was overruled by the last known lowest value to correctly reflect the previously described assumption of the non-reversible neurodegenerative process. Values were truncated to fall within the minimum and maximum scale value of the MMSE (0-30) and Katz (0-6). Furthermore, values were truncated to prevent them from being higher than the sampled MMSE and Katz at MCI since a symptomatic improvement from MCI to dementia was assumed not able to occur.

Figure 4: Average MMSE and Katz over time (observed = average of observed scores in the Kungsholmen Project database; predicted = average of predicted scores in same subject sample as observed)
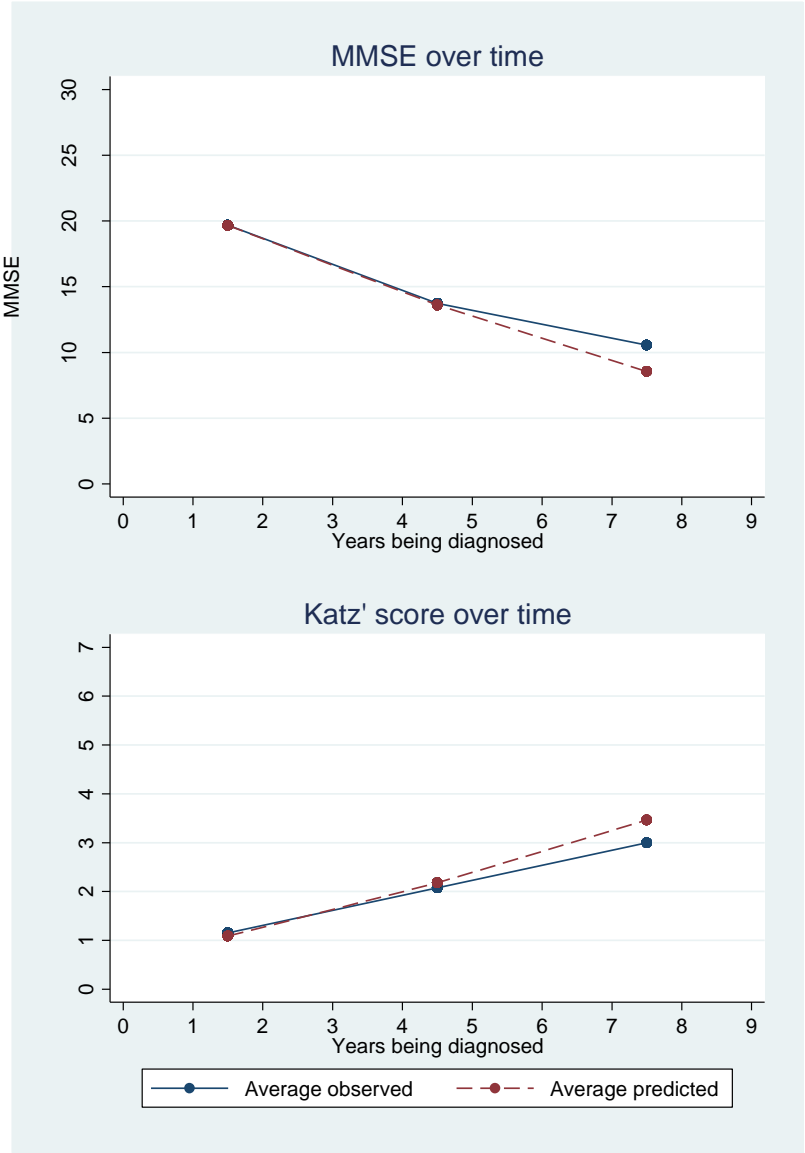
3.5   Treatment effect

A hypothetical treatment with a constant efficacy was applied in the MCI and dementia phase. Treatment effect was assumed to result in a 50% [29] reduction of disease progression. Once treatment was started it was assumed not to stop. The treatment effect was applied as a time increase on the estimated time from MCI to dementia conversion (e.g. an estimated 4 years from MCI to dementia without treatment became 8 years if treatment was provided). Within the dementia phase the treatment effect was applied on the time component of the MMSE and Katz to decrease the effect of time on progression with 50% (e.g. the beta-coefficient of time of -3.26 was decreased with 50% to -1.63). Treatment effect in dementia was assumed to be present in all disease severities (mild, moderate and severe). Treatment side effects were assumed to reduce QOL with 5% (disutility of 0.05) as long as treatment was provided.

When running the deterministic model comparing no treatment with test & treat (using 0.77 sensitivity and 0.68) treatment was cost-effective with a NMB= € 103,981). Treat all as intervention compared to a control strategy of test & treat (sensitivity is 0.77 and specificity is 0.68) was not cost-effective with a NMB= € -12,857.

The effect of DMT on survival was not modelled since it has a complex interaction with cost-effectiveness which has been extensively researched by Sköldunger et al. [29].

3.6   Lawton to Katz using the DAD

Several scales to measure activities of daily living were used in the different data sources. The LeARN study included the Disability Assessment for Dementia [30] (DAD) and in a subsample the Lawton [31], the MEDICIE study included the Lawton and Kungsholmen Project the Katz index scale. Because the disease progression formulas were already established the DAD and Katz scales were transformed.

None of the Lawton items matched the Katz and part of the DAD items matched the Katz. Therefore a DAD based Katz score was calculated using only the DAD items that corresponded to the Katz in the LeARN study. These items were: wash and dry completely (bathing), dress himself/herself completely (dressing), undress himself/herself completely (dressing), decide to use the toilet at appropriate time (toileting), go out and reach a familiar destination (moving), use toilet without accidents (continence), eat his/her meals at a normal pace (eating) [32].

Next the correlation between the Lawton total score (representing iADL, range 0-31) and the Katz surrogate score (that was based on the corresponding DAD items, representing ADL, range 0-6), was assessed using linear regression in the LeARN study. The correlation between the Katz surrogate score and Lawton was 0.42. The betas from the univariate linear regression in which the Katz surrogate total score (using DAD-matching variables only) was predicted by the Lawton total score resulted in a constant of -0.83 and effect of Lawton of 0.12 (p<0.01), see Figure 5. As expected the lower bound of the Lawton was related to an intermediate value on the Kats, indicating that being highly iADL dependent relates to being intermediate ADL dependent. This was expected since iADL is commonly affected earlier than ADL. To transform the Lawton to a Katz-score Equation 4 that was based on the regression results was used. This transformation was applied in the MEDICIE study where only the Lawton was measured. This Lawton-based Katz surrogate score variable was further used in the cost and quality of life assessments of the MEDICIE study.

$$Katzs_i = -0.831 + 0.116 Lawton_i$$

Figure 5: Correlation between Lawton and DAD total score based on Katz-matching variables only
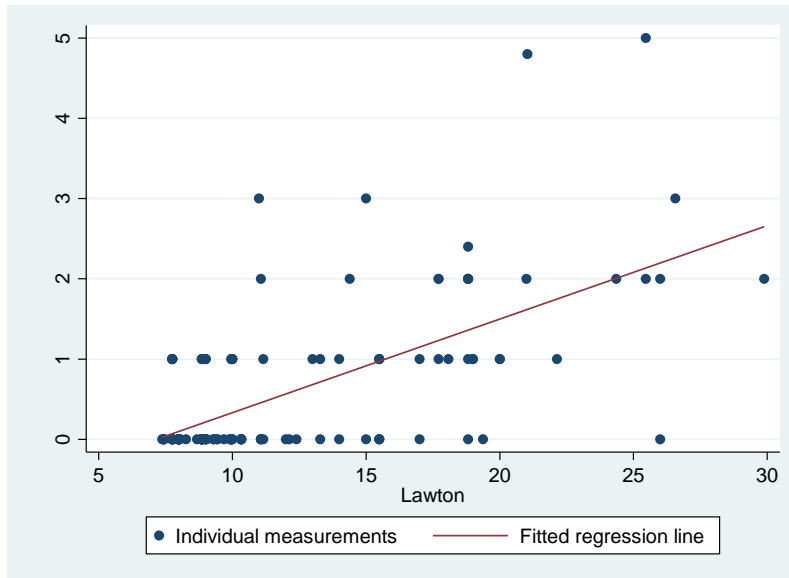
3.7   Costs

Cost data was not available from the Kungsholmen database and the LeARN study insufficiently covered the disease severity spectrum of demented subjects which made generalization to moderate and severe dementia limited. Therefore costs were analysed in the MEDICIE data [23,33] which contained subjects with MCI and dementia with MMSE ranging from 4-29. Costs were calculated based on the results of 201 patient and proxy interviews (patients recruited via primary GP centres were excluded due to limited generalizability) on the 1-year health care resource use and quality of life. The study adopted a societal perspective including standard diagnostic, hospital, medication, admission, day care, home care, general practitioner, informal care utilization and patient and family out of pocket costs. The study measured activities of daily living using the Lawton scale, which was transformed to a Katz surrogate score as explained in paragraph 3.6.

The cost data in the MEDICIE cohort distribution was skewed and by visual evaluation square root transformation improved toward a normal distribution. A generalized linear model was build using a square root link function. Costs were predicted by gender, age, MMSE, a Katz surrogate scale and all 2-way interactions. Education was not available in the database. A stepwise procedure was applied by removing interactions with highest p-values first until $p<0.05$ followed by the predictors. Analyses were performed separately for MCI and dementia subjects.

In the MCI sample only the Katz was significant. In the dementia sample both the MMSE and Katz were significant (see Table 6). Annual costs were predicted by exponentiation of the regression formula outcome in each individual to retransform back from the square root transformation (see equation 1 and 2 in the manuscript). Extreme value analysis retrieved no predicted costs lower than 0.

153

Table 6: Predictors of individual annual formal and informal care costs from a backward generalized linear regression model, separately for MCI and dementia phase

	MCI		Dementia	
Variable	*Beta*	*P*	*Beta*	*P*
Constant	140.16	NA	232.81	NA
MMSE	NA	NA	-2.54	0.013
Katz[1]	62.63	0.000	38.04	0.000

1Katzs = Katz surrogate scale based on the Lawton (see paragraph 3.6)

The total costs of CSF were estimated € 211 based on expert opinion. This estimate included the performance of a lumbar puncture. Furthermore it included the costs for the assay run, the analyses of total Tau, phosphorelated Tau and Amyloid Beta42 protein, reporting and quality check. This was estimated in the situation of a high turn-over. The costs of the diagnostic tests as performed in standard clinical practice (patient medical information, a physical examination, a clinical and neuropsychological examination, and the 3 tesla MRI conclusions on atrophy and white matter lesions) were included in the resource utilization of the MEDICIE study. To prevent double counting and because the costs of current practice tests does not differ between the control and intervention strategy they were not explicitly priced. DMT costs were estimated on € 5,853 per year [29]. All costs were expressed in euros at 2012 values (at that time, €1.00 was equivalent to US $1.29 and British £0.81) using the price index table from the Dutch Central Bureau for Statistics.

3.8 Quality of life years

Similar to the cost analyses, quality of life analyses were performed in the MEDICIE data [23]. The proxy's perception of the patient's health-related quality of life was measured by the EQ-5D. The health utility score was calculated using the Dutch tariff [34].

Linear regression was applied using the predictors gender, age, MMSE, a Katz surrogate scale (see paragraph 3.6) and all 2-way interactions. A stepwise procedure was applied by removing interactions with highest p-values first until p<0.05 followed by the predictors. Analyses were performed separately for MCI and dementia subjects. In the MCI sample only the Katz was significant, in the dementia sample gender, MMSE, Katz and the interaction between gender and MMSE were significant (see Table 7 and see equation 3 and 4 in the manuscript). Extreme value analysis retrieved utility predictions between -0.431 and 0.831 for MCI and between -0.578 and 1.035. Utilities were truncated at maximum of 1.

Table 7: Predictors of individual health-related quality of life utility from a backward linear regression, separately for MCI and dementia phase

	MCI		Dementia	
Variable	*Beta*	*P*	*Beta*	*P*
Constant	0.831	NA	0.653	NA
Gender (=female)	NA	NA	0.382	0.020
MMSE	NA	NA	0.007	0.288
Katz[1]	-0.210	0.000	-0.205	0.000
Gender x MMSE	NA	NA	-0.018	0.026

1Katz = Katz surrogate scale based on the Lawton (see paragraph 3.6)

DMT was assumed to have side effects. Disutility due to DMT side effects was estimated by expert opinion (FV) at a constant 0.05 as long as the treatment was provided.

The disutility of the medical consequences for receiving an incorrect diagnostic test was captured by the disutility of the treatment side effects. The non-medical consequences of an incorrect diagnosis were assumed to be caused by coping with the personal decision that were made based on the wrong information (e.g. unnecessary personal long-term planning). This only accounted for the MCI phase since in the dementia phase such non-medical consequences were assumed not to differ between AD dementia and other NDD dementias. The negative impact of incorrectly labelling a person with AD was considered higher than incorrectly labelling a person as no AD (expert opinion by FV). A correct (either AD-positive or AD-negative diagnosis) was used as point of departure. The disutility of an incorrect AD positive diagnosis was estimated 0.25 for 1 year by expert opinion (FV). The disutility of an AD negative diagnosis was estimated 0.15 for 1 year.

The final health utility estimate was retrieved by adding the estimated utility by regression equation 3 and 4 of the manuscript with the disutility of an incorrect diagnosis and the disutility of treatment. Using the trapezium rule the health utility was multiplied with the corresponding time period to derive quality adjusted life years (QALY).

3.9 Discounting

Costs were discounted at 4% and effects at 1.5% according to the guideline for pharmaco-economic research [35]. Formulas were used for the discounted costs in a specific year (Equation 5) and the total discounted costs stable annual costs (Equation 6).

Equation 5: Formula to discount outcome in a specific year (Od = discounted outcome (total costs or effects); dr = discount rate; t = time (years) between model start and the year in which the On (non-discounted outcome) occurred)

$$O_d = \frac{O_n}{(1 + dr)^t}$$

Equation 6: Formula to discount outcome over a specific period (Od = discounted outcome (total costs or effects) over a specific period in which the On (non-discounted outcome) per year is assumed stable; dr = discount rate; t = time of the period since model start)

$$O_d = \frac{O_n}{dr}\left(1 - \frac{1}{(1 + dr)^t}\right)$$

4 Uncertainty analyses

4.1 Probabilistic sensitivity analysis

The correlations between the parameters within the statistical model were estimated using the variance-covariance matrix for the models to predict: MMSE, Katz, costs, utility, time to dementia, time to death in MCI, and the simulation MCI cohort. The probabilistic sensitivity analysis (PSA) was based on the Cholesky decomposition of these variance-covariance matrices. In brief, the Cholesky decomposition matrix (T) was derived from the variance covariance matrix (V) such that T multiplied by its transpose gives the covariance matrix V (TT' = V). Next the vector of correlated variables (x) was calculated by multiplying the Cholesky decomposition matrix (T) to a vector of independent randomly drawn standard normal values (z) and adding the mean parameter values to this (y) such that x = y + Tz [36].

Furthermore a normal distribution was also applied to mixed model random effects in the MMSE and ADL regression functions. An overview of the parameters subject to PSA and their chosen distributions can be found in table 2 of the manuscript. The number of PSA iterations was 10,000.

Part of the parameters (medication effect, sensitivity and specificity of the base case, treatment costs, intervention costs, disutility due to false positive or false negative, and disutility of treatment side effects) were not subject for the probabilistic sensitivity analysis because no empirical data was available.

4.2 Univariate sensitivity analysis

In univariate sensitivity analyses several fixed values were varied (see Table 8) and the incremental NMB outcomes were compared to the base case (DMT costs=5,855 & DMT efficacy=0.5, AD prevalence=49%, CSF costs=211, disutility FP & FN only in the first MCI year, disutility DMT side effects =0.05).

Table 8: Input parameters of the base-case and the changes in input parameters in the univariate sensitivity analysis.

Scenario	DMT costs	DMT efficacy	AD prevalence	CSF costs	Disutility FP	Disutility FN	Disutility side effects
Base case	€ 5,855	0.5	49%	€ 211	0.25 only in first year	0.15 only in first year	0.05
More efficient treatment	€ 1,171	0.75					
Less efficient treatment	€ 35,119	0.25					
High prevalence			75%				
Low prevalence			25%				
Higher CSF costs				€ 2,112			
Higher disutility non-medical consequences					0.25 during whole MCI phase	0.15 during whole MCI phase	
Higher disutility DMT							0.20

Because the current practice predictive accuracy of DMT response was unknown and the estimates derived from the LeARN study highly uncertain, the relative accuracy change (from current practice sensitivity and specificity to the intervention 'verify AD' or to the 'verify non-AD'

156

sensitivity and specificity) instead of absolute accuracy values was used as input value in the model. The generated results can be applied in any situation of the current practice versus an intervention with accuracy estimates for any biomarker in AD when DMT becomes available. Technically, for the incremental NMB analysis the sensitivity and specificity were set to 75% sensitivity and 75% specificity. Because the incremental NMBs are linearly proportionally related to the sensitivity and specificity of the current practice strategy (e.g. a sensitivity increase from 75% to 85% results in the same incremental proportion of subjects correctly identified as AD as a sensitivity increase from 55% to 65%) the accuracy estimates of the control do not influence the incremental NMB. Using the 75% estimate enabled an incremental sensitivity of the 'verify AD' to range between -25% to 0% (resulting in 75% to the highest possible 100%) and specificity from 0% to 25% (resulting in the lowest possible 50% to 75%), and vice versa for the 'verify non-AD'.

5 Validation

Validation was based on guidance by the ISPOR report on model transparency and validation [37]. However, not all items including cross validity and predictive validity could be performed.

5.1 Face validity

Face validity was performed by several experts. Face validity in terms of model structure and data sources was considered sufficient by Prof. Dr. A. Wimo though remarking the limitations of not including the disease progression of behavioural problems, and the use of multiple data sources from populations with different characteristics.

5.2 Internal validity

5.2.1 Internal validation of input parameters within its observed cohort

The 2 regression formulas for MMSE and Katz and the 2 survival functions were validated against their own source by visual assessment of the graphs by RH (see paragraph 3.4). Furthermore, the observed costs in the MEDICIE study were compared to the predicted costs using the subjects in the database, and similar for utility (see Figure 6). Visual assessment of these graphs by RH indicated valid predictions. Status of validation: November 29th 2013, passed.

Figure 6: Validation of costs and utility regression formulas within cohort by comparing the observed with the predicted values.
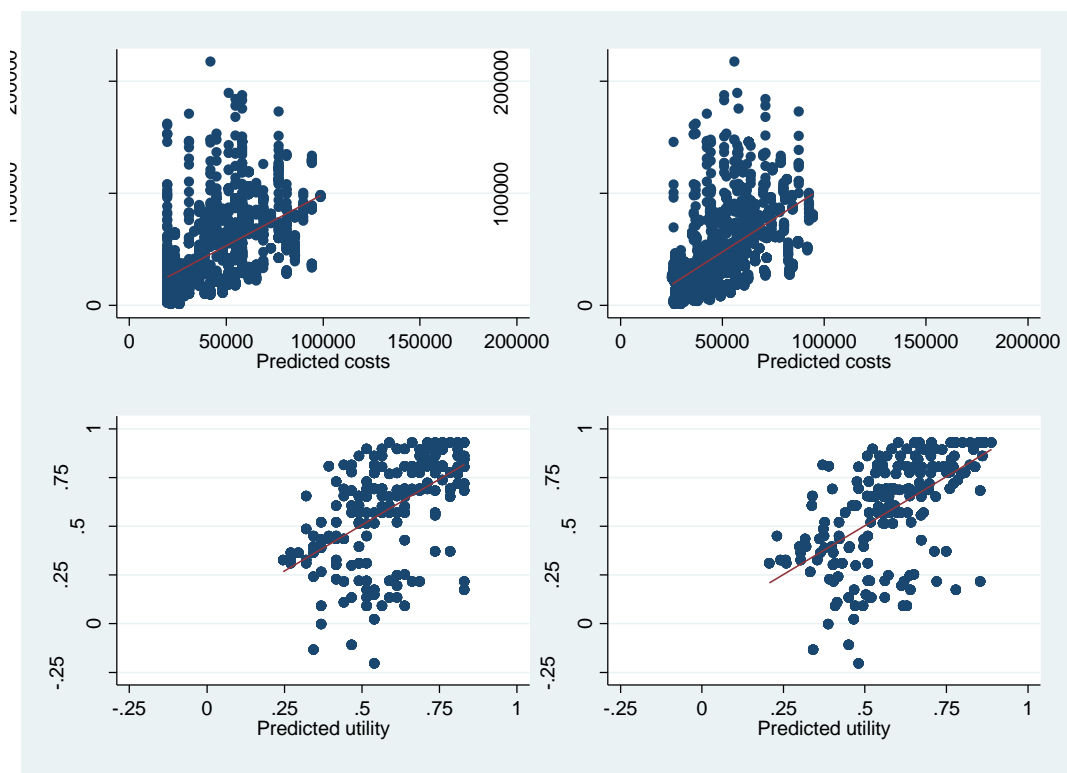
5.2.2 Extreme value testing

Extreme values were tested on diagnostic accuracy parameters of which the effects on other outcomes could be predicted by logic reasoning and observed in the simulation. A sensitivity of 0 and a specificity of 1 (for both the control strategy as for the intervention strategy) resulted in every subject being diagnosed as no-AD. Vice versa (sensitivity 1 and specificity 0) resulted in every subject being diagnosed as AD. Status of validation: October 24th, 2013, passed.

5.2.3 Identical scenarios

Input parameters that define the difference between control and intervention were set identical. These parameters were: sensitivity and specificity in both MCI and dementia and costs of the diagnostic intervention. The deterministic version of the model was run to prevent differences by the PSA random values of parameters. Status of validation: October 24th, 2013, passed.

5.2.4 Structured walkthrough

A complete and up-to-date documentation of the code was provided to an external researcher (AT) for reviewing. Furthermore, the code was explained in detail and a walkthrough meeting was held to check the code for errors. Status of validation: November 7th, 2013, passed.

6   Sources of funding

7   References

[1]   Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:257–62.

[2]   Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280–92.

[3]   Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment 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 2011;7:270–9.

[4]   McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:335–7.

[5]   American Psychiatric Association, editor. DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association; 1994.

[6]   Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–8.

[7]   Petersen RC, Roberts RO, Knopman DS, et al. Mild cognitive impairment: Ten years later. Arch Neurol 2009;66:1447–55.

[8]   Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010;9:119–28.

[9]   Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand 2009;119:252–65.

[10]   Cohen JT, Neumann PJ. Decision analytic models for Alzheimer's disease: state of the art and future directions. Alzheimers Dement 2008;4:212–22.

[11]   Green C, Shearer J, Ritchie CW, Zajicek JP. Model-based economic evaluation in Alzheimer's disease: a review of the methods available to model Alzheimer's disease progression. Value Health 2011;14:621–30.

[12]   Waldemar G, Dubois B, Emre M, Georges J, McKeith IG, Rossor M, et al. Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. Eur J Neurol 2007;14:e1–26.

[13]   Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001;56:1143–53.

[14]   American Psychiatric Association, editor. Diagnostic and Statistical Manual of Mental Disorders: DSM-III-R. Washington DC: American Psychiatric Association; 1987.

[15]   McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS--ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;77:939–44.

[16] Nederlandse Vereniging voor Klinische Geriatrie. Diagnostiek en medicamenteuze behandeling van dementie. 2005.

[17] Sabbagh MN. Drug development for Alzheimer's disease: where are we now and where are we headed? Am J Geriatr Pharmacother 2009;7:167–85.

[18] Cushman L, Scherer MJ. Psychological assessment in medical rehabilitation. 1st ed. Washington, DC: American Psychological Association; 1995.

[19] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of adl: a standardized measure of biological and psychosocial function. JAMA 1963;185:914–9.

[20] Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4. Value Health n.d.;15:821–7.

[21] Handels RL, Aalten P, Wolfs CA, Olde Rikkert M, Scheltens P, Visser PJ, et al. Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease: the research protocol of a prospective cohort study. BMC Neurol 2012;12:72.

[22] Handels RLH, Xu W, Rizzuto D, Caracciolo B, Wang R, Winblad B, et al. Natural progression model of cognition and physical functioning among people with mild cognitive impairment and alzheimer's disease. J Alzheimers Dis 2013;37:357–65.

[23] Wolfs CAG, Dirksen CD, Kessels A, Severens JL, Verhey FRJ. Economic evaluation of an integrated diagnostic approach for psychogeriatric patients: results of a randomized controlled trial. Arch Gen Psychiatry 2009;66:313–23.

[24] Hulstaert F, Blennow K, Ivanoiu A, Schoonderwaldt HC, Riemenschneider M, De Deyn PP, et al. Improved discrimination of AD patients using beta-amyloid(1-42) and tau levels in CSF. Neurology 1999;52:1555–62.

[25] Carrieri MP, Serraino D. Longevity of popes and artists between the 13th and the 19th century. Int J Epidemiol 2005;34:1435–6.

[26] Hanley JA, Carrieri MP, Serraino D. Statistical fallibility and the longevity of popes: William Farr meets Wilhelm Lexis. Int J Epidemiol 2006;35:802–5.

[27] Mendiondo MS, Ashford JW, Kryscio RJ, Schmitt FA. Modelling mini mental state examination changes in Alzheimer's disease. Stat Med 2000;19:1607–16.

[28] Mohs RC, Schmeidler J, Aryan M. Longitudinal studies of cognitive, functional and behavioural change in patients with Alzheimer's disease. Stat Med 2000;19:1401–9.

[29] Sköldunger A, Johnell K, Winblad B, Wimo A. Mortality and treatment costs have a great impact on the cost-effectiveness of disease modifying drugs in Alzheimer's disease. Curr Alzheimer Res 2012:1–10.

[30] Gélinas I, Gauthier L, McIntyre M, Gauthier S. Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. Am J Occup Ther 1999;53:471–81.

[31] Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9:179–86.

[32] Wimo A, Gaudig M, Schäuble B, Jedenius E. The economic impact of galantamine vs placebo: an analysis based on functional capacity in a Swedish cohort study. J Med Econ 2012;15:1019–24.

[33] Handels RLH, Wolfs CAG, Aalten P, Verhey FRJ, Severens JL. Determinants of care costs of patients with dementia or cognitive impairment. Alzheimer Dis Assoc Disord 2013;27:30–6.

[34] Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJ V. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006;15:1121–32.

[35] Hakkaart- van Roijen L, Tan SS, Bouwmans CAM. Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg 2010.

[36] Briggs AH, Claxton K, Sculpher MJ. Decision Modelling for Health Economic Evaluation. Oxford University Press; 2006.

[37] Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value Health n.d.;15:843–50.

8 Abbreviations

'verify AD'	CSF is performed under the condition that AD is expected by standard current clinical diagnostic workup
'verify non-AD'	CSF is performed under the condition that AD is not expected by standard current clinical diagnostic workup
AD	Alzheimer's disease
AD–	No Alzheimer's disease pathology
AD+	Alzheimer's disease pathology
ADL	Activities of daily living
Aβ	Amyloid plaques
CDR	Clinical dementia rating
CSF	Cerebrospinal fluid
DAD	Disability assessment dementia
DES	Discrete event simulation
DMT	Disease modifying treatment
FN	False negative test outcome
FP	False positive test outcome
ICER	Incremental cost-effectiveness ratio
LeARN	LeARN study
MCI	Mild cognitive impairment
MEDICIE	MEDICIE study
MMSE	Mini-mental state examination
MRI	magnetic resonance imaging
NDD	Neurodegenerative disease
NMB	Net monetary benefit
PSA	Probabilistic sensitivity analysis
QALY	quality adjusted life years
RAD–	Result of the diagnostic workup is non-AD
RAD+	Result of the diagnostic workup is AD
Tau	neurofibrillary tangles
TN	True negative test outcome
TP	True positive test outcome

CHAPTER 9

GENERAL DISCUSSION

With the arrival of in vivo identification of AD hallmarks by biomarkers in MRI, CSF and PET, their value in clinical practice is both welcomed and questioned [1–4]. The balance of patient-important outcomes, such as reducing uncertainty, planning support, anxiety, labelling and stigma, and predicting treatment response is unknown. Improved patient health as a result of an AD biomarker can be evaluated by comparing the medical and non-medical consequences of a biomarker-driven diagnostic pathway to its best alternative diagnostic pathway [5,6]. This is of major importance due to the increasing use of biomarkers, and CSF in particular [7], in clinical practice.

We aimed to evaluate the prognostic accuracy of CSF and the cost-utility of CSF in the context of diagnostic workup and treatment in cognitive disorders. Firstly, a consensus protocol to establish a reproducible independent reference standard for a prognosis was established. This made it possible to evaluate of the prognostic value of the CSF Amyloid-Beta Tau ratio from a clinical perspective. Secondly, a decision-analytic model was built using the recommendations generated by the systematic review of health economic evidence, the formulas that describe disease progression in a population of MCI and demented subjects, and the resource use in subjects with cognitive disorders. The model evaluated the selective use of CSF applied to decide on a hypothetical disease-modifying treatment in MCI subjects.

SUMMARY OF FINDINGS

The LeARN study was described in chapter 2, in which a cohort of memory clinic patients suspected of having a primary neurodegenerative disease was followed for two years and a decision model was built to synthesise the study results with other evidence to estimate the long-term outcomes of using CSF.

To determine the diagnostic accuracy of CSF, a consensus procedure to set a reference diagnosis was piloted in chapter 3. The data from a limited number of patients (n=11) with 2-year follow up were assessed by 3 clinical experts in a panel discussion. The consensus procedure was considered to be feasible for research purposes. Several adaptations were made, among which the inclusion of a summary of the patient and informal caregiver history to increase the extent to which the consensus procedure reflects clinical practice.

In chapter 4 the consensus procedure was applied among the participants in the LeARN study (n=111) to compare a diagnostic workup to predict functional decline according to current practice with a workup to predict functional decline combining current practice and CSF. The CSF ratio of Amyloid-Beta and Tau was used to verify clinically expected decline due to AD or to disprove no expected decline. The results of CSF led to 25 out of 111 subjects being reclassified, of whom 6 correctly (i.e. clinicians expected decline while CSF correctly changed their conclusion to no decline as no decline was observed after 2 years, or vice versa) and 19 incorrectly (i.e. clinicians expected decline while CSF incorrectly changed their conclusion to no decline as decline was observed). Subgroup analysis did not reveal an added value of CSF in any of the syndromes SMC, MCI or dementia.

The added value of selective use of CSF to current clinical practice was determined either after a positive or after a negative diagnosis of the current clinical practice. Therefore a decision-analytic model was developed, populated and analysed to estimate the cost-utility of a CSF biomarker in

combination with hypothetical disease-modifying treatment compared to current practice diagnostic workup in combination with a hypothetical disease-modifying treatment.

The methodological characteristics of current economic evaluations of diagnostic interventions for AD were reviewed in chapter 5 to help develop the decision-analytic model. The quality of the studies was considerable, but various aspects were insufficiently addressed. CSF was not evaluated and some of the models focused on a timely diagnosis of the dementia syndrome. This deviated from our thesis objective of an early diagnosis and prognosis in MCI. Furthermore, the possible positions of a biomarker in future scenarios with available disease-modifying treatment were not covered fully by the current studies. Finally, some recommendations on including non-medical consequences of diagnostic tests have been described in recent research that could improve the current decision models [6,8–10]. These findings aided in the development of a new decision model to accomplish the thesis aim.

To populate the model, disease progression and resource use related to disease severity state were analysed. In chapter 6 the data for 476 persons with incident MCI or dementia from Kungsholmen Project were summarised in mathematical regression functions that describe disease progression. Thirty-one per cent of the MCI subjects developed AD dementia after a median time of 3.03 years. Women had a significantly shorter time for MCI to AD type dementia conversion than men. Disease progression in the dementia phase was described by the change in cognition and functioning over time. On average, subjects decreased 1.84 MMSE points per year, which worsened with higher age, and increased 0.38 Katz points per year, which also worsened with higher age in combination with lower MMSE and higher education.

The analyses in chapter 7 of care resource used the data for 219 subjects from the MEDICIE study on the cost-effectiveness of an integrated multidisciplinary diagnostic facility for diagnosing dementia [11]. It revealed the costs associated with a 1-point change of cognition and functioning and associated with other changes in patient and informal caregiver characteristics due to health care resource usage. MMSE and informal caregiver's burden, measured by the "Self-Perceived Pressure from Informal Care" instrument, were significant predictors of the 1-year care costs. Patient gender and instrumental activities of daily living partially explained the 1-year care costs.

The decision model in chapter 8 has aggregated the mathematical formulas from the analyses of the previous 2 chapters and combined them with additional analyses on health related quality of life and diagnostic accuracy. The headroom analysis revealed 0.39 gained QALYs and potential savings of €33,622 per subject if a perfect test existed for predicting disease-modifying treatment response. Furthermore, the CSF biomarker was positioned as an add-on when the current practice diagnostic workup concluded that AD was present. CSF was also positioned as an add-on when the current practice concluded that AD was absent. In the second position (after absence of AD), CSF was cost-effective in more possible combinations of sensitivity and specificity than in the first position (after presence of AD)

The results generated by the studies in this thesis did not indicate an added value of the CSF Beta Amyloid total tau ratio information in current clinical practice by a decision rule for the 2-year prediction of cognitive or functional decline in subjects with SMC, MCI or dementia. It was found that CSF is likely to be cost-effective for the selection of subjects eligible for treatment when future disease-modifying treatment becomes available. Given the assumptions regarding disease-modifying treatment, the potential benefits of preventing undertreatment of AD patients were higher than the potential benefits from preventing overtreatment of healthy subjects, because

disease-modifying treatment postponed dementia conversion and the related need for care resources. CSF was therefore best positioned to reveal subjects who suffer from AD pathology but are overlooked in current practice. This suggests aiming at increasing the diagnostic workup's sensitivity which could be done by using a CSF test to verify the non-AD diagnoses set by clinical practice.

METHODOLOGICAL CONSIDERATIONS

Reflection of current clinical practice

Our studies intended to reflect the current clinical practice of specialised centres for the diagnosis of cognitive disorders. Therefore, several methodological characteristics were adopted to maximise generalizability to this setting. This included using all the diagnostic information available in clinical practice to represent the current practice diagnostic workup. Furthermore, this diagnostic information was interpreted by clinical experts who made the diagnosis. Finally, the CSF Amyloid-Beta Tau ratio was explicitly positioned as an add-on test to the current clinical practice diagnostic workup. The methodological design in this study improved on previous studies, which mainly represented current practice by a single test and did not rely on expert clinical interpretations of diagnostic test information [12–18]. The increasing AD-related use of CSF in practice [7] emphasises the importance of assessing CSF's performance when exposed to clinical practice circumstances before adoption of CSF in the clinical routine is justified.

Non-medical consequences

A methodological limitation in chapter 4 is that the balance of a diagnostic test's non-medical consequences that go beyond diagnostic test accuracy was not taking into account [5] (e.g. planning support and future life were not balanced against anxiety and stigma [19]). Improved intermediate outcomes of test accuracy were implicitly assumed to be valuable for the patient. Furthermore, the consequences of an incorrect diagnosis of expected decline due to AD could cause more harm than the consequences of an incorrect diagnosis of no expected decline. CSF could be dedicated to correct the false diagnosis with the highest impact at the expense of not correcting the false diagnosis with the lowest impact (e.g. if a false positive is worse than a false negative, CSF could be dedicated to prevent a false positive at the expense of correcting false negative diagnoses). Although this could result in a worse numerical balance of overall correct diagnoses, it could decrease the overall negative impact of the non-medical consequences in a group of patients.

In the cost-utility study (chapter 8) the impact of non-medical consequences of an early diagnosis on quality of life were estimated by expert opinion as a disutility due to a false positive or false negative diagnosis of MCI due to AD. While the level of evidence of this type of source is considered to be low, it was assumed to more accurately estimate the consequences than neglecting these effects. Taking into account the non-medical consequences of diagnoses, even if the true value is unknown, complied with the recommendations based on a literature review by Fang et al. [10].

Follow-up period

The reference diagnosis in this thesis was based on the observation of decline within a period of 2 years. The value of such prognosis may be limited in practice because it has an 'expiration date' of 2 years for a patient. A patient who is reassured that they will not decline will probably return to uncertainty after 2 years, and will require retesting to unravel the prognosis for the succeeding time period. A longer follow-up period, for example of 10 years, seems more relevant. However, this also entails uncertainty because it can occur at any moment within this long period. This emphasises the importance of estimating the timing of an individual's expected decline [3], which was not addressed in this thesis.

Comparability of the populations studied

Evidence from the populations in the LeARN, MEDICIE and Kungsholmen Project studies were combined in the decision model in chapter 8. These populations are not fully comparable and this might have limited the external validity of the model [20]. Subjects from the LeARN and MEDICIE studies were recruited in secondary care centres. These subjects experienced a sign or symptom that prompted them to visit a primary care centre that triaged them to a specialised centre. This was opposite to the Kungsholmen Project study in which all inhabitants aged 75+ of the Kungsholmen district in Stockholm were invited to participate in the study, regardless of whether they suffered from complaints. The two populations reflect a different spectrum of AD. The majority of the general population (reflected by the Kungsholmen Project) would probably not have sought medical attention (e.g. due to less aggressive AD progression that enables them to better cope with the disease and manage AD up to a more severe state at home or due to well-organised family support). Even if they had sought medical attention, they would not all have been referred by their GP to a specialist centre. Disease progression in AD subjects from a memory clinic population, such as was estimated by Mendiondo et al. [21], is therefore faster than estimated by the Kungsholmen Project formulas. Disease-modifying treatment might act better in a clinical population and improve cost-utility there. However, when disease-modifying treatment becomes available it is likely that subjects from the general population will more easily be convinced to seek medical attention for mild memory complaints. In that case, the diagnostic accuracy study as performed in a memory clinic population does not fit the disease progression estimates in the general population study. Most likely it overestimates the diagnostic accuracy of biomarkers in the general population.

In this early technology assessment, these limitations of the external validity are unlikely to threaten the value of the results considering the uncertainty around the effect of disease-modifying treatment and the extensive uncertainty analysis of the diagnostic accuracy. However, when disease-modifying treatment becomes available, it will be necessary to match the disease spectrum of the populations from which diagnostic accuracy and disease progression are retrieved to ensure external validity of the decision model.

The gold standard paradox

The gold standard for AD is often defined in terms of neuropathology or (conversion to) the clinical expression of AD-type dementia. Each reflects a different concept, and there is no one-to-

one relationship between them [22]. The choice of which is the actual gold standard, however, is a paradox. Medical tests have a variety of applications among which risk factor screening, diagnosis, prognosis, decide upon therapy, or response monitoring. The best reference standard for AD mainly depends on the purpose of the medical test and this can vary among studies. A study on the validity of a biomarker to determine pathology is thus best evaluated by neuropathology as it aims to expose the in vitro molecular hallmarks of AD in vivo. A study on the prognostic value of a biomarker is best evaluated by the course of the clinical phenotype of AD, because it aims to predict a patient's future health state. Such research can for example reveal the added value of a biomarker in clinical practice when assuming that a prognosis can help a patient to create a representation of the disease and the symptoms that can be expected and the timing of care demands. A patient is probably not interested in the neuropathology in his/her brain per se, and a biomarker at an early stage of development is not necessarily best evaluated by the development of clinical symptoms over time. Therefore, no single gold standard exists because it would require the assumption that a biomarker that can perfectly predict neuropathology is fully correlated to clinical expression or vice versa, which have been proven false [22].

IMPLICATIONS FOR PRACTICE AND POLICY

Current practice

The newly developed diagnostic research criteria [3] emphasise that biomarkers must predict individual progression in a specific time period before they can be adopted in the core clinical criteria for MCI in the clinical routine. The results in chapter 4 suggest that there was no improvement in progression prediction when CSF was added to the current practice diagnostic workup by means of a decision rule in subjects with an average age of 67. The adoption of the CSF Amyloid Beta Tau ratio in the current clinical routine for short-term prognosis in a population similar to our research sample can therefore not be recommended at present.

Other applications of CSF, for example to test for inflammation or malicious cells or prognoses in other age groups, were not evaluated in this thesis and the thesis results have no implications for these applications.

Future clinical practice

Several plausible scenarios of applying CSF in MCI patients in future clinical practice, in the event that a disease-modifying treatment that acts upon the pathological mechanism as reflected by CSF Amyloid Beta and Tau becomes available, were found to be beneficial in this thesis. The results of the decision model in chapter 8 showed that the relatively inexpensive CSF test (estimated at € 211) improves the efficiency of the allocation of resources required for disease-modifying treatment and dementia-related care. When CSF was positioned to verify the expectation of no AD by the current practice to predict treatment response, it was beneficial over a large range of plausible values of sensitivity and specificity. Furthermore, the model showed the complex interaction between the test and its clinical climate. For example, if the AD prevalence among MCI subjects who visit the memory clinic decreases, the model indicated that the CSF test should move its position from verifying a clinical expectation that AD is not the cause of the

complaints to verifying a clinically expected AD in order to be efficient in the balance between resource allocation and quality of life. These results support the design of the future diagnostic process.

Current and future research practice

The results of the early health technology assessment's cost-utility analysis have several implications for current research practice. They suggest increasing the sensitivity of current or newly developed CSF tests. This might be done in various ways, for example by changing the cut-off value or by developing a more sensitive biomedical lab procedure. The impact of the CSF's position in the diagnostic pathway on the results indicated the relevance of aiming at a specific balance of sensitivity and specificity. This demotes the use of the Youden index (deriving the cut-off value at which the combined sensitivity and specificity are maximised regardless of their specific combination of values). The limitation of the Youden index is that it does not take the consequences of a correct positive diagnosis versus correct negative diagnosis from a patient or health economic perspective into account.

The early health technology assessment also attempts to look into the future and sketch the possible impact of disease-modifying treatment for AD in its interaction with the diagnostic process. The sensitivity analyses provide insight when conditions such as prevalence change in practice. Low prevalence stresses the importance of a test's specificity to prevent a loss of resources by overtreating a large number of non-AD patients.

Several research fields fell outside the scope of this thesis for which the results have no intended implications. No implications were intended for the role of CSF in research for unravelling the AD mechanism, disease-modifying treatment development, or the role of biomarkers in supporting subject selection for AD treatment trials.

Consequences of the Alzheimer's label

The analyses within this thesis exposed the lack of data on the consequences of labelling MCI subjects as AD, regardless of whether this label was produced by the current practice diagnostic workup or by an AD biomarker. Early diagnosis of AD could reduce uncertainty and provide an opportunity to plan support and future life. On the other hand, the stigma around the AD label could provoke anxiety or a depressive reaction [19]. Furthermore, extensive evaluation within a memory clinic might be experienced as emotionally burdening by some patients, while others may experience harm from not receiving an early diagnosis to reduce their anxiety due to uncertainty. The various emotional reactions to early assessment and diagnosis are likely to depend on personal factors. A shared decision making environment has been proposed to perform careful counselling before performing an early diagnosis. This provides the opportunity to weight the expected benefits and harms [23]. In such an approach a patient can, supported by a clinical expert, consider the personal consequences to them of a possible positive test result, negative test result, or no further testing. Such a patient-centred approach takes into account patient preferences and may provide an alternative to a single standardised diagnostic workup offer. However, these hypotheses have not been tested. A diagnostic test is no exception to other

medical technologies and its benefit to health outcome must therefore be proven benefit before adoption in practice is appropriate.

FUTURE DIRECTIONS

An AD biomarker should prove effective and efficient in terms of patient-important outcomes and care resource use compared to the standard diagnosis workup before being disseminated into the clinical routine. This was already emphasised by Fineberg et al. [24] in 1978 by stating "Diagnosis is not an end in itself. In general, medicine is directed toward the goal of improved health outcome." Up to now, insufficient data from systematic evaluations have been available on improved health outcomes of AD biomarkers to eliminate the uncertainty that impedes decision makers from deciding on whether dissemination into practice is appropriate. Future research to generate more evidence that enables this decision could take place in 2 different scenarios: one without disease-modifying treatment and one with disease-modifying treatment available to AD patients.

Scenario 1: No cure available in the near future

In the first scenario, which also represents the current situation, only the non-medical consequences of using a biomarker can affect patient-important outcomes of the various possible diagnostic test outcomes: a patient who received verification of AD, assurance of no AD, staying in uncertainty of not knowing their AD probability, or receiving one of these diagnoses incorrectly. Although these non-medical consequences, such as anxiety, relief or stigmatisation, can be measured by specific instruments, their translation into health economic outcomes comes with limitations [6,9]. For example, Neumann et al. [8] estimated the willingness to pay for an imperfect AD test which was valued at an average of $428. This represents the value of information from a diagnostic test in the absence of treatment. However, estimates of willingness to pay are subject to incentives to misrepresent preferences and to difficulties in placing a price value on risks in life [9]. Alternatively, non-medical consequences could be measured with quality of life scales. However, these rarely capture all the possible test consequences such as stigma or reassurance [6]. These issues are struggled with by other disease research fields as well [10] and call for generic methodological developments. Recently-developed quality of life scales such as the ICECAP-O [25] and research on the value of information and patient preferences are important steps towards quantifying patient-important outcomes for health technology assessment. Validation studies should be performed to evaluate whether such scales are sensitive enough to detect the subtle changes in non-medical consequences.

The study design for new research that applies a quality of life scale to quantify the non-medical consequences would ideally be randomised to generate the highest level of evidence. However, as the results of this thesis indicate, the optimal position in the clinical pathway is subject to several uncertain factors. This requires the comparison of many possibilities, each requiring scarce research resources when evaluated in a trial. As an alternative, the non-medical consequences can be measured and averaged by diagnostic conclusion (correctly receiving a verification of MCI due to AD, correct reassurance of MCI not due to AD, staying in uncertainty of not knowing the probability of AD, or receiving one of these diagnoses incorrectly) in a cohort of

MCI patients in which everyone receives the test or tests under evaluation and a reference standard. With the consequences measured for each of these test outcomes (and not per test), alternative positions of CSF that lead to different distributions of the diagnostic conclusions can be evaluated retrospectively without having to observe a new cohort over time. Furthermore, various combinations with related biomarkers in MRI and PET or sequences of biomarker combinations can be tested to find the optimal balance in diagnostic outcomes that maximises patient benefits.

A cohort design is however prone to confounding. Quality of life assessed over time is subject to influences aside from the impact of a specific diagnostic outcome such as disease progression. Subjects with an aggressive form of AD will decline more quickly and have a higher probability of being diagnosed correctly since the disease is more evident. Because disease progression is correlated to decreased quality of life [26], it confounds the relationship between a correct diagnosis and quality of life by suggesting that a correct diagnosis decreases quality of life. Besides, the test procedure itself can have an impact on a patient such as a lumbar puncture headache, radiation from a PET tracer or claustrophobia in an MRI scan. Furthermore, knowledge of the type of biomarker could affect downstream treatment management irrespective of the link between test outcome and treatment. For example a clinician might be reassured more by an advanced AD biomarker than a neuropsychological profile. It is difficult to blind for these factors or to develop a placebo that is ethically justified. Some solutions might be helpful such as cloaking the type of test to the clinician by presenting standardised test results.

Scenario 2: Disease-modifying treatment is available in the near future

In the second scenario, an AD biomarker will have both non-medical consequences as well as medical consequences due to disease-modifying treatment. In this scenario the medical consequences will probably overshadow non-medical consequences and a biomarker mainly becomes a means to manage treatment instead of establishing a diagnosis. Clinical trials with biomarkers in combination with disease-modifying treatment are much more resource-intensive than diagnostic studies on biomarkers alone. This requires a smart research design to evaluate the test-treatment pathway. Similar to the cohort design to reveal the non-medical consequences as described in the previous paragraph, the test can be disconnected from the test outcome. However, in contrast to evaluating the consequence of the diagnostic label, a new experimental treatment can be randomised without ethical issues. In such case, the treatment can be randomised in all subjects of a cohort regardless of the biomarker result to measure effect differences in an analyses stratified by biomarker outcome [27]. In the event that the treatment's status has evolved from experimental to established, it can be randomised only in those in subjects whose clinical practice test result and biomarker result are discordant. Only in this group could a biomarker initiate a change in treatment management [28]. The latter study design can also be applied if multiple biomarkers are under evaluation. All subjects receive all biomarkers under evaluation, and a researcher can retrospectively apply any strategy of test combinations and evaluate the difference in treated and non-treated subjects [27]. This reduces the possibility of bias and at the same time maximises the knowledge generated with a high level of evidence without individual studies for each possible test combination.

Long-term health economic consequences

Estimates from trials can reveal the short-term impact of medical and non-medical consequences of a biomarker on patient-important outcomes. Long-term estimates are important for the evaluation of a biomarker from a societal perspective to enable societal reimbursement decisions. The current decision models to estimate these effects are of considerable quality as concluded from the systematic review (chapter 5). Furthermore, future model developments no longer suffer from technical restraints with the availability of advanced modelling methods such as discrete event simulation which enable more subtlety in the timing of events. On the other hand, such models are restricted by the limited availability of data on predictors of, for example, the initiation of professional care, institutionalisation and survival. These events can have a major impact on the cost-effectiveness of a (diagnostic) intervention, as has been shown by other research [29]. This urges the need for estimates that reveal the complex interaction between informal care, the patient's disease characteristics and professional (diagnostic) care and their impact on quality of life and resource use. Registries could provide such data and this is probably more efficient than setting up dedicated trials.

Conclusion

The results of this thesis show that the routine use of a CSF diagnostic decision rule in current practice is expected not to improve the prognosis of memory clinic patients similar to the LeARN participants. If in future practice a disease-modifying treatment becomes available, selective use of CSF or related biomarkers has the potential to be cost-effective in treatment response prediction. Diagnostic research on CSF and related biomarkers for AD must prove improved patient-important outcomes before it can be decided to adopt CSF and related biomarkers in current clinical practice.

REFERENCES

[1]   Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:257–62.

[2]   Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280–92.

[3]   Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment 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 2011;7:270–9.

[4]   McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:335–7.

[5]   Bossuyt PMM, Reitsma JB, Linnet K, Moons KGM. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem 2012;58:1636–43.

[6]   Bossuyt PMM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Med Decis Making 2009;29:E30–8.

[7]   Spies PE, Slats D, Ramakers I, Verhey FRJ, Olde Rikkert MGM. Experiences with cerebrospinal fluid analysis in Dutch memory clinics. Eur J Neurol 2011;18:1014–6.

[8]   Neumann PJ, Cohen JT, Hammitt JK, Concannon TW, Auerbach HR, Fang C, et al. Willingness-to-pay for predictive tests with no immediate treatment implications: a survey of US residents. Health Econ 2012;21:238–51.

[9]   Lee DW, Neumann PJ, Rizzo JA. Understanding the medical and nonmedical value of diagnostic testing. Value Health 2010;13:310–4.

[10]   Fang C, Otero HJ, Greenberg D, Neumann PJ. Cost-utility analyses of diagnostic laboratory tests: a systematic review. Value Health 2011;14:1010–8.

[11]   Wolfs CAG, Dirksen CD, Kessels A, Severens JL, Verhey FRJ. Economic evaluation of an integrated diagnostic approach for psychogeriatric patients: results of a randomized controlled trial. Arch Gen Psychiatry 2009;66:313–23.

[12]   Van Rossum IA, Vos S, Handels R, Visser PJ. Biomarkers as predictors for conversion from mild cognitive impairment to Alzheimer-type dementia: implications for trial design. J Alzheimers Dis 2010;20:881–91.

[13]   Vos S, Van Rossum I, Burns L, Knol D, Scheltens P, Soininen H, et al. Test sequence of CSF and MRI biomarkers for prediction of AD in subjects with MCI. Neurobiol Aging 2012.

[14]   Buerger K, Teipel SJ, Zinkowski R, Blennow K, Arai H, Engel R, et al. CSF tau protein phosphorylated at threonine 231 correlates with cognitive decline in MCI subjects. Neurology 2002;59:627–9.

[15]   Palmqvist S, Hertze J, Minthon L, Wattmo C, Zetterberg H, Blennow K, et al. Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up study. PLoS One 2012;7:e38639.

[16]   Richard E, Schmand BA, Eikelenboom P, Van Gool WA. MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer's disease in patients with mild cognitive impairment: a diagnostic accuracy study. BMJ Open 2013;3.

[17]   Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR, et al. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. Neurobiol Aging 2012;33:1203–14.

[18]   Heister D, Brewer JB, Magda S, Blennow K, McEvoy LK. Predicting MCI outcome with clinically available MRI and CSF biomarkers. Neurology 2011;77:1619–28.

[19]   Iliffe S, Manthorpe J, Eden A. Sooner or later? Issues in the early diagnosis of dementia in general practice: a qualitative study. Fam Pract 2003;20:376–81.

[20]   O'Brien B. Economic evaluation of pharmaceuticals. Frankenstein's monster or vampire of trials? Med Care 1996;34:DS99–108.

[21]   Mendiondo MS, Ashford JW, Kryscio RJ, Schmitt FA. Modelling mini mental state examination changes in Alzheimer's disease. Stat Med 2000;19:1607–16.

[22]   Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001;56:1143–53.

[23]   Visser PJ, Wolf H, Frisoni G, Gertz H-J. Disclosure of Alzheimer's disease biomarker status in subjects with mild cognitive impairment. Biomark Med 2012;6:365–8.

[24]   Fineberg H V. Evaluation of computed tomography: achievement and challenge. AJR Am J Roentgenol 1978;131:1–4.

[25]   Flynn TN, Chan P, Coast J, Peters TJ. Assessing quality of life among British older people using the ICEPOP CAPability (ICECAP-O) measure. Appl Health Econ Health Policy 2011;9:317–29.

[26]   Jönsson L, Andreasen N, Kilander L, Soininen H, Waldemar G, Nygaard H, et al. Patient- and proxy-reported utility in Alzheimer disease using the EuroQoL. Alzheimer Dis Assoc Disord 2006;20:49–55.

[27]   Lijmer JG, Bossuyt PMM. Various randomized designs can be used to evaluate medical tests. J Clin Epidemiol 2009;62:364–73.

[28] Lord SJ, Irwig L, Bossuyt PMM. Using the principles of randomized controlled trial design to guide test evaluation. Med Decis Making 2009;29:E1–E12.

[29] Sköldunger A, Johnell K, Winblad B, Wimo A. Mortality and treatment costs have a great impact on the cost-effectiveness of disease modifying drugs in Alzheimer's disease. Curr Alzheimer Res 2012:1–10.

There has been a growing interest in identifying Alzheimer's disease (AD) pathology using medical tests in the cerebrospinal fluid (CSF), Positron Emission Tomography and Magnetic Resonance Imaging in patients who suffer from mild cognitive impairment (MCI) or dementia. Recently developed research diagnostic criteria hypothesise that these medical tests have an added value in clinical practice in predicting cognitive and functional decline, or in deciding on therapies when effective treatments become available in MCI patients. Despite the research status of these biomarkers they are increasingly being applied in clinical practice, especially CSF. It is therefore important to evaluate the added value of CSF for the diagnosis of AD as well as the balance between the impact on the society's health and the costs of the required care resource when employing CSF in the current clinical practice.

The general aim of this thesis was to evaluate the prognostic accuracy of CSF and the cost-utility of CSF in the context of diagnostic workup and treatment for cognitive disorders. In part I the methodology to determine the added value of a medical test was explored and subsequently applied to estimate the added value of CSF in clinical practice. In part II several studies were performed of which the results were used to build a decision analytic model. The model was used to simulate MCI subjects to determine the cost-utility of CSF.

In **chapter 2** the LeARN study was described, including the study design, methodological considerations and choices of measurement instruments to enable the studies of prognostic accuracy and cost-utility of CSF. In the LeARN study 241 patients suspected of having a primary neurodegenerative disease were approached in four Dutch academic memory clinics and followed for two years. Clinical data and data on quality of life, care use and emerging medical tests were measured.

In **chapter 3** a protocol was developed for an expert panel consensus diagnosis. The protocol enabled to determine both the usual clinical practice as well as a reference diagnosis, in absence of a gold standard for AD. An adjusted Delphi method was used in which 3 clinical experts arrived at a consensus reference diagnosis for 11 patients in a pilot study. The results showed that this protocol was feasible in research practice.

In **chapter 4** the added diagnostic value of CSF to a clinical judgment for the prediction of cognitive or functional decline was determined by applying the developed protocol from chapter 3. The diagnosis reflecting current clinical practice without CSF was compared to a decision rule in which the CSF test reclassified the clinical diagnoses. Correctness of reclassifications was determined using the reference diagnosis based on the expert panel interpretation of the 2-year follow-up data. Adding the CSF Beta Amyloid1-42 total tau ratio information to the clinical diagnosis did not improve the prediction of 2-year cognitive or functional decline based in memory clinic patients with subjective memory complaints, MCI or dementia.

In **chapter 5** the literature was systematically reviewed for research on economic evaluations of interventions for the early diagnosis of AD. The general and methodological characteristics were described. Eight decision-analytic modelling studies and one trial-based economic evaluation were found and showed considerable diversity among the study objective and (methodological) characteristics. Recommendations were focused on the diagnostic aspects and the applicability of existing models for the evaluation of recently revised diagnostic research criteria for AD.

In **chapter 6** AD-free survival time in people with MCI and decline of cognitive and physical function in persons with AD was estimated. Within the Swedish Kungsholmen Project 153 persons with MCI and 323 with AD, who were not diagnosed before, were identified using a screening procedure that applied international criteria in a general population of cognitively healthy elderly people aged ≥75. The median AD-free survival time in MCI was 3.5 years. In persons with dementia cognitive decline was 1.84 points on the Mini-Mental State Examination per year and physical functioning declined by 0.38 Katz points per year. Age and level of education were associated with speed of decline.

In **chapter 7** factors that determined health care resource use costs and patient and family care spending were explored in a cost of illness study. The average 1-year health care sector costs were €26,140 and €11,931 for patient and family, including the costs of informal care. The analyses indicated that cognitive functioning, caregiver burden, patient gender, and instrumental activities of daily living were significantly associated with care costs.

In **chapter 8** the cost-utility of routine use of CSF in clinical practice was determined. This was done by simulating individuals over various diagnostic strategies and comparing their average quality adjusted life years (QALY) and care resource spending from a societal perspective. A decision analytic model was built using mathematical equations generated from the results of chapter 4, 5, 6 and 7. The model compared the current clinical practice to 3 intervention strategies in which CSF was positioned after receiving a clinical diagnosis or as a perfect test. In all strategies it was assumed that a future hypothetical disease-modifying treatment (DMT) was available that slows AD progression by 50%. A perfect CSF test resulted in 0.39 (0.26 to 0.54) QALYs gained and € 33,622 (21,232 to 50,780) savings per individual and represented the room for improvement from the current clinical practice. The other analyses indicated more potential benefit from CSF when it was used to rule-out AD in individuals who were not expected to have AD (i.e. a treatment selection approach to prevent undertreatment) rather than to verify the AD diagnosis as set in clinical practice (a treatment selection approach to prevent overtreatment), given the assumption on the availability of a hypothetical DMT.

In the **final chapter** the findings of this thesis were discussed. The importance was raised of using a clinician's interpretation of a medical test result in diagnostic research to ensure that the study results reflect actual clinical practice. Several limitations were mentioned including the mixed use of persons from the general population and patients who visited a memory clinical, who may represent different expressions of AD. Next, the implications of the thesis results were discussed. The CSF decision rule did not improve the prognosis and can therefore not be recommended in the clinical practice. If a DMT becomes available in the future, a CSF test is likely able to efficiently allocate it to patients with MCI. Furthermore, several research directions were provided including the relevance to estimate the non-medical consequences of the AD-label in absence of treatment before the newly developed medical tests are adopted in current practice.

At last, this thesis contained a summary, a description of knowledge valorisation, a list of publications, information about the author, and acknowledgements.

SAMENVATTING

Evaluatieonderzoek naar diagnostiek voor de zieke van Alzheimer

Er is een toenemende belangstelling naar medische testen voor het identificeren van de pathologie van de ziekte van Alzheimer (AD) bij patiënten met een lichte cognitieve stoornis (MCI). Biomarkers in het hersenvocht (CSF), Positron Emissie Tomografie (PET) en Magnetic Resonance Imaging (MRI) zijn voorbeelden van zulke testen. Onlangs zijn richtlijnen voor de diagnostiek van AD in wetenschappelijk onderzoek voorgesteld. Deze nemen de hypothese aan dat biomarkers een toegevoegde waarde hebben bovenop de reguliere diagnostiek, enerzijds voor het voorspellen van functionele achteruitgang en anderzijds voor het nemen van therapeutische beslissingen zodra effectieve behandelingen voor patiënten met MCI voor handen zijn. Vooral CSF wordt reeds in de reguliere zorg toegepast hoewel deze test nog wetenschappelijk wordt onderzocht. Daarom is het belangrijk de meerwaarde van CSF voor de prognose van MCI vast te stellen. Ook is het belangrijk de balans te bepalen tussen maatschappelijke gezondheid en zorgkosten als gevolg van het gebruik van AD biomarkers.

Deze thesis heeft als doel de accuraatheid ten aanzien van prognose en de kosten-utiliteit van routinematig gebruik van CSF te bepalen. In deel I werd de methodologie verkend om de meerwaarde van een medische test te bepalen. Vervolgens werd deze methodologie toegepast om CSF te evalueren. In deel II werden verschillende studies uitgevoerd waarvan de resultaten werden gebruikt om een beslismodel te bouwen. Met dat model werden MCI patiënten gesimuleerd waarmee de kosten-utiliteit van CSF kon worden berekend.

Om het doel van deze thesis te bereiken werd de LeARN studie opgezet. De studieopzet, de methodologische overwegingen en de keuzes voor meetinstrumenten werden in **hoofdstuk 2** beschreven. In de LeARN studie werden in 4 Nederlandse academische geheugenpoliklinieken 241 patiënten benaderd bij wie men een neurodegeneratieve ziekte vermoedde. De klinische gegevens, kwaliteit van leven, biomarker informatie en het zorggebruik werd gedurende 2 jaar gemeten.

In **hoofdstuk 3** werd een protocol ontwikkeld waarmee een reguliere diagnose en prognose kon worden gesteld door een panel van klinische experts. In dit protocol werd ook de referentiestandaard beschreven waarmee de correctheid van een diagnose en prognose kon worden bepaald. Deze standaard was gebaseerd op informatie over het beloop van iemands klachten over een tijdsperiode van 2 jaar. Een aangepaste Delphi methode werd gehanteerd waarin 3 klinisch experts voor 11 patiënten in een proefstudie tot overeenstemming kwamen over de diagnose en prognose. Uit de resultaten bleek dat het protocol uitvoerbaar was voor wetenschappelijke doeleinden.

In **hoofdstuk 4** werd de toegevoegde waarde van CSF voor het voorspellen van cognitieve en functionele achteruitgang bepaald door het protocol uit hoofdstuk 3 toe te passen. Deze prognose, die de huidige reguliere zorg zonder CSF weerspiegelt, werd vergeleken met een beslisalgoritme waarin de CSF ratio Beta Amyloïde1-24 en t-tau de prognose kon herzien. Het beslisalgoritme leidde niet tot een betere voorspelling van cognitieve of functionele achteruitgang bij patiënten met subjectieve klachten, MCI of dementie.

In **hoofdstuk 5** werden de algemene en methodologische kenmerken van economische evaluaties van interventies voor een vroege diagnose van AD beschreven in een systematische review. Er werden 8 beslismodellen en 1 trialstudie gevonden die aanzienlijke diversiteit in studiedoel en (methodologische) kenmerken vertoonden. Aanbevelingen werden gericht op de diagnostische aspecten en de toepasbaarheid van de bestaande beslismodellen voor de evaluatie van de onlangs voorgestelde richtlijnen voor AD diagnostiek.

In **hoofdstuk 6** werd het natuurlijk beloop van AD onderzocht. Daarvoor werd gebruik gemaakt van het Zweedse Kungsholmen Project waarin 153 personen met MCI en 323 met AD werden geïdentificeerd uit een groep cognitief gezonde ouderen van 75 jaar of ouder uit de algemene populatie. De mediaan van de tijd tot ontwikkeling van AD-type dementie bij MCI was 3.5 jaar. Personen met AD-type dementie gingen 1.84 punten per jaar achteruit op de Mini-Mental State Examination en 0.38 punten per jaar op de Katz schaal voor lichamelijk functioneren. Leeftijd en opleidingsniveau waren geassocieerd met de snelheid van achteruitgang.

In **hoofdstuk 7** werden factoren onderzocht die de kosten van zorggebruik bij mensen met een cognitieve stoornis verklaren. Per patiënt bedroegen de gemiddelde jaarlijkse kosten voor de zorgsector €26,140 en voor patiënt en familie €11,931, inclusief de kosten voor informele zorg. De cognitie, de ervaren belasting van de mantelzorger, het geslacht van de patiënt, en de mate waarin activiteiten van het dagelijks leven konden worden uitgevoerd waren geassocieerd met de totale zorgkosten.

In **hoofdstuk 8** werd de kosten-utiliteit van CSF bij mensen met MCI bepaald in een toekomstscenario waarin een hypothetisch medicijn voorhanden is. Daarvoor werden individuen gesimuleerd over verschillende diagnose strategieën. Hun gemiddeld aantal levensjaren, gecorrigeerd voor gezondheidsgerelateerde kwaliteit van leven (QALY), en hun zorgkosten werden vergeleken vanuit een maatschappelijk perspectief. Op basis van wiskundige vergelijkingen gebaseerd op hoofdstuk 4, 5, 6 en 7 werd een beslismodel ontwikkeld. Het model vergeleek de reguliere zorg met 3 alternatieve strategieën waarin CSF werd gepositioneerd. In de eerste strategie werd CSF als perfecte test doorgerekend. In de twee andere strategieën werd CSF doorgerekend nadat een voorlopige klinische diagnose was gesteld. In alle berekeningen werd een hypothetisch medicijn verondersteld dat de AD progressiesnelheid halveert en alleen wordt verstrekt na een positieve diagnose. Een perfecte CSF test resulteerde in 0.39 additionele (0.26 tot 0.54) QALYs en een besparing van € 33,622 (21,232 tot 50,780) per individu en representeert daarmee de maximale ruimte voor verbetering van de huidige zorg. Uit de evaluatie van de 2 andere strategieën kwam een potentieel groter netto voordeel naar voren wanneer CSF werd ingezet om AD aan te tonen bij individuen bij wie geen AD werd verwacht (een behandelselectiebenadering ter voorkoming van onderbehandeling) dan wanneer een voorlopige klinische diagnose van AD werd bevestigd (het voorkomen van overbehandeling).

In het **laatste hoofdstuk** werden de bevindingen van dit proefschrift bediscussieerd. Het belang van het weerspiegelen van de klinische praktijk in diagnostisch onderzoek werd ter sprake gebracht. Verschillende beperkingen van de thesis werden genoemd waaronder het gebruik van zowel personen uit de algemene populatie als uit een geheugenpolikliniek. Tevens werden de implicaties van het proefschrift bediscussieerd. CSF, toegepast in een beslisalgoritme, verbeterde de prognose niet en daarom kan dit algoritme niet worden aangeraden in de reguliere zorg. Wanneer een medicijn beschikbaar is dat ingrijpt op de AD pathologie kan een CSF test dit medicijn waarschijnlijk op een efficiënte wijze toewijzen aan patiënten met MCI. Daarnaast werd

advies gegeven voor toekomstig wetenschappelijk onderzoek. Onder andere werd aanbevolen de niet-medische consequenties te meten van een Alzheimer diagnose bij mensen met MCI alvorens te besluiten om biomarkers onderdeel uit te laten maken van de reguliere zorg omdat er op dit moment nog geen werkzaam medicijn is bij MCI.

Ten slotte bevat dit proefschrift een samenvatting, een beschrijving van kennisvalorisatie, een publicatielijst, informatie over de auteur, en een dankwoord.

KNOWLEDGE VALORIZATION

SOCIETAL RELEVANCE

Alzheimer's disease (AD) causes dementia, which has an increasing impact on a human being's capacity to live independently. Furthermore, dementia affects the caregiving role of families and impacts their lives as well. It has been estimated that 35.6 million people lived with dementia worldwide in 2010 and this is expected to double in the next 20 years. In the Netherlands 65,000 individuals with dementia are known by general practitioners. Worldwide it caused 0.8% of all years lost due to death or equivalent healthy years lost due to disability in 2004. The associated societal economic impact was estimated at US$604 billion in 2010. Over the years, the proportion of the national budgets that governments spend at health care is increasing worldwide.

To control costs governments are forced to make choices in the wide range of newly developed medical technologies. In order to maximize societal health, as governments are obligated by their constitutions, decision makers choose those technologies with the largest health gain at the lowest costs. This prevents opportunity losses if a limited budget would be used up on obtaining technologies affecting only few people or result in only marginal gains. Making such choices involves many disciplines and requires independent research on the health benefits of interventions and cost estimates of the required care resource.

The results from this thesis are highly relevant in the decision making process. Technological progress resulted in novel biomarkers to detect AD pathology, among which markers in Magnetic Resonance Imaging (MRI), tracers for Positron Emission Tomography (PET) and markers in the Cerebrospinal Fluid (CSF). These biomarkers have the potential to improve the accuracy of a clinical diagnosis and the early prognosis of AD. They are increasingly being used in the clinical practice. To aid in deciding whether to adopt these markers in routine practice the full test-treat pathway and the impact on patient important outcomes should be evaluated. The studies in this thesis cover the full test-treat pathway and both patient-important outcomes in terms of health-related quality of life as well as care resource utilization. Besides the added value of biomarkers in current practice it also evaluates future scenarios. The evidence from this thesis is highly relevant in the decision making process aimed at improving health and reduce costs.

TARGET AUDIENCE

The results of this research are relevant to various stakeholders involved in dealing with AD and related disorders.

Patients may want to know the cause of their cognitive complaints. They may also want to know if they are at risk of increased care dependency. A biomarker indicating the probability of worsening disease progression would be helpful, but early diagnosis may have downside such as stigmatization and anxiety. The results of this research indicate the consequences of testing, which could be helpful for patients to make an informed decision whether or not to undergo an AD medical test.

The families of patients form the cornerstone of the support system. Often, they are consulted by the patient when important decisions have to be made and sometimes they are authorized to

make these decisions for patients. This makes the results of this thesis as important to patients as to their families or other members of their support network.

General practitioners are among the first medical professionals who are concerned with the memory complaints of a patient. They often have to deal with great uncertainty in concluding whether a patient should be referred for further testing. The results of this thesis are helpful for making decisions on referring for advanced AD medical testing.

Medical specialists, especially those working in memory clinics, are one of the core target audiences of the research from this thesis. Similar to GPs they deal with diagnostic uncertainty and have to decide what tests to perform to diminish this. The results of this thesis informs on the added value of biomarkers to routine clinical diagnostics. Furthermore, the subgroup analyses indicate what type of patient benefits the most from routine testing. The thesis also increases the understanding of the natural progression of AD, which could improve the accuracy of a prognosis in a patient.

Decision makers working in government organizations, such as the Dutch National Health Care Institute (Zorginstituut Nederland) or similar entities in Europe or worldwide, can add the results of this thesis to other evidence to form an advice on whether or not to adopt routine use of biomarkers in clinical practice. Other bodies such as the Dutch Healthcare Authority (Nederlandse Zorgautoriteit) could use it to determine the maximum amount for reimbursement e.g. in a diagnostic-treatment combination (DBC/DOT). The decision to adopt a biomarker is particularly important if it can be used to decide on a treatment that has a high budget impact.

Commercial organizations concerned with medical test development can use the results to improve their products. For example it was indicated that prevention of undertreatment could be realized by improving a test's sensitivity. This parameter can be adjusted by improving for example the lab procedure of analyzing the CSF samples.

Pharmaceutical companies are provided a future scenario in which treatment is cost-effective. This means the thesis gives an indication of the borders of minimal treatment effects and maximum treatment costs. Furthermore, it puts the treatment in its full test-treatment perspective and sketches opportunities to improve treatment provision and thereby make it more efficient.

At last, health insurance companies are key players in the access to care. They can discuss the use of advanced biomarkers by means of the results of this thesis with government decision makers, patients, specialists and the memory clinic board. It could aid in defining a financial structure for reimbursement of advances testing in order to maintain efficient care provision.

BUSINESS

The main product of this thesis is the decision analytical model that simulates persons with memory problems and their lifetime disease progression, quality of life and care resource use. This decision model can be used by pharmaceutical, biomedical or bioinformatics companies to simulate the effects of AD related interventions such as new treatments and diagnostic tests. By licensing, patenting or selling the model it can be commercialized.

The model can also be exploited as a service in which modeling experts help a company by generating health-economic evidence of their products. The service could be used to model any intervention that affects the disease progression in AD.

It can not only be of interest by commercial organizations but also by national or international research groups for example in a scientific collaboration. The service could be established at the department of the university or a startup-company for example at the Maastricht Health Campus.

INNOVATION

Several studies on advanced AD medical tests have been performed in the past such as ADNI (Alzheimer's Disease Neuroimaging Initiative in North America). It is aimed to identify neuroimaging measures and biomarkers associated with cognitive and functional changes in healthy elderly and persons with mild cognitive disorder (MCI) or AD. Furthermore, the DESCRIPA study (Development of Screening Guidelines and Clinical Criteria for Pre-dementia AD) is aimed to develop screening guidelines for pre-dementia AD in the general population. The uniqueness of this study is the assessment of resource utilization and health-related quality of life to enable an economic evaluation. Furthermore, the decision analytic model enables the evaluation of the optimal diagnostic strategy. Also, without a disease modifying treatment, the added value of biomarkers is uncertain. Therefore, the availability of such treatment was explored in the sensitivity analysis.

It is unique that this study relied on the clinical interpretation of all information available in standard clinical practice (history, physical, neurological, and neuropsychological examination, and MRI). This makes the results better reflect real-life clinical practice than currently available evidence that relied on statistical models that simplified the combination of medical tests. The results of this thesis closer reflect the value of routine use of biomarkers in practice than previous research. It emphasizes on care-dependency in its outcomes which is relevant from a clinical perspective.

Another innovation of the research described in this thesis is the improvement in AD decision analytic modeling. The model simulates natural progression in an innovative way. It does not simplify progression in terms a limited number of disease stages but describes the complex interrelation between cognition and functioning on the original measurements scales. Furthermore, other progression models in AD have been developed mostly among clinical samples or prevalent AD dementia cases. However, disease modifying treatments are supposed to be effective in early (pre-dementia) AD. Thus long-term data on the natural course, which were available for the analysis of this thesis, are required to evaluate their effectiveness. Besides this the model combines the pre-dementia with the dementia stage. None of the economic evaluations that evaluated the added value of AD biomarkers in clinical practice focussed on this early diagnosis in the MCI phase.

At last, the research in this thesis was performed within the Centre of Translational Molecular Medicine (CTMM). CTMM has innovation stated in its vision. Its innovative character can be found in the combination of both the development of new technologies as well as the early health-economic evaluation of it. This stimulates the translation and valorization of the newly developed techniques in clinical practice.

VALORIZATION PROCESS

Several steps are required before the evidence generated for this thesis is ready to be applied as described in the above sections. Two types of further research are required. First, the impact of biomarkers on patient important outcomes in current care practice with the currently available treatments should be estimated. Second, a process evaluation applying the results in our local memory clinic should be performed. This should be combined with the results of the thesis to compile the evidence base of biomarkers for AD.

For patients and their relatives the combined evidence should be translated to an information leaflet. The language of the leaflet should be accessible for persons with memory complaints to enable them to make an informed choice on further testing. It should contain a balanced description of the advantages and disadvantages of advanced testing.

For general practitioners and medical specialists the pile of evidence should be transformed into clinical guidelines. These can be the Dutch Orde Medisch Specialisten and international criteria such as from the NIA&AA (National Institute on Aging and Alzheimer's Association). In these criteria both the accuracy of biomarkers as well as the impact on patient important outcomes should be explained. Additional research could indicate related important aspects such as conditions when to stop testing (i.e. the threshold of sufficient diagnostic certainty). Also information on the optimal sequence of the available biomarkers, such as 'cheapest first' or 'most informative first' or patient-depending sequences could be added.

These recommendations do not come from single studies and require validation by other research groups and discussion in the literature, social media and on international conferences. It is therefore important to spread the research results in the scientific community as much as possible.

Decision makers need a different type of recommendation report. It should include a summary of evidence of the full spectrum of health technology assessment, including for example effectiveness, cost-effectiveness, ethical and legal considerations. Such report should balance the technical details. The government's decision makers should also be provided a budget impact analysis on the national available care budget.

To generate business from the decision model it should be transformed into a software package with a user-friendly interface. Although companies could pick up the possibilities of the model on international conferences, an active process of contacting companies for a collaboration using or exploitation of the model would probably be more effective to valorize the evidence.

PUBLICATIONS

INTERNATIONAL PEER REVIEWED JOURNALS

Wimo A, Ballard C, Brayne C, Gauthier S, **Handels R,** Jones RW, Jonsson L, Khachaturian AS, Kramberger M. Health economic evaluation of treatments for Alzheimer's disease: impact of new diagnostic criteria. J Intern Med. 2014 Mar;275(3):304-16.

Handels RL, Xu W, Rizzuto D, Caracciolo B, Wang R, Winblad B, Verhey FR, Severens JL, Fratiglioni L, Joore MA, Wimo A. Natural Progression Model of Cognition and Physical Functioning among People with Mild Cognitive Impairment and Alzheimer's Disease. J Alzheimers Dis. 2013;37(2):357-65.

Handels RL, Wolfs CA, Aalten P, Joore MA, Verhey FR, Severens JL. Diagnosing Alzheimer's disease: A systematic review of economic evaluations. Alzheimers Dement. 2014 Mar;10(2):225-37.

Ossenkoppele R, Prins ND, Pijnenburg YA, Lemstra AW, van der Flier WM, Adriaanse SF, Windhorst AD, **Handels RL**, Wolfs CA, Aalten P, Verhey FR, Verbeek MM, van Buchem MA, Hoekstra OS, Lammertsma AA, Scheltens P, van Berckel BN. Impact of molecular imaging on the diagnostic process in a memory clinic. Alzheimers Dement. 2013 Jul;9(4):414-21.

Handels RL, Wolfs CA, Aalten P, Verhey FR, Severens JL. Determinants of Care Costs of Patients With Dementia or Cognitive Impairment. Alzheimer Dis Assoc Disord. 2013 Jan-Mar;27(1):30-6.

Handels R, Aalten P, Wolfs CA, Olderikkert M, Scheltens P, Visser PJ, Joore MA, Severens JL, Verhey FR. Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease: the research protocol of a prospective cohort study. BMC Neurol. 2012 Aug 10;12:72.

Vos S, van Rossum I, Burns L, Knol D, Scheltens P, Soininen H, Wahlund LO, Hampel H, Tsolaki M, Minthon L, **Handels R**, L'italien G, van der Flier W, Aalten P, Teunissen C, Barkhof F, Blennow K, Wolz R, Rueckert D, Verhey F, Visser PJ. Test sequence of CSF and MRI biomarkers for prediction of AD in subjects with MCI. Neurobiol Aging. 2012 Oct;33(10):2272-81.

van Rossum IA, Vos S, **Handels R**, Visser PJ. Biomarkers as predictors for conversion from mild cognitive impairment to Alzheimer-type dementia: implications for trial design. J Alzheimers Dis. 2010;20(3):881-91.

PUBLISHED ABSTRACTS & PRESENTATIONS AT (INTER)NATIONAL CONFERENCES

Handels RLH, Joore M, Tran-Duy A, Wolfs CAG, Verhey FRJ, Severens JL. Cost-utility of Alzheimer's disease biomarkers. Alzheimer's Association International Conference (AAIC) 2014 Copenhagen (Denmark). *Oral presentation & session chair*

Handels RLH, Joore M, Vos SJB, Aalten P, Ramakers I, Olde-Rikkert M, Scheltens P, van Berckel BNM, Leentjens A, Wolfs CAG, Severens JL, Verhey FRJ. Added diagnostic value of cerebrospinal fluid in predicting decline in memory clinic subjects in clinical practice. Alzheimer's Association International Conference (AAIC) 2014 Copenhagen (Denmark). *Poster presentation*

Handels RLH, Joore M, Vos SJB, Aalten P, Ramakers I, Olde-Rikkert M, Scheltens P, van Berckel BNM, Leentjens A, Wolfs CAG, Severens JL, Verhey FRJ. Advanced diagnostic techniques for neurodegenerative diseases: are they ready to adopt in the clinical routine? Karel V symposium 2014, Utrecht (Netherlands). *Oral presentation*

Handels RLH, Joore MA, Tran-Duy A, Wolfs CAG, Verhey FRJ, Severens JL. Early cost-utility analysis of Alzheimer's disease diagnostic tests. The Low Lands Health Economists' Study Group (LoLaHESG) 2014, Oostvoorne (the Netherlands). *Oral presentation & session chair*

Vos S, van Rossum I, Burns L, Knol D, van der Flier W, Blankenstein M, Barkhof F, Scheltens P, Soininen H, Wahlund L, Hampel H, Tsolaki M, Minthon L, Blennow K, Verhey F, **Handels R**, Wolz R, Rueckert D, Visser PJ. CSF and MRI biomarkers as predictors for Alzheimer-type dementia in subjects with MCI. International Conference on Alzheimer's Disease (AAIC) 2011, Paris (France).

Handels RL, Wolfs CA, Aalten P, Verhey FR, Severens JL. Determinants of care costs of patients with dementia or cognitive impairment. International Psychogeriatric Association (IPA) 2011, den Haag (the Netherlands). International Psychogeriatrics (2011), Vol. 23, Supplement 1, S139–S400. PS02.23. *Poster presentation*

CURRICULUM VITAE

Ron Handels is born in Kerkrade on March 15 1986 and grew up in Heel (Netherlands). He graduated from high school at St. Ursula in Horn in 2004. In the next year he received his propaedeutic certificated in mechanical engineering at the Technical University of Eindhoven. After that he studied health sciences at Maastricht University, followed the honors program and was member of the board of examiners, education Committee, board of the Student Council Health and Life Sciences, and board of the study association. In 2008 he started the Health Sciences Research Master at Maastricht University and specialized in Health Technology Assessment. He also worked as a research assistant on two projects. After he graduated he started working as a PhD candidate at the Department of Psychiatry and Neuropsychology at Maastricht University under supervision of Prof. Dr. Frans Verhey and Prof. Dr. Hans Severens. As part of his PhD project, he worked for 3 months in Stockholm at the Karolinska Institutet to collaborate with Prof. Dr. Anders Wimo and colleagues. Presently, Ron Handels works as a postdoctoral researcher at both Maastricht University and Karolinska Institutet.

188

Allereerst wil ik graag alle patiënten en hun naasten bedanken die hebben deelgenomen aan de onderzoeken waarop dit proefschrift is gebaseerd. Het is verre van vanzelfsprekend dat jullie je inzetten voor de wetenschap. Het is dan ook een eer om met de gegevens die jullie beschikbaar stellen te mogen werken.

In het bijzonder wil ik mijn promotoren, Prof. Dr. Verhey en Prof. Dr. Severens bedankten.

Beste Frans, ik zie nog steeds de verbazing op je gezicht toen ik vlak voor het einde van het promotietraject meldde dat ik ging trouwen en in verwachting was van mijn dochter. Ik heb geen moment druk ervaren in de combinatie privé-werk en dat heb ik aan jou te danken. Jij weet een prettige sfeer te creëren waardoor ik met plezier uitkijk naar onze toekomstige overleggen en congressen. Daarnaast heb jij me geleerd me op wetenschappelijke wijze uit te drukken, zowel in geschreven als gesproken woord. Ik dank je daar hartelijk voor.

Beste Hans, jij enthousiasmeerde me 7 jaar geleden tijdens de bachelor opleiding voor het doelmatigheidsonderzoek en dat is alleen maar toegenomen. Daarbij verbaast het me telkens weer hoe effectief jij kunt overleggen met diepgang en humor. Ik vond het dan ook jammer dat je naar Rotterdam bent gegaan. Toch bleef je altijd betrokken en was je er op cruciale momenten. Die efficiëntie is precies wat mij aan jou blijft inspireren.

Beste Manuela, door jou durf ik me met goed fatsoen gezondheidswetenschapper in doelmatigheidsonderzoek te noemen. Jij hebt me alle fijne kneepjes van het vak geleerd. Je wist altijd een antwoord op mijn vragen of je netwerk daarvoor in te schakelen. Bedankt dat je de discussie opzocht en me hielp om kwesties tot in detail uit te zoeken. Daar ben ik veel wijzer van geworden.

Beste Claire, jij hebt me de basics van de geheugenpoli en de zorg rondom dementie bijgebracht; onder andere het verschil tussen een psychiater en een psycholoog. Daarbij vond ik het fijn een inhoudelijk maatje op de afdeling te hebben. Je was er altijd voor me als ik weer eens de weg kwijtraakte in mijn ongestructureerde ideeën. Bedankt voor jouw geduld met mijn chaotisch karakter.

Dear Anders, let's bake those breads! Meanwhile we can apply our insights from research to our personal lives and generate a one-liner. My nervousness to go abroad evaporated after my first day at the Karolinska Institutet. If your enthusiasm is viral it infected me in a cost-effective way.

Also my special thanks to your colleagues Laura Fratiglioni, Bengt Winblad, Gunilla Johansson and Weili Xu, you are thé example of international collaboration. And thank you Niels Andreasen and Erik Jedenius for providing me with a nice home (and the Danish bitter that give me a day off). Anders Sköldunger and my roommates and close colleagues at Karolinska Institutet, thank you for your warm welcome and all the fun we had.

Pauline en Inez, dankzij jullie organisatorische vaardigheden in de LeARN studie kon ik me volledig focussen op de inhoud. Ik heb af een toe mijn hoofd eens boven de boeken uit gestoken en

Henny Bruinewoud, jij bent een organisatorische inspiratiebron. Bedankt dat je open stond voor mijn ideeën en altijd met me meedacht.

Rik, Sofie, Marissa, Petra, Saskia en Sharon, jullie maken het Alzheimer-onderzoekers-wereldje een must. Bedankt voor het vlot beantwoorden van mijn talloze e-mails over de LeARN data. Tevens alle collega's aan de 4 LeARN centra bedankt voor de studieopzet en dataverzameling.

Alzheimer Nederland en jullie donateurs, bedankt dat jullie mijn bezoek aan het Karolinska Institutet mogelijk hebben gemaakt. Ik waardeer jullie vertrouwen in mijn onderzoek ideeën.

Christel van Gool, destijds had jij me klaargestoomd voor dit promotiewerk en ik hoop dat mijn opgebouwde kennis en kunde ooit van pas mag komen in een jouw masteronderwijs.

Bakir, we waren een goed team tijdens het studentenberaad. Je hebt toen mijn enthousiasme voor werken binnen een universiteit gekatalyseerd.

Tom, Yori en Ivo, dankzij jullie was er ook enige ontspanning buiten mijn werk.

Hub, Nellie, Mark en Erik, jullie hebben me gevormd tot wie ik ben. Ik ben een geluksvogel met jullie.

Lore, ik had nooit durven dromen hoeveel energie ik van jou krijg. Jou nieuwsgierigheid lijkt onuitputtelijk en is een ware inspiratiebron.

Inge, jij bent alles voor mij wat de wetenschap niet is. Jouw onvoorwaardelijke vertrouwen in mij is een belangrijke basis waardoor ik dit werk kan doen.

Inge, je hebt me altijd gesteund, zelfs als dat betekende dat ik enkele maanden van huis was.

Inge, ik kan je nooit genoeg bedanken.

"It's always too early to evaluate a technology... until suddenly it's too late."

Martin Buxton