

Positronenemissionstomographie (PET) und PET/CT bei Alzheimer-Demenz

Berichtsplan

Auftrag: D06-01N
Version: 1.0
Stand: 05.06.2013

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Positronenemissionstomographie (PET) und PET/CT bei Alzheimer-Demenz

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

21.12.2006

Interne Auftragsnummer:

D06-01N

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8 (KölnTurm)
50670 Köln

Tel.: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

Schlagwörter: Alzheimer Krankheit, Positronen-Emissions-Tomographie, Computertomographie, systematische Übersicht

Keywords: Alzheimer Disease, Positron-Emission Tomography, Tomography, X-Ray Computed, Systematic Review

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	vi
Abbildungsverzeichnis	vii
Abkürzungsverzeichnis	viii
1 Hintergrund	1
1.1 Definition des Krankheitsbildes	1
1.2 Epidemiologie und individuelle Krankheitslast	1
1.3 Verlauf der Erkrankung	2
1.4 Derzeit übliche Therapien	4
1.5 Diagnostische Verfahren	5
1.5.1 Übliche diagnostische Verfahren	5
1.5.2 Biomarker	8
1.6 PET und PET/CT	10
1.6.1 Empfehlungen wissenschaftlicher Fachgesellschaften zur PET bzw. PET/CT bei Alzheimer-Demenz.....	11
2 Ziele der Untersuchung	12
3 Projektbearbeitung	14
3.1 Zeitlicher Verlauf des Projekts	14
3.2 Dokumentation der Änderungen im Projektverlauf	14
4 Methoden	16
4.1 Ermittlung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT	17
4.1.1 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT anhand von systematischen Übersichten.....	18
4.1.2 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT anhand von Primärstudien.....	18
4.1.2.1 Populationen	18
4.1.2.2 Prüf- und Vergleichsinterventionen.....	18
4.1.2.3 Patientenrelevante Endpunkte.....	19
4.1.2.4 Studientypen	20
4.1.2.5 Sonstige Studiencharakteristika.....	21
4.1.3 Tabellarische Darstellung der Einschlusskriterien („Nutzenstudien“).....	21
4.2 Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT	22

4.2.1	Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Güte anhand von systematischen Übersichten	23
4.2.1.1	Populationen	23
4.2.1.2	Prüftechnologien (Indextest I), Vergleichstechnologien (Indextest II) und Referenzstandard	23
4.2.1.3	Zielgrößen	24
4.2.1.4	Den systematischen Übersichten zugrunde liegende Studientypen.....	24
4.2.1.5	Sonstige Charakteristika der systematischen Übersichten.....	25
4.2.1.6	Tabellarische Darstellung der Einschlusskriterien (systematische Übersichten zur diagnostischen und prognostischen Güte).....	26
4.2.2	Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Güte anhand von Primärstudien („Ergänzungsrecherche“).....	27
4.2.2.1	Populationen	27
4.2.2.2	Prüftechnologien (Indextest I), Vergleichstechnologien (Indextest II) und Referenzstandard	27
4.2.2.3	Zielgrößen	27
4.2.2.4	Studientypen	27
4.2.2.5	Sonstige Studiencharakteristika.....	27
4.2.2.6	Tabellarische Darstellung der Einschlusskriterien („Ergänzungsrecherche“ nach Primärstudien zur diagnostischen und prognostischen Güte)	27
4.2.3	Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen.....	28
4.3	Informationsbeschaffung.....	28
4.3.1	Bibliografische Literaturrecherche	28
4.3.2	Suche nach weiteren publizierten und nicht publizierten Studien.....	29
4.3.3	Selektion relevanter Studien und systematischer Übersichten	29
4.4	Informationsbewertung	30
4.4.1	Bewertung von Studien zum patientenrelevanten Nutzen und Schaden	30
4.4.2	Bewertung von systematischen Übersichten zur diagnostischen und prognostischen Güte	31
4.4.3	Bewertung von Primärstudien zur diagnostischen und prognostischen Güte	32
4.5	Informationssynthese und -analyse	32
4.5.1	Gegenüberstellung der Ergebnisse der Einzelstudien	33
4.5.1.1	Gegenüberstellung der Ergebnisse von (Primär-)Studien zur Bewertung des patientenrelevanten Nutzens und Schadens.....	34
4.5.1.2	Gegenüberstellung der Ergebnisse von systematischen Übersichten zur diagnostischen und prognostischen Güte.....	34
4.5.1.3	Gegenüberstellung der Ergebnisse von Primärstudien zur diagnostischen und prognostischen Güte („Ergänzungsrecherche“).....	35

4.5.1.4	Gegenüberstellung der Ergebnisse aus systematischen Übersichten und der Ergänzungsrecherche („Robustheitsprüfung“)	35
4.5.2	Meta-Analysen	35
4.5.2.1	Meta-Analysen für vergleichende Interventionsstudien	35
4.5.2.2	Meta-Analysen für Prognosestudien	36
4.5.2.3	Meta-Analysen für Diagnosestudien	36
4.5.3	Sensitivitätsanalysen	37
4.5.4	Subgruppenmerkmale und andere Effektmodifikatoren	38
5	Literaturverzeichnis	39
	Anhang A – Offenlegung potenzieller Interessenkonflikte von Patientenvertretern	47

Tabellenverzeichnis

	Seite
Tabelle 1: ICD-10-Kriterien für eine Alzheimer-Demenz	5
Tabelle 2: DSM-IV-Kriterien für eine Alzheimer-Demenz	6
Tabelle 3: Kriterien der National Institute of Neurologic, Communicative Disorders and Stroke-Alzheimer's Disease and related Disorders Association Work Group (NINCDS-ADRDA)	7
Tabelle 4: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren.....	16
Tabelle 5: Evidenzklassifizierung des G-BA für Unterlagen zu diagnostischen Methoden	17
Tabelle 6: Übersicht der Einschlusskriterien für Primärstudien zur Nutzenbewertung der PET bzw. PET/CT.....	22
Tabelle 7: Übersicht der Einschlusskriterien für systematische Übersichten zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT.....	26
Tabelle 8: Übersicht der Einschlusskriterien für die Primärliteratur zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT (Ergänzungsrecherche).....	28

Abbildungsverzeichnis

	Seite
Abbildung 1: Pathophysiologische Kaskade der Alzheimer-Demenz	3
Abbildung 2: Biomarker zur Diagnose der Alzheimer-Demenz.....	9

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AA	Alzheimer's Association
A β	β -Amyloid-Peptid
AD	Alzheimer-Demenz
ADL	Activities of Daily Living (Aktivitäten des täglichen Lebens)
ADRDA	Alzheimer's Disease and Related Disorders Association
APP	Amyloid-Precursor-Protein
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BMI	Body-Mass-Index
CCT	Controlled Clinical Trial (kontrollierte klinische Studie)
CT	Computertomographie
DEGAM	Deutsche Gesellschaft für Allgemeinmedizin
DGN	Deutsche Gesellschaft für Neurologie
DGN	Deutsche Gesellschaft für Nuklearmedizin
DGPPN	Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, 4 th edition (4. Auflage des Diagnostischen und Statistischen Handbuchs Psychischer Störungen)
EFNS	European Federation of Neurological Sciences
EMA	European Medicines Agency
FDG	Fluor-18-2-Fluoro-2-Deoxy-D-Glukose
fMRT	Funktionelle MRT
G-BA	Gemeinsamer Bundesausschuss
G-I-N	Guidelines International Network (internationales Leitlinien-Netzwerk)
HTA	Health Technology Assessment
ICD-10-GM	International Statistical Classification of Diseases and Related Health Problems, 10 th edition (Deutschsprachige Version der 10. Ausgabe der internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme)
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
LR	Likelihood Ratio
MCI	Mild Cognitive Impairment (leichte kognitive Beeinträchtigung)
MRT	Magnetresonanztomographie
NGC	National Guideline Clearinghouse
NIA	National Institute on Aging

Abkürzung	Bedeutung
NINCDS	National Institute of Neurological and Communicative Disorders and Stroke
PET	Positronenemissionstomographie
PET/CT	Integration von PET und CT in einem Untersuchungsgerät
PIB	Pittsburgh compound-B
PICO	Patient, Intervention, Comparison (Vergleichsintervention), Outcome (patientenrelevanter Endpunkt)
QOL-AD	Quality of Life – Alzheimer’s Disease (Instrument zur Messung der gesundheitsbezogenen Lebensqualität von Patienten mit Alzheimer-Demenz)
RI	Reagan Institute
SMI	Subjective Memory Impairment (subjektive Gedächtnisstörungen)
SN	Sensitivität
SP	Spezifität
SPECT	Single-Photon-Emissions-Computertomographie
RCT	Randomized controlled trial (randomisierte kontrollierte Studie)
ROC	Receiver Operating Characteristic (Methode zur Bewertung und Optimierung von Analysestrategien)
VOPT	Verification of Only Positive Testers

1 Hintergrund

1.1 Definition des Krankheitsbildes

Mit dem Begriff der Demenz wird eine chronische und meist progressive Fehlfunktion des Gehirns bezeichnet, die zu einer Verschlechterung des Gedächtnisses und anderer kognitiver Funktionen, zu einer Beeinträchtigung der Aktivitäten des täglichen Lebens (im Folgenden: ADL [„Activities of Daily Living“]) und zu wechselnd ausgeprägten begleitenden psychopathologischen Symptomen führt [1].

Die deutsche Version der internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD-10-GM) definiert die Alzheimer-Demenz (AD) als eine primär degenerative zerebrale Krankheit mit unbekannter Ätiologie und charakteristischen neuropathologischen und neurochemischen Merkmalen. Sie beginnt meist schleichend und entwickelt sich langsam, aber stetig über einen Zeitraum von mehreren Jahren [2].

Die AD stellt mit einem Anteil von ca. 50 bis 70 % das häufigste alleinige Krankheitsbild eines demenziellen Syndroms dar. Als zweithäufigste Form gilt die frontotemporale Demenz. Daneben kommen vaskulär bedingte Demenzen, auch als Mischform zusammen mit der AD, vor. Andere Ursachen der Demenz wie Morbus Parkinson, Lewy-Körper-Demenz, Creutzfeldt-Jakob-Krankheit sowie reversible demenzielle Zustandsbilder im Gefolge anderer Erkrankungen sind deutlich seltener [3-7].

1.2 Epidemiologie und individuelle Krankheitslast

In Deutschland sind derzeit rund 1 Million Personen von Demenz betroffen. Davon lassen sich etwa 700 000 Fälle auf die AD zurückführen [4,5]. Frauen sind häufiger betroffen als Männer, Prävalenz und Inzidenz steigen mit dem Alter [7].

Verschiedene Risikofaktoren der Entstehung von Demenzen sind bekannt. Dabei kann zwischen unabänderlichen und beeinflussbaren Faktoren unterschieden werden [8].

Zu den unabänderlichen Risikofaktoren zählen Alter, Geschlecht und genetische Faktoren. Das Risiko des Auftretens einer Alzheimer-Demenz steigt mit dem Alter, welches als wichtigster Faktor für die Entstehung der Alzheimer-Demenz angesehen wird. Auch nach Korrektur der unterschiedlichen Lebenserwartungen ist das Risiko des Auftretens einer AD bei Frauen höher als bei Männern [9]. Das Vorliegen einer Demenz bei Verwandten ersten Grades ist ebenfalls mit einem erhöhten Risiko verbunden [10,11]. Genetische Faktoren sind selten (ca. 1 % aller Alzheimer-Demenzen) und bewirken üblicherweise ein Krankheitsauftreten vor dem 55. Lebensjahr [8,12].

Zu den beeinflussbaren Faktoren zählen u. a. Rauchverhalten, Ernährungsgewohnheiten und metabolisch-vaskuläre Einflüsse sowie geistige bzw. soziale Stimulation. Das Risiko des Auftretens einer Alzheimer-Demenz ist bei Rauchern höher als bei Nichtrauchern [13,14].

Welchen Einfluss eine Hormonersatztherapie auf den Zeitpunkt des Auftretens einer Alzheimer-Demenz hat, ist umstritten [15,16]. Einige Autoren geben an, dass das Risiko für das Auftreten einer AD durch eine Hormonersatztherapie erhöht sein könnte [17]. Eine Meta-Analyse ermittelt einen statistisch nicht signifikanten Vorteil durch die Behandlung von Patienten mit Bluthochdruck hinsichtlich des Auftretens von Demenzen (RR 0,80; 0,63-1,02). Getrennte Daten für Alzheimer-Demenz wurden nicht berichtet [18].

Weitere Faktoren werden diskutiert. So sollen erhöhte Werte des Apolipoproteins E und erhöhte Cholesterinwerte, ein erhöhter Body-Mass-Index (BMI), Diabetes mellitus, Kopfverletzungen und Depressionen das Risiko für AD erhöhen, wohingegen regelmäßige geistige bzw. soziale Stimulation sowie sportliche Aktivität einen positiven Einfluss haben sollen [8,14,19,20]. Die Rolle von Vitaminen für die Prävention bzw. ein verspätetes Auftreten einer Alzheimer-Demenz ist nicht vollständig geklärt [21].

Für die leichte kognitive Beeinträchtigung (Mild cognitive impairment, MCI), die als frühes Stadium nicht notwendigerweise in eine Alzheimer-Demenz übergeht, werden teilweise andere Risikofaktoren, wie z. B. Bildungsgrad und genetische Faktoren, beschrieben [22].

1.3 Verlauf der Erkrankung

In der Pathogenese der Alzheimer-Demenz spielen der Metabolismus des Amyloid-Precursor-Proteins (APP) und sein Abbau zu β -Amyloid-Peptiden ($A\beta$) eine herausragende Rolle. Die Evidenz hierfür wurde anhand von Mutationen in den humanen Genen des Amyloid-metabolismus und Untersuchungen an transgenen Mäusen gewonnen [23-25]. $A\beta$ lagert sich in Form extrazellulärer Aggregate, sogenannter „Plaques“, im Gehirn ab. Neben einem primären Defekt im APP-Stoffwechsel mit einer vermehrten Bildung von $A\beta$ besonders bei früh erkrankten, genetisch belasteten Patienten wird auch ein primärer Defekt im Abbau von $A\beta$ als Ursache der AD diskutiert. Neben diesen extrazellulären Proteinablagerungen treten als weiteres pathologisches Kennzeichen der AD Ablagerungen aus abnorm phosphoryliertem Tau(τ)-Protein auf. Dieses Tau-Protein bildet nicht mehr abbaubare Konglomerate, sogenannte Neurofibrillenbündel (Tangles), die intrazellulär abgelagert werden und zum Absterben der Nervenzellen führen [26].

Insbesondere im höheren Alter beziehungsweise bei fortschreitender Pathologie sind weitere gewebeschädigende Prozesse nachweisbar (oxidativer Stress, Zellschädigung durch glutamaterge Exzitotoxizität, entzündliche Vorgänge, neuronale Apoptose). Die Anfänge der neuropathologischen Veränderungen gehen den manifesten kognitiven Störungen um viele Jahre voraus [27,28].

Sperling et al. schlagen in einer aktuellen Arbeit ein Modell der pathophysiologischen Kaskade vor, die zur Alzheimer-Demenz führt [28] (Abbildung 1). Darin wird die Ablagerung von Plaques als Schlüsselereignis in der Pathophysiologie der AD angesehen. Die Autoren betonen aber, dass die Ätiologie der AD nicht vollständig geklärt ist. Dazu verweisen sie auf mögliche synaptische, mitochondriale, metabolische, entzündliche, neuronale, zytoskeletale

und altersbedingte Einflüsse hin, die von anderen Arbeitsgruppen vorgeschlagen wurden [29,30]. Im Vergleich zu Plaques-Ablagerungen könnten diese einen früheren Einfluss auf den Verlauf der Erkrankung haben bzw. eine zentrale Rolle in der Pathogenese einnehmen [28].

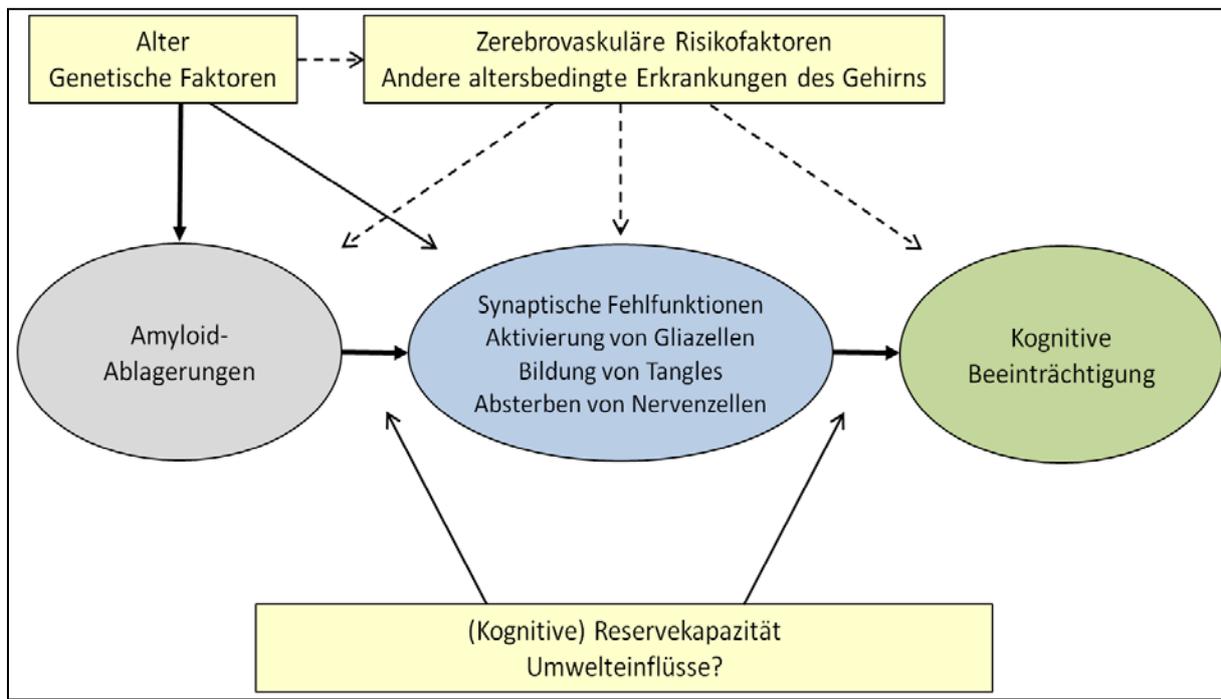


Abbildung 1: Pathophysiologische Kaskade der Alzheimer-Demenz (Modifiziert nach Sperling et al. 2011) [28])

Die Alzheimer-Demenz fällt zwar meist zuerst wegen kognitiver Symptome, insbesondere der Merkfähigkeitsstörung, auf. Die Beeinträchtigungen der Aktivitäten des täglichen Lebens, besonders der basalen Fähigkeiten (wie zum Beispiel Körperpflege), bestimmen aber den Pflegeaufwand. Daneben spielen begleitende psychopathologische Symptome wie Apathie, Depression, Agitiertheit, Ängstlichkeit, Schlafstörung und paranoide Symptome eine wesentliche Rolle im Hinblick auf die Lebensqualität und können auch die pflegenden Angehörigen in hohem Maß belasten [31].

Eine Studie gibt an, dass die mediane Überlebenszeit von Patienten mit einer wahrscheinlichen AD vergleichbar ist mit der Überlebenszeit von Patienten mit chronischen Erkrankungen (z. B. koronare Herzerkrankung) [32].

Bei der Erkrankung können 3 Stadien unterschieden werden [28,33,34]:

1) Prä-klinische Phase

Diese Phase kann mehrere Jahre dauern. Dabei können Veränderungen von Biomarkern, wie z. B. eine erhöhte Amyloid-Tracer Retention in der Positronenemissionstomographie (PET), gesehen werden, welche jedoch nicht mit einer klinischen Veränderung des Patienten, insbesondere einer Beeinträchtigung der kognitiven Fähigkeiten, einhergeht [28,33].

2) Leichte kognitive Beeinträchtigung (Mild cognitive impairment, MCI)

Die vorwiegend klinisch gestellte Diagnose kann durch pathologische Werte weiterer Biomarker unterstützt werden. Andere Ursachen einer MCI, wie z. B. Depressionen, Flüssigkeitsmangel oder altersassoziierter kognitiver Abbau, müssen ausgeschlossen werden.

3) Alzheimer-Erkrankung

Letztes Stadium der Erkrankung, in dem die Symptome – insbesondere auch die psychopathologischen – am deutlichsten ausgeprägt sind.

Es ist zu beachten, dass der Übergang zwischen den 3 Stadien nicht klar abgrenzbar ist. Ein Vorliegen der Stadien 1 und 2 ist mit einer erhöhten Wahrscheinlichkeit für eine Alzheimer-Erkrankung verbunden, muss jedoch nicht zwangsläufig darin übergehen.

1.4 Derzeit übliche Therapien

Behandlungs- und Therapiemaßnahmen können angesichts der komplexen und sich mit zunehmender Krankheitsprogression verändernden Bedürfnislage in 3 Bereiche gegliedert werden:

- Psychopharmakologische Therapien (Antidementiva oder andere psychotrope Medikamente) zielen auf eine direkte Veränderung kognitiver oder psychologischer Leistungen des Demenzerkrankten.
- Psychologisch-verhaltenstherapeutische Interventionen setzen im Frühstadium der Erkrankung auf modifizierbare Verhaltensmuster oder Ersatzstrategien.
- Interventionen im Sozialbereich (Angehörigenberatung und -schulung, Tagesstätten, Pflegedienste, betreute Wohnformen) treten ab einem mittleren Erkrankungsstadium in den Vordergrund.

Alle 3 Interventionsebenen interagieren dabei in hohem Maße.

1.5 Diagnostische Verfahren

1.5.1 Übliche diagnostische Verfahren

Es bestehen deutliche Schwierigkeiten, die Beeinträchtigungen und Bedürfnisse der von der Alzheimer-Demenz Betroffenen zu erfassen. Probleme können fluktuieren oder nur in bestimmten Situationen in Erscheinung treten. Zudem können krankheitsbedingte Verhaltens-, Aufmerksamkeits- und Gedächtnisstörungen und der krankheitsbedingte Verlust kommunikativer Fähigkeiten die adäquate Äußerung von Bedürfnissen oder Präferenzen durch den Erkrankten beeinträchtigen. Ein weiteres spezielles Problem demenzieller Erkrankungen ist die bei einem Teil der Patienten fehlende Selbstwahrnehmung eigener Defizite und Hilfsbedürftigkeit. Diese krankheitsbedingt fehlende Selbstwahrnehmung kann bei Selbsteinschätzung zu einer von außen gesehen groben Fehleinschätzung möglicher Beeinträchtigungen und Bedürfnisse führen [35].

Die Auswirkungen der AD und die Wirksamkeit einer Therapie können außer vom Betroffenen selbst (mit den oben erwähnten Einschränkungen) vom behandelnden Personal und von den Angehörigen oder Pflegenden beurteilt werden. Die Bewertung der Ausprägung der Krankheitssymptome und der dadurch bedingten Beeinträchtigungen kann aber je nach Bereich und Beurteiler variieren [36,37].

Für die Diagnose einer AD sind zurzeit keine ausreichend sensitiven und spezifischen Kriterien in der klinischen Routine verfügbar. Neben dem Nachweis des typischen klinischen Bildes einer Demenz erfolgt die Diagnose auch durch den Ausschluss möglicher Alternativursachen. Eine Differenzierung zwischen einer Demenz vom Alzheimer-Typ und einer vaskulären Demenz aufgrund klinischer Kriterien kann jedoch im Einzelfall zu einer Fehldiagnose führen [38]. Die Diagnose einer AD wird international anhand der ICD-10 (Tabelle 1) oder der verwandten DSM-IV-Kriterien (Tabelle 2) gestellt.

Tabelle 1: ICD-10-Kriterien für eine Alzheimer-Demenz [39]

Vorliegen einer Demenz	1. Gedächtnisstörung
	2. weitere kognitive Störung
	3. dadurch bedingte Beeinträchtigung im Alltag
	4. Verminderung der Affektkontrolle und des Antriebs oder verändertes Sozialverhalten
	5. Bestehen der kognitiven Symptome seit mehr als 6 Monaten
Ausschluss anderer Erkrankungen	Ausschluss von zum Beispiel zerebrovaskulären und endokrin-metabolischen Erkrankungen

Tabelle 2: DSM-IV-Kriterien für eine Alzheimer-Demenz [40]

Kriterium A1	Gedächtnisstörung
Kriterium A2	Weitere kognitive Störung A2a Aphasie A2b Apraxie A2c Agnosie A2d exekutive Störung
Kriterium B	Abfall der Leistungen von einem früher erreichten Niveau mit Behinderung der Alltagsaktivität
Kriterium C	Schleichender Beginn und langsame Zunahme
Kriterium D	Die Leistungsminderung ist nicht bedingt durch: D1 andere Erkrankungen des zentralen Nervensystems D2 systemische Erkrankungen D3 Substanzmissbrauch
Kriterium E	Die Leistungsminderung ist nicht nur im Rahmen eines Delirs vorhanden.
Kriterium F	Die Symptome lassen sich nicht durch eine andere psychiatrische Erkrankung erklären, zum Beispiel eine Depression oder Schizophrenie.

Die ICD-10 und das DSM-IV definieren 2 Subtypen der AD: die AD mit frühem Beginn (Beginn der Demenz vor dem beziehungsweise bis zum Ende des 65. Lebensjahres) und die AD mit spätem Beginn (erste klinische Symptome ab dem beziehungsweise nach dem 65. Lebensjahr). Weitere Kriterien wurden bereits 1984 durch die National Institute of Neurologic, Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association Work Group (NINCDS-ADRDA) veröffentlicht [8,41]. Bei der Diagnose einer „wahrscheinlichen AD“ werden eine Sensitivität von 81 % und eine Spezifität von 70 % angegeben. Für die Diagnose einer „möglichen AD“ werden Sensitivität und Spezifität mit 83 % bzw. 48 % geschätzt [42]. Allerdings werden auch niedrigere Werte für die Spezifität angegeben, etwa für die Unterscheidung von AD und frontotemporalen Demenzen (Sensitivität 93 %, Spezifität 23 %) [43]. Die Kriterien aus dem Jahr 1984 sind in Tabelle 3 dargestellt.

Im Jahre 2009 haben das National Institute on Aging (NIA) und die Alzheimer's Association (AA) eine Arbeitsgruppe gegründet, um die 1984 aufgestellten NINCDS-ADRDA-Kriterien zu überarbeiten. Die Publikationsserie zu dieser Überarbeitung [28,33,44-46] erschien 2011 und berücksichtigte neuere Erkenntnisse zu unterschiedlichen Aspekten der Krankheit [47]. Obwohl mittlerweile zu vielen Subtypen der Alzheimer-Demenz neuere Kriterien vorliegen, sind die NINCDS-ADRDA-Kriterien von 1984 nach wie vor relevant, da sich auch jüngste Studien noch auf diese älteren Kriterien beziehen.

Tabelle 3: Kriterien der National Institute of Neurologic, Communicative Disorders and Stroke-Alzheimer's Disease and related Disorders Association Work Group (NINCDS-ADRDA) [41]

<p style="text-align: center;"><u>Klinisch mögliche Alzheimer-Demenz („possible AD“)</u></p> <ul style="list-style-type: none">▪ demenzielles Syndrom mit atypischer Symptomatik oder atypischem Verlauf ohne erkennbare andere neurologische oder internistische Demenzursache▪ demenzielles Syndrom mit gleichzeitig vorliegender anderer Erkrankung, die auch eine Demenz erzeugen kann, in diesem Fall aber nicht als entscheidende Ursache angesehen wird▪ progredientes Defizit in nur einem kognitiven Bereich
<p style="text-align: center;"><u>Klinisch wahrscheinliche Alzheimer-Demenz („probable AD“)</u></p> <p>I. notwendige Voraussetzungen:</p> <ul style="list-style-type: none">▪ Zeichen einer Demenz in klinischer Untersuchung und bei neuropsychologischen Tests (zum Beispiel MMSE)▪ Defizite in 2 oder mehr kognitiven Bereichen▪ fortschreitende Verschlechterung des Gedächtnisses und anderer kognitiver Funktionen▪ keine Bewusstseinstörung▪ Beginn zwischen dem 40. und 90. Lebensjahr <p>II. Unterstützende Befunde:</p> <ul style="list-style-type: none">▪ Fortschreitende Verschlechterung der Sprache (Aphasie), Motorik (Apraxie) und Wahrnehmung (Agnosie)▪ Beeinträchtigung des Alltagslebens und Verhaltensänderungen▪ positive Familienanamnese für Alzheimer-Demenz, besonders falls neuropathologisch gesichert▪ Normalbefund einer Liquoranalyse, unspezifische EEG-Veränderungen, CCT-gesicherte Progression einer zerebralen Atrophie <p>III. Mit der Diagnose einer wahrscheinlichen Alzheimer-Demenz vereinbar:</p> <ul style="list-style-type: none">▪ Plateaus im Krankheitsverlauf▪ Begleitsymptome wie Depression, Schlaflosigkeit, Inkontinenz, Wahn, Verkennungen, Halluzinationen, „katastrophisierende Reaktionen“, Störungen des Sexualverhaltens, Gewichtsverlust▪ Besonders bei fortgeschrittener Erkrankung: erhöhter Muskeltonus, Myoklonien, Gangstörungen, Krampfanfälle▪ Normales CCT <p>IV. Befunde und anamnestische Angaben, die die Diagnose einer Alzheimer-Demenz unwahrscheinlich erscheinen lassen:</p> <ul style="list-style-type: none">▪ plötzlicher Beginn (apoplexartig)▪ früh auftretende fokal-neurologische Ausfälle: Hemiparesen, Gesichtsfeldausfälle, Ataxien▪ früh auftretende Krampfanfälle und Gangstörungen
<p style="text-align: center;"><u>Definitive Diagnose einer Alzheimer-Demenz („definite AD“)</u></p> <ul style="list-style-type: none">▪ Kriterien einer klinisch wahrscheinlichen Alzheimer-Demenz und▪ histopathologische Belege mittels Biopsie oder Autopsie

Seit der Publikation der NINCDS-ADRDA-Kriterien aus 1984 [41] gab es Empfehlungen, die Definition des „sicheren AD“ auszudehnen. So haben Dubois et al. vorgeschlagen, dass klinische und genetische Belege einer AD für eine „sichere“ AD-Diagnose schon zu Lebzeiten ausreichend seien [48]. Eine Reihe von kritischen Arbeiten zur Gewichtung der postmortalen neuropathologischen Begutachtung des Gehirns in der NINCDS-Diagnose der sicheren AD folgte [49-51], darunter auch Studien, die zeigten, dass mit zunehmendem Alter die Korrelation zwischen dem Vorhandensein von AD-typischen neuropathologischen Läsionen und klinischer Demenz abnimmt [52]. 2011 wurde empfohlen, der Neuropathologie eine bescheidenere Rolle in der AD-Diagnose beizumessen [50]. Sogar die Empfehlungen der Autoren der NINCDS-ADRDA-Kriterien zeigen in diese Richtung: Symptome, die 1984 als „definite AD“ bezeichnet wurden, werden in den NIA-AA-Empfehlungen aus 2011 lediglich als „pathophysiologisch bestätigte AD“ klassifiziert [45].

1.5.2 Biomarker

Unter Biomarkern versteht man üblicherweise die Bestimmung (bio-)chemischer Substanzen bzw. Merkmale, um einen Aufschluss über Körperprozesse zu erhalten. Neuere Definitionen beinhalten auch die Messung bzw. Darstellung von Organsfunktionen bzw. -strukturen [53]. Diese Definition schließt somit auch die Ergebnisse bildgebender Verfahren ein und ist diesem Bericht zugrunde gelegt.

In Abbildung 2 ist der mögliche Einsatz verschiedener Biomarker zur Diagnose der verschiedenen Stadien der Alzheimer-Demenz dargestellt [28]. Dabei ist zu berücksichtigen, dass diese Marker nicht Alzheimer-spezifisch sind. So könnte die Erhöhung von Tau-Proteinen auch auf neuronale Verletzungen hindeuten. Ein positiver Befund, der anhand von Biomarkern gestellt wurde, stellt somit nicht automatisch einen „richtig-positiven“ Befund dar und sollte daher noch in weiteren Tests bestätigt bzw. widerlegt werden.

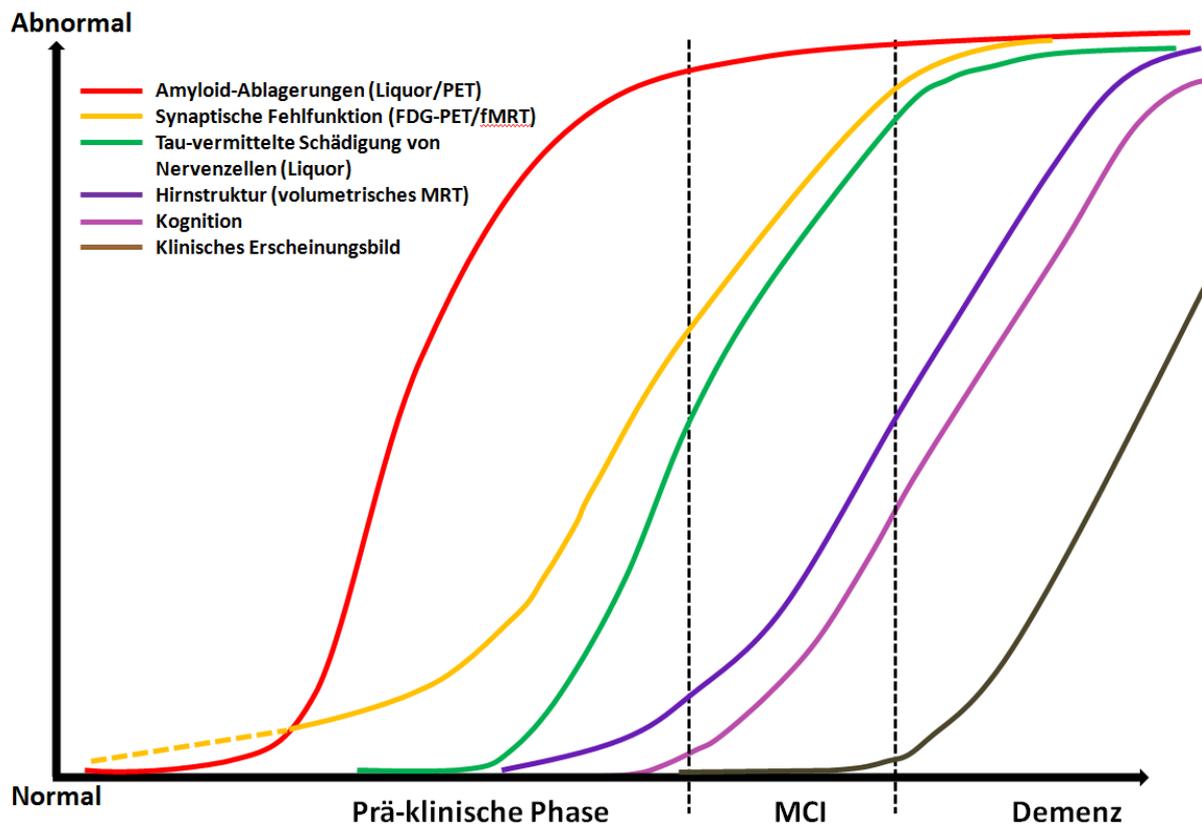


Abbildung 2: Biomarker zur Diagnose der Alzheimer-Demenz (modifiziert nach Sperling et al. 2011 [28]). Die x-Achse stellt in dieser Abbildung den Krankheitsfortschritt dar, die y-Achse die Ausprägung des entsprechenden Biomarkers. Die gestrichelte Linie weist darauf hin, dass bei bestimmten Patienten mit genetischen Veranlagungen synaptische Fehlfunktionen durch FDG-PET bzw. fMRT eher dargestellt werden könnten als Amyloid-Ablagerungen im Liquor.

Aus der Abbildung ist ersichtlich, dass verschiedene Marker in den unterschiedlichen Stadien der Erkrankung eingesetzt werden können, um eine Diagnosestellung zu ermöglichen bzw. zu erleichtern [28].

Im Folgenden sind mögliche Anwendungsgebiete für eine PET bei AD im präklinischen Stadium sowie der leichten kognitiven Beeinträchtigung (MCI) beschrieben.

Insbesondere im präklinischen Stadium der Erkrankung könnten PET bzw. PET/CT eingesetzt werden, um eine asymptotische zerebrale Amyloidose zu diagnostizieren. FDG-PET bzw. PET/CT könnte zudem eine Erkennung von synaptischen Fehlfunktionen ermöglichen [28]. Einige Autoren vermuten, dass der Nachweis von Amyloid-Ablagerungen, ggf. mit gleichzeitiger Diagnose neurologischer Degeneration, prognostisch mit kognitivem Abbau assoziiert ist [54,55]. PET bzw. PET/CT können auch zur Überprüfung eines Behandlungsansprechens (sog. Therapiemonitoring) verwendet werden. Da jedoch die Validität dieses Surrogatendpunktes bislang nicht untersucht ist, können aus Studien, die beim Therapiemonitoring alleine

die PET einsetzen, keine Aussagen zum patientenrelevanten Nutzen der untersuchten Therapien abgeleitet werden.

Die Diagnose einer leichten kognitiven Beeinträchtigung (MCI) wird vorwiegend klinisch und nach Ausschluss weiterer Ursachen, wie z. B. Depressionen, Flüssigkeitsmangel oder altersassoziierter kognitiver Abbau, gestellt. Dabei könnten Biomarker und bildgebende Verfahren eine Rolle spielen [28]. Allerdings ist eine allgemein anerkannte Definition der MCI nicht vorhanden [56].

Wie zuvor beschrieben, ist der Übergang zwischen den verschiedenen Stadien nicht klar abgrenzbar. Ein Vorliegen von frühen Stadien ist mit einer erhöhten Wahrscheinlichkeit einer AD verbunden, muss jedoch nicht zwangsläufig darin übergehen. Zudem berichtet eine Übersichtsarbeit, die Daten aus 50 Studien zu 5620 an Demenz erkrankten Personen zusammenfasst, eine Prävalenz von 9 % für potenziell reversible Demenzerkrankungen [57].

Die Autoren der kürzlich veröffentlichten Empfehlungen des „National Institute on Aging“ und der „Alzheimer’s Association workgroup“ weisen darauf hin, dass weitere Studien nötig sind, um die Bedeutung der Biomarker, ggf. im Zusammenhang mit anderen Testresultaten, zu verstehen und ihren Einsatz in der Praxis zu verbessern [30,35,58]. Dies wird auch von anderen Autoren gefordert [59].

1.6 PET und PET/CT

PET und PET/CT sind diagnostische Verfahren, die freigesetzte Photonen registrieren. Dazu werden radioaktive Substanzen (sogenannte Tracer) über eine Vene appliziert. Durch den Zerfall des Tracers werden Positronen freigesetzt. Diese kollidieren mit Elektronen, die im Körper vorhanden sind. Mittels ringförmig um den Patienten angeordneter Photonen-detektoren werden die dabei diametral entgegengesetzt freigesetzten Photonen erfasst. Aus der räumlichen Verteilung der registrierten Photonen wird auf die räumliche Verteilung der radioaktiven Substanz (des sogenannten Tracers) im Körperinneren geschlossen.

In der Anfangszeit der PET wurden sogenannte Nichtvollringsysteme hergestellt. Die Empfindlichkeit der später entwickelten Vollring-PET-Systeme war den ursprünglichen so stark überlegen, dass nur diese neueren Systeme eine breite Anwendung fanden. Die ersten Vollring-PET-Systeme hatten nur einen Detektorring; bei modernen Systemen werden heute mehrere Ringe nebeneinander angeordnet, wodurch die Empfindlichkeit des Gesamtsystems noch weiter gesteigert wurde.

Im Gegensatz zu anderen bildgebenden Verfahren, bei denen Informationen über die anatomische Struktur und damit die Lokalisation von Läsionen erhoben werden (sogenannte morphologische Bildgebung), kann die PET komplementär dazu Funktion und Metabolismus der Gewebe charakterisieren (sogenannte metabolische Bildgebung). Dadurch können biochemische und physiologische Prozesse sichtbar gemacht werden, weswegen man die PET zu den Verfahren der funktionellen Bildgebung zählt.

Als Tracer finden unterschiedliche radioaktiv markierte Substanzen Verwendung, die bei den verschiedensten Stoffwechselfvorgängen eine Rolle spielen. Je nach Fragestellung variiert die Auswahl des Tracers. Zur Diagnostik der Alzheimer-Demenz wird üblicherweise Fluor-18-2-Fluoro-2-Deoxy-D-Glukose (FDG) eingesetzt, um eine durch neuropathologische Prozesse bedingte Abnahme des Glukosestoffwechsels darzustellen (siehe auch Abbildung 1) [60,61]. Anreicherungen neuerer Substanzen, wie Pittsburgh compound-B (PIB) und 18F-markiertes Fluorbetapir, sollen eine Darstellung von Amyloid-Ablagerungen ermöglichen [62].

1.6.1 Empfehlungen wissenschaftlicher Fachgesellschaften zur PET bzw. PET/CT bei Alzheimer-Demenz

Die 2008 veröffentlichte S3-Leitlinie „Demenz“ der Deutschen Gesellschaft für Allgemeinmedizin (DEGAM) sieht „nach dem heutigen Wissensstand bzgl. der Demenzdiagnose keinen zusätzlichen Nutzen“ für die Positronenemissionstomographie (PET) bzw. die Single-Photon-Emissions-Computertomographie (SPECT) [19]. Hierbei folgt die DEGAM den Empfehlungen der *American Academy of Neurology* [42].

Eine weitere, im November 2009 veröffentlichte S3-Leitlinie befasst sich mit „Demenzen“. Diese wurde von der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) und der Deutschen Gesellschaft für Neurologie (DGN) in Zusammenarbeit mit der Deutschen Alzheimer Gesellschaft e. V. – Selbsthilfe Demenz herausgeben [59]. Danach können „FDG-PET und HMPAO-SPECT bei Unsicherheit in der Differenzialdiagnostik von Demenzen zur Klärung beitragen“. Einen regelhaften Einsatz in der Diagnostik empfiehlt die Leitlinie nicht. Die Empfehlung basiert auf einer Adaptation der englischen *National Clinical Practice Guideline Number 42*, die vom *National Collaborating Centre for Mental Health* nach Beauftragung durch das *Social Care Institute for Excellence National Institute for Health and Clinical Excellence* erstellt wurde und 2007 veröffentlicht wurde [8]. Dieser Bericht gibt eine ähnliche Empfehlung ab (vgl. S. 154-155 des Reports).

Die 2010 veröffentlichten Empfehlungen der *European Federation of Neurological Sciences (EFNS)* sehen den Nutzen der PET darin, die klinische Diagnosestellung bzw. die Ergebnisse anderer bildgebender Verfahren zu ergänzen, um eine höhere Spezifität zu erzielen [63]. Zudem weisen andere Autoren auf den Nutzen der im Vergleich zur Spezifität (63 %) höheren Sensitivität (> 90 %) zur Erkennung von Demenzen hin. [64]. Zusammenfassend empfiehlt die EFNS den Einsatz von FDG PET für Patienten, bei denen nach klinischer Diagnostik und klassischen bildgebenden Verfahren die Diagnose ungesichert ist [63].

2 Ziele der Untersuchung

Die vorliegende Untersuchung soll als Erstes prüfen, welcher patientenrelevante Nutzen oder Schaden mit der bildgebenden Diagnostik mittels PET bzw. PET/CT im Vergleich oder in Ergänzung zur konventionellen Diagnostik bei der Differenzialdiagnostik manifester Demenzen im Hinblick auf die Detektion von Alzheimer-Demenz, die (frühere) Detektion einer Alzheimer-Demenz oder ein Therapiemonitoring verbunden ist (Fragestellung 1).

Falls dies aufgrund der vorliegenden Datenlage nicht möglich ist, weil keine oder zu wenige aussagekräftige Studien vorliegen, soll als Zweites untersucht werden, ob sich durch den Einsatz der PET bzw. PET/CT die prognostische bzw. diagnostische Güte der Diagnosestrategie verbessert (Fragestellung 2).

Folgende Technologievergleiche werden betrachtet:

- Diagnostische Strategie mit PET versus diagnostische Strategie ohne PET (einzelne Verfahren, ihre Kombination) oder keine weitere Diagnostik.
- Diagnostische Strategie mit PET/CT versus diagnostische Strategie ohne PET/CT (einzelne Verfahren, ihre Kombination) oder keine weitere Diagnostik.
- Diagnostische Strategie mit PET/CT versus diagnostische Strategie mit PET.

Unter konventioneller diagnostischer Strategie wird im vorliegenden Bericht eine Kombination verschiedener Prozeduren wie anamnestische Angaben (durch den Betroffenen bzw. durch Angehörige bzw. Pflegepersonen), körperliche, neuropsychologische und kognitive Untersuchungen, Labortests, EEG sowie bildgebende Verfahren wie CT, MRT, verstanden, die beitragen, die Diagnose anhand aktueller Klassifikationen (DSM-4, ICD-10 oder NINCDS-ADRDA) zu stellen. Es sollen aber darüber hinaus alle weiteren in den gefundenen Studien angestellten diagnostischen Vergleiche dargestellt werden.

Es erfolgt keine Eingrenzung der Fragestellungen auf bestimmte Tracer. Alle Fragestellungen werden auch im Rahmen der deutschen Strahlenschutzgesetzgebung erörtert.

Allgemeine Voraussetzungen für Studien zur Nutzenbewertung („Nutzenstudien“)

Unter „patientenrelevantem Nutzen und Schaden“ (im Weiteren kurz als „Nutzen und Schaden“ bezeichnet) werden Veränderungen verstanden, die für den Patienten fassbare Konsequenzen haben, wie z. B. die Auswirkung auf die Mortalität, Morbidität und gesundheitsbezogene Lebensqualität. Mit dem Begriff „Nutzen“ werden kausal begründete positive Effekte, mit dem Begriff „Schaden“ kausal begründete negative Effekte einer medizinischen Intervention auf patientenrelevante Endpunkte (siehe Abschnitt 4.1.2.3) bezeichnet. Der Begriff „Nutzenbewertung“ bezeichnet den gesamten Prozess der Evaluation medizinischer Interventionen hinsichtlich ihrer kausal begründeten positiven und negativen Effekte im Vergleich mit einer klar definierten anderen Intervention, einem Placebo oder keiner Intervention [65].

Allgemeine Voraussetzungen für Studien zur Bewertung der diagnostischen und prognostischen Güte

Falls ausreichend Evidenz zum patientenrelevanten Nutzen und Schaden vorliegt, ist eine Nutzenbewertung alleine anhand der Beantwortung dieser Fragestellung möglich. In diesem Fall ist die Nutzenbewertung abgeschlossen. Falls jedoch keine ausreichende Evidenz zur Bewertung des patientenrelevanten Nutzens und Schadens vorliegt, erfolgt eine Bewertung der diagnostischen und der prognostischen Güte der PET bzw. der PET/CT.

Unter „diagnostischer Güte“ wird eine korrekte Zuordnung der Patienten zum jeweiligen Gesundheitszustand verstanden, unter „prognostischer Güte“ eine korrekte Vorhersage der patientenrelevanten Zielgrößen.

Indikationen für den Einsatz der PET bzw. PET/CT laut Konkretisierung des G-BA

Im Folgenden werden die laut Konkretisierung des G-BA festgelegten Indikationen für den Einsatz der PET und PET/CT dargestellt. Im Rahmen des vorliegenden Berichts soll sowohl der patientenrelevante Nutzen (Fragestellung 1) als auch die diagnostische und prognostische Güte (Fragestellung 2) für diese Indikationen untersucht werden:

- Diagnose einer Alzheimer-Demenz (insbesondere Genauigkeit und Zeitpunkt der Diagnostik). Insbesondere folgende Unterindikationen:
 - Diagnostische Abgrenzung einer Alzheimer-Demenz von anderen demenziellen Erkrankungen (z. B. vaskuläre Demenz)
 - Diagnostische Abgrenzung einer Alzheimer-Demenz von anderen nicht-demenziellen Erkrankungen (z. B. Depression)
 - Frühzeitige Abgrenzung von Patienten mit Risiko für Alzheimer-Demenz (z. B. Patienten mit MCI oder SMI), die eine AD entwickeln
- Monitoring einer Therapie einer Alzheimer-Demenz

3 Projektbearbeitung

3.1 Zeitlicher Verlauf des Projekts

Der Gemeinsame Bundesausschuss (G-BA) hat mit Schreiben vom 21.12.2006 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Recherche, Darstellung und Bewertung des aktuellen medizinischen Wissensstandes zur PET und PET/CT bei 14 verschiedenen Krankheitsbildern beauftragt. Der vorliegende Berichtsplan bezieht sich auf den Teilauftrag zur Diagnose von Alzheimer-Demenz.

Die endgültigen Konkretisierungsunterlagen wurden dem Institut mit dem Schreiben des G-A vom 19.06.2008 übersandt. In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Während der Erstellung des Berichtsplans wurden am 03.05.2011 Patientenvertreter der Deutschen Alzheimer Gesellschaft e. V. zur Festlegung patientenrelevanter Zielgrößen konsultiert. Die Selbstangaben der Patientenvertreter zu potenziellen Interessenkonflikten sind in Anhang A dargestellt.

Der vorläufige Berichtsplan in der Version 1.0 vom 09.01.2012 wurde am 18.01.2012 auf der Website des IQWiG veröffentlicht und zur Anhörung gestellt. Bis zum 15.02.2012 konnten schriftliche Stellungnahmen eingereicht werden. Unklare Aspekte aus den schriftlichen Stellungnahmen zum vorläufigen Berichtsplan wurden am 16.04.2012 in einer wissenschaftlichen Erörterung mit den Stellungnehmenden diskutiert. Die Dokumentation und Würdigung der Anhörung zum Berichtsplan ist auf der Website des IQWiG veröffentlicht.

Der vorliegende Berichtsplan beinhaltet die Änderungen, die sich aus der Anhörung ergeben haben.

Auf Basis des Berichtsplans wird die vorläufige Bewertung vorgenommen. Diese wird in einem Vorbericht veröffentlicht, der zur Anhörung gestellt wird. Der Vorbericht wird zusätzlich einem externen Review unterzogen. Im Anschluss an die Anhörung zum Vorbericht erstellt das IQWiG einen Abschlussbericht. Dieser Bericht wird an den G-BA übermittelt und 8 Wochen später auf der Website des IQWiG veröffentlicht. An selber Stelle wird auch die Dokumentation und Würdigung der Anhörung zum Vorbericht veröffentlicht.

3.2 Dokumentation der Änderungen im Projektverlauf

Berichtsplan im Vergleich zum vorläufigen Berichtsplan

- Der Abschnitt 1.6.1 zu den Empfehlungen wissenschaftlicher Fachgesellschaften wurde auf Basis der in der Anhörung angesprochenen Aspekte präzisiert und ergänzt.
- Die Einschlusskriterien wurden auf Patienten mit erhöhtem Risiko für bzw. Verdacht auf eine Alzheimer-Demenz und Patienten mit einer diagnostizierten Demenz (Einsatz der PET bzw. PET/CT zur Differenzialdiagnostik) erweitert. Diese Erweiterung gilt für beide Berichtsfragestellungen (vgl. 4.1.2.1 und 4.2.2.1).

- Die Methoden zur Informationssynthese diagnostischer Studien wurden gemäß dem aktuellen Stand der Forschung zu bivariaten Meta-Analysen angepasst (4.5.2).
- Der Abschnitt 4.1.2.3 zu patientenrelevanten Endpunkten wurde bezüglich Managementänderungen und des Therapiemonitorings spezifiziert.
- Der Abschnitt 4.2 wurde überarbeitet. Bezüglich der Recherche nach Studien zur diagnostischen und prognostischen Güte wurde eine Einschränkung auf die einschlägigen Publikationssprachen Englisch, Deutsch sowie Französisch vorgenommen. Aufgrund zahlreicher bereits durchgeführter Nutzenbewertungen der PET und PET/CT (IQWiG-Projekte D06-01A–K) wird hierdurch kein relevanter Informationsverlust erwartet. Des Weiteren wurde die Formulierung des Einschlusskriteriums „Zielgrößen“ für Studien zur diagnostischen Güte in Bezug auf die notwendige Darstellung „geeigneter statistischer Maßzahlen“ verallgemeinert.

4 Methoden

In der Vergangenheit verband man mit dem Denken über die Bewertung von diagnostischen Tests primär die Beschreibung der technischen Merkmale des Tests und die Evaluation seiner diagnostischen Güte, d. h. seiner Fähigkeit, kranke von gesunden Personen zu unterscheiden [66]. Damit verband man die Erwartung, dass sich eine bessere diagnostische Güte automatisch in eine bessere Patientenversorgung umsetzen würde. Seitdem hat sich das Denken über den Stellenwert diagnostischer Tests im therapeutischen Management weiterentwickelt, wobei neben der verbesserten diagnostischen Güte die Konsequenzen, die aus dem diagnostischen Test für den Arzt, den Patienten und die Gesellschaft resultieren, in das Zentrum der Betrachtung rücken [67,68]. Fryback und Thornbury haben dazu in ihrer Publikation „The efficacy of diagnostic imaging“ von 1991 [67] 6 Stufen identifiziert, die von den technischen Auswirkungen („technical efficacy“) als unterste Stufe über die diagnostische Genauigkeit („diagnostic accuracy efficacy“) als 2. Stufe bis hin zu den therapeutischen Auswirkungen („therapeutic efficacy“) als 4. Stufe und den patientenrelevanten Auswirkungen („patient outcome efficacy“) als 5. Stufe reichen. Als oberste Stufe betrachten sie schließlich die gesellschaftlichen Konsequenzen („societal efficacy“) der diagnostischen Technologie (siehe auch Tabelle 4).

Tabelle 4: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren

	Köbberling et al. [68]	Fryback & Thornbury [67]	Evidenzklassifizierung G-BA [69]
Patientenrelevante Zielgrößen	Phase 4: Wirksamkeit	6. Stufe: Auswirkung auf Systemebene 5. Stufe: Auswirkung im Hinblick auf patientenrelevante Endpunkte	Evidenzstufe I
Surrogatzielgröße (Management-änderung)		4. Stufe: Auswirkung im Hinblick auf das therapeutische Denken des Behandelnden 3. Stufe: Auswirkung im Hinblick auf das (differenzial)diagnostische Denken („Nachtstwahrscheinlichkeit“)	Evidenzstufe II
Sensitivität (SN), Spezifität (SP), Likelihood Ratio (LR), prädiktive Werte	Phase 3: Diagnostische Genauigkeit bei nicht bekanntem Krankheitsstatus	2. Stufe: Auswirkung im Hinblick auf Diskriminationsfähigkeit	Evidenzstufe III
Sensitivität (SN), Spezifität (SP), Likelihood Ratio (LR)	Phase 2: Diagnostische Genauigkeit bei bekanntem Krankheitsstatus		
Analytische Sensitivität, Spezifität, Reproduzierbarkeit	Phase 1: Technische Voruntersuchungen	1. Stufe: Technische Auswirkungen	Evidenzstufe IV

Diese erweiterte Betrachtungsweise findet zunehmend ihren Niederschlag in der klinischen Forschung, wo vermehrt die Auswirkungen eines neuen Tests auf die Entscheidungsfindung des Arztes, auf eine verbesserte Behandlung für den Patienten und auf die Konsequenzen für die Gesellschaft empirisch untersucht werden. Die Mehrzahl der westlichen Gesundheitssysteme zieht Überlegungen zu den Konsequenzen und dem Nutzen und Schaden diagnostischer Tests für Arzt, Patient und Gesellschaft in Betracht, wenn sie über den Wert eines diagnostischen Verfahrens nachdenkt. Auch der G-BA hat sich in seiner Verfahrensordnung bei der Bewertung diagnostischer Tests (2. Kapitel, § 11 Abs. 2) dieser Sichtweise angeschlossen [69]. In Tabelle 4 sind verschiedene Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren einander gegenübergestellt. Neben der Einteilung von Fryback und Thornbury und der Evidenzklassifizierung des G-BA wurde auch das Phasenschema nach Köbberling et al. [68] mit aufgenommen.

Tabelle 5: Evidenzklassifizierung des G-BA für Unterlagen zu diagnostischen Methoden

Evidenzklasse nach G-BA [69]	Kriterien
I a	Systematische Übersichtsarbeiten von Studien der Evidenzstufe I b
I b	Randomisierte kontrollierte Studien
I c	Andere Interventionsstudien
II a	Systematische Übersichtsarbeiten von Studien zur diagnostischen Testgenauigkeit der Evidenzstufe II b
II b	Querschnitt- und Kohortenstudien, aus denen sich alle diagnostischen Kenngrößen zur Testgenauigkeit (Sensitivität und Spezifität, Wahrscheinlichkeitsverhältnisse, positiver und negativer prädiktiver Wert) berechnen lassen
III	Andere Studien, aus denen sich die diagnostischen Kenngrößen zur Testgenauigkeit (Sensitivität und Spezifität, Wahrscheinlichkeitsverhältnisse) berechnen lassen
IV	Assoziationsbeobachtungen, pathophysiologische Überlegungen, deskriptive Darstellungen, Einzelfallberichte u. Ä.; nicht mit Studien belegte Meinungen anerkannter Experten, Berichte von Expertenkomitees und Konsensuskonferenzen

4.1 Ermittlung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT

Um den patientenrelevanten Nutzen und Schaden der bildgebenden diagnostischen Technologien PET bzw. PET/CT für die Diagnose der Alzheimer-Demenz zu bewerten, genügt es nicht festzustellen, inwieweit die PET bzw. die PET/CT den gängigen diagnostischen Methoden überlegen ist. Vielmehr geht es bei einer erweiterten Bewertung darum festzustellen, inwieweit sich eine präzisere Diagnosestellung auf die Behandlung und das weitere Management des Patienten auswirkt und inwieweit, in letzter Konsequenz, das

veränderte Management patientenrelevante Endpunkte wie z. B. Mortalität, Morbidität oder die Lebensqualität beeinflusst.

Diese Fragen lassen sich beispielsweise durch Studien untersuchen, in denen die Patienten z. B. je nach PET-Befund randomisiert entweder der bisherigen, von der PET unabhängigen Standardtherapie oder einer aufgrund des PET-Befundes adaptierten Therapie zugeteilt werden (sogenannte „Managementstudien“). Aber auch andere Studiendesigns kommen infrage [70]. Wie sich die Behandlungsoptionen verändern und welche Auswirkungen die unterschiedlichen Strategien schließlich auf patientenrelevante Endpunkte haben, lässt sich am besten durch Studien mit ausreichender Nachbeobachtungszeit beantworten (siehe auch Abschnitt 4.1.2.3).

4.1.1 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT anhand von systematischen Übersichten

Im Folgenden werden die Begriffe Sekundärliteratur, systematische Reviews oder HTA-Berichte unter dem Begriff „systematische Übersichten“ zusammengefasst.

Vorrecherchen haben ergeben, dass vermutlich keine qualitativ ausreichende systematische Übersicht zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT bei der Diagnose der Alzheimer-Demenz vorliegt. Daher soll der patientenrelevante Nutzen in dem zu erstellenden Bericht nicht auf Basis systematischer Übersichten, sondern anhand von Primärstudien untersucht werden.

4.1.2 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT anhand von Primärstudien

4.1.2.1 Populationen

Die Primärstudien müssen Angaben zu mindestens 10 Patienten mit erhöhtem Risiko für bzw. Verdacht auf eine Alzheimer-Demenz oder Patienten mit einer diagnostizierten und nicht näher differenzierten Demenz (Einsatz der PET zur Differenzialdiagnostik) machen.

4.1.2.2 Prüf- und Vergleichsinterventionen

Als Prüfintervention gilt eine diagnostische Strategie mit Anwendung einer Vollring-PET bzw. einer Vollring-PET/CT (Indextest I). Als Vergleichsintervention (Indextest II) wird eine diagnostische Strategie ohne Anwendung der PET bzw. der PET/CT betrachtet. Ein weiterer Vergleich beinhaltet eine diagnostische Strategie mit Anwendung der Vollring-PET/CT als Prüfintervention und eine diagnostische Strategie mit Vollring-PET (ohne CT) als Indextest II. Folgende Vergleiche werden betrachtet:

- Diagnostische Strategie mit PET versus diagnostische Strategie ohne PET (einzelne Verfahren, ihre Kombination) oder keine weitere Diagnostik.
- Diagnostische Strategie mit PET/CT versus diagnostische Strategie ohne PET/CT (einzelne Verfahren, ihre Kombination) oder keine weitere Diagnostik.

- Diagnostische Strategie mit PET/CT versus diagnostische Strategie mit PET.

Verschiedene (randomisierte) Vergleiche nach vorheriger Diagnostik mit der PET bzw. PET/CT (z. B. Randomisierung der diskordanten Ergebnisse der PET und CT) sind ebenfalls möglich [70].

4.1.2.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevanten Endpunkte verwendet, die eine Beurteilung der Interventionsziele ermöglichen:

- Aktivitäten des täglichen Lebens (alltagspraktische Fähigkeiten),
- kognitive Leistungsfähigkeit,
- gesundheitsbezogene Lebensqualität,
- andere mit der Erkrankung verbundene Symptome, zum Beispiel Depression, Schlaf-Wach-Umkehr, Wahnhaftigkeit, Agitiertheit (begleitende psychopathologische Symptome),
- Notwendigkeit einer vollstationären oder institutionalisierten ambulanten Pflege (Institutionalisierung),
- Mortalität,
- unerwünschte Arzneimittelwirkungen,
- andere unerwünschte Wirkungen.

Subjektive Endpunkte (z. B. gesundheitsbezogene Lebensqualität) werden nur dann berücksichtigt, wenn sie mit validen Messinstrumenten (z. B. QOL-AD) erfasst wurden.

Folgende Zielgrößen werden ergänzend betrachtet:

- Lebensqualität der betreuenden Angehörigen,
- Betreuungsaufwand durch eine oder mehrere betreuende Person(en) oder Institution(en),
- interventions- und erkrankungsbedingter Aufwand,
- Zufriedenheit der Patienten mit der Behandlung,
- klinischer Eindruck (klinisches Globalurteil).

Ein Nutzen oder Zusatznutzen kann sich alleine auf Basis dieser ergänzend dargestellten Zielgrößen jedoch nicht ergeben.

Managementänderungen

Eine Änderung im Patientenmanagement alleine (ohne Verknüpfung zu den oben genannten patientenrelevanten Endpunkten) stellt keine Zielgröße im Sinne dieses Berichts dar und wird daher ebenfalls nur ergänzend betrachtet. In Ausnahmefällen kann eine Managementänderung alleine Patientenrelevanz entfalten, wenn folgende Bedingungen erfüllt sind:

- Sie wirkt sich unmittelbar auf patientenrelevante Endpunkte aus. Beispiele für solche Managementänderungen können die Vermeidung von unwirksamen chirurgischen Eingriffen und die Vermeidung unnötiger invasiver Diagnostik sein.
- Mit der Managementänderung dürfen bezüglich der oben genannten patientenrelevanten Endpunkte keine inakzeptablen Verschlechterungen einhergehen.

Für den Einschluss in den Bericht wird lediglich gefordert, dass Änderungen im Management gleichzeitig mit den patientenrelevanten Endpunkten Mortalität, Morbidität oder Lebensqualität erfasst werden.

PET-Parameter als Endpunkte (Therapiemonitoring)

Studien, in denen die PET zum Therapiemonitoring eingesetzt wird, verwenden häufig die PET-Ergebnisse als Endpunkte. Das Ansprechen auf eine Therapie wird daran festgemacht, dass sich bestimmte PET-Parameter in Abhängigkeit von bestimmten therapeutischen Interventionen verändern. In diesen Fällen entspricht das Ergebnis der PET einem Surrogatendpunkt, dessen Validität bisher nicht untersucht wurde. Sofern in derartigen Studien neben den PET-Parametern auch patientenrelevante Endpunkte untersucht wurden, werden die Studien eingeschlossen.

4.1.2.4 Studientypen

Randomisierte kontrollierte Studien (RCTs) sind, sofern sie methodisch adäquat und der jeweiligen Fragestellung angemessen durchgeführt wurden, mit der geringsten Ergebnisunsicherheit behaftet. Sie liefern daher die zuverlässigsten Ergebnisse für die Bewertung des Nutzens einer medizinischen Intervention. Es werden deshalb primär RCTs betrachtet.

Ist die auf randomisierten kontrollierten Studien basierende Datenlage nicht hinreichend, um den patientenrelevanten Nutzen und Schaden der PET bzw. PET/CT mit ausreichender Ergebnissicherheit schätzen zu können (z. B. kleine Fallzahl, wenige Ereignisse [geringe Präzision], hohes Verzerrungspotenzial), werden gegebenenfalls auch verfügbare eindeutig prospektive, jedoch nicht randomisierte Interventionsstudien mit zeitlich parallelen Kontrollgruppen herangezogen (sogenannte controlled clinical trials: CCTs), sofern das Problem einer möglichen Strukturungleichheit (unfairer Vergleich) der Beobachtungsgruppen adäquat in der Planung und Auswertung der entsprechenden Studien berücksichtigt wurde. Solche Studien können zwar die Aussage von aggregierten Ergebnissen aus qualitativ belastbaren RCTs in der Regel nicht qualitativ ändern, diese aber ggf. bestärken.

Finden sich auch in der 2. Kategorie (CCTs) keine ausreichenden Daten, werden gegebenenfalls auch verfügbare eindeutig prospektive, jedoch nicht randomisierte Studien ohne Zuteilung zu den Diagnosegruppen durch die Untersuchungsleiter (Beobachtungsstudien) mit zeitlich parallelen Kontrollgruppen herangezogen, sofern das Problem einer möglichen Strukturungleichheit (unfairer Vergleich) der Beobachtungsgruppen adäquat in der Planung und Auswertung der entsprechenden Studien berücksichtigt wurde. Auch diese Studien können die Aussage von aggregierten Ergebnissen aus qualitativ belastbaren RCTs in der Regel nicht qualitativ ändern, diese aber ggf. bestärken.

Dieses Vorgehen scheint für die zugrunde liegende Fragestellung gerechtfertigt, da sich erst sehr langsam der Paradigmenwechsel in der Bewertung diagnostischer Verfahren hin zu vergleichenden Studien mit patientenrelevanten Endpunkten vollzieht. Zudem bilden möglicherweise nicht randomisierte kontrollierte Interventionsstudien – wie oben skizziert – einen Ausgangspunkt für die Ausgestaltung zukünftiger aussagekräftigerer Studien.

4.1.2.5 Sonstige Studiencharakteristika

Es erfolgt keine weitere Einschränkung.

4.1.3 Tabellarische Darstellung der Einschlusskriterien („Nutzenstudien“)

In die Nutzenbewertung werden alle Studien einbezogen, die alle nachfolgenden Einschlusskriterien erfüllen (siehe Tabelle 6).

Tabelle 6: Übersicht der Einschlusskriterien für Primärstudien zur Nutzenbewertung der PET bzw. PET/CT

Einschlusskriterien	
E1a	Mindestens 10 Patienten mit erhöhtem Risiko für bzw. Verdacht auf Alzheimer-Demenz oder Patienten mit diagnostizierten und nicht näher differenzierten Demenz (Einsatz der PET zur Differenzialdiagnostik; siehe auch Abschnitt 4.1.2.1)
E2a	Vergleich von 2 (oder mehr) diagnostischen Strategien wie z. B. PET versus Strategie ohne PET PET/CT versus Strategie ohne PET/CT PET/CT versus PET oder Vergleich von 2 (oder mehr) therapeutischen Strategien, die in Abhängigkeit vom PET- bzw. PET/CT-Befund eingesetzt werden. (siehe auch Abschnitt 4.1.2.2)
E3a	Patientenrelevante Endpunkte wie in Abschnitt 4.1.2.3 formuliert
E4a	Randomisierte kontrollierte Studien (Bei nicht hinreichend auf randomisierten kontrollierten Studien basierender Datenlage werden ggf. prospektive nicht randomisierte kontrollierte Studien [mit und ohne experimentelle Gruppenzuteilung], wie in Abschnitt 4.1.2.4 beschrieben, eingeschlossen.)
E5a	Keine Mehrfachpublikationen ohne relevante Zusatzinformationen
E6a	Vollpublikation verfügbar ^a
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch die nicht vertrauliche Weitergabe eines Studienberichts an das Institut oder die nicht vertrauliche Bereitstellung eines Berichts über die Studien, der den Kriterien des CONSORT-Statements [71] genügt und eine Bewertung der Studie ermöglicht. CONSORT: Consolidated Standards of Reporting Trials</p>	

4.2 Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT

In den orientierenden Vorrecherchen fanden sich systematische Übersichten, die Teilaspekte der Fragestellungen zur diagnostischen Güte der PET oder PET/CT bei Alzheimer-Demenz untersuchten. Mit dem Ziel einer möglichst raschen Ergebniserhebung und Nutzung bereits vorliegender wissenschaftlicher Evidenz auf höchster Evidenzstufe soll die Bewertung der diagnostischen Güte der PET und PET/CT primär auf Grundlage von systematischen Übersichten erfolgen, sofern sichergestellt ist, dass sie bestimmte Mindestanforderungen erfüllen. Findet sich keine geeignete systematische Übersicht zu einer Fragestellung, werden die diagnostische und die prognostische Güte anhand von Primärstudien beurteilt. Sind die identifizierten systematischen Übersichten nicht mehr aktuell, wird für den Zeitraum und für Fragestellungen, die sie nicht abdecken, relevante Primärliteratur recherchiert und im Bericht berücksichtigt (Ergänzungsrecherche).

Nachfolgend werden zunächst die Kriterien für einen Einschluss von systematischen Übersichten in die geplante Bewertung aufgeführt (Abschnitt 4.2.1). Anschließend erfolgt die Darstellung für einzuschließende Primärstudien („Ergänzungsrecherche“; Abschnitt 4.2.2).

4.2.1 Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Güte anhand von systematischen Übersichten

4.2.1.1 Populationen

Die den systematischen Übersichten zugrunde liegenden Primärstudien müssen Angaben zu mindestens 10 Patienten mit erhöhtem Risiko für bzw. Verdacht auf eine Alzheimer-Demenz oder Patienten mit einer diagnostizierten und nicht näher differenzierten Demenz (Einsatz der PET zur Differenzialdiagnostik) machen.

4.2.1.2 Prüftechnologien (Indextest I), Vergleichstechnologien (Indextest II) und Referenzstandard

Die den systematischen Übersichten zugrunde liegenden Primärstudien müssen die Angaben zur diagnostischen Güte anhand folgender Vergleiche ermittelt haben:

Prüftechnologie („Indextest I“) ist die Vollring-PET bzw. PET/CT. Als Vergleichstechnologie („Indextest II“) dient jegliche diagnostische Strategie ohne PET bzw. PET/CT („konventionelle Diagnostik“). Außerdem wird der Vergleich PET vs. PET/CT zugelassen (s. Abschnitt 4.1.2.2).

Indextest I und Indextest II müssen zu einem Referenzstandard (= der besten zur Verfügung stehenden diagnostischen Strategie) in Beziehung gebracht worden sein (direkter Vergleich), der für eine adäquate Interpretation der Ergebnisse unabhängig von den Indextests erhoben worden ist, da es anderenfalls zu Verzerrungen der Güte- und / oder Variabilitätsschätzer kommen kann (sogenannter Verifikationsbias oder Inkorporationsbias) [72,73].

Referenzstandard

Die Bewertung und Klassifizierung von AD im Rahmen wissenschaftlicher Studien ist eine schwierige Fragestellung, da kein akzeptierter Goldstandard für die Diagnose der Erkrankung existiert. Daher kann bei Studien zur diagnostischen Güte bzw. bei Studien, die auf einem Diagnosetest beruhen, nur ein Referenzstandard, d. h. der beste zur Verfügung stehende Test, benutzt werden. Durch das Fehlen eines Goldstandards kann die Güte der angewandten Referenzstandards nicht eingeschätzt werden.

Als der derzeit best-verfügbare Referenzstandard wird die Kombination verschiedener einzelner Tests angesehen [74]. Bei diesem gemischten bzw. zusammengesetzten Referenzstandard (engl. composite reference standard) wird der Krankheitsstatus nach vordefinierten und festgelegten (deterministischen) Regeln anhand der jeweiligen Versionen von ICD, DSM oder NINCDS-ADRDA bestimmt. Bei Vorformen der Alzheimer-Demenz (SMI, MCI, etc.)

sollte die Erhebung des Krankheitsstatus bei Einschluss in die Studie und nach einer Nachbeobachtung von durchschnittlich 18 Monaten durchgeführt werden.

Der postmortale Nachweis übermäßiger neurofibrillärer Verklumpungen und neuritischer Plaques, welcher zur Diagnosesicherung, etwa im Rahmen der NINCDS-ADRDA, eingesetzt werden kann [39], wird von neueren Arbeiten kritisch betrachtet [49-51]. In einer Arbeit wird berichtet, dass lediglich 40 % bis 57 % der untersuchten Gehirne von Patienten, die eine wahrscheinliche Alzheimer-Demenz aufwiesen, auch die neuropathologischen Befunde einer reinen AD aufwiesen [75]. Dabei bezogen sich die Autoren auf die Kriterien des US-amerikanischen National Institute on Aging und des Reagan Institute (NIA-RI) [43]. Aus diesen Gründen werden im Rahmen des Berichts nur klinische Diagnosen am lebenden Patienten als Referenzstandard herangezogen.

Zur Differenzialdiagnostik bei manifesten Demenzformen werden daher die Kriterien einer „klinisch wahrscheinlichen Alzheimer-Demenz“ („probable AD“) und / oder der genetischen Analyse ohne histopathologische Befunde herangezogen. Eine Nachbeobachtung ist dafür nicht erforderlich.

4.2.1.3 Zielgrößen

Es werden systematische Übersichten eingeschlossen, deren Ergebnisse auf Primärstudien beruhen, die relevante Angaben zu folgenden Zielgrößen enthalten:

- Beschreibung der diagnostischen Güte anhand geeigneter statistischer Maßzahlen. Dabei müssen patientenbezogene Daten zu den Indextests und zum Referenzstandard eindeutig ableitbar sein.
- Prognostische Aussagen hinsichtlich zukünftiger klinischer Ereignisse (z. B. Abnahme der kognitiven Leistungsfähigkeit oder der Aktivitäten des täglichen Lebens, erhöhte Pflegebedürftigkeit, Übergang zu höherem Stadium der Erkrankung) anhand geeigneter statistischer Maßzahlen (z. B. aus Kaplan-Meier-Schätzungen). Es sollten dazu adäquate Maßnahmen zur Berücksichtigung von Störgrößen durchgeführt worden sein.

4.2.1.4 Den systematischen Übersichten zugrunde liegende Studientypen

Berücksichtigt werden Ergebnisse systematischer Übersichten, die auf Daten beruhen, die anhand folgender Primärstudientypen erhoben wurden:

- Diagnostische Güte
Querschnitt- oder prospektive Kohortenstudien, ggf. mit klinischer Nachbeobachtung, durchgeführt in der Anwendungssituation (gemäß Phase 3 nach Köbberling et al. [68]).
- Prognostische Güte
Prospektive Kohortenstudien: Um die prognostische Güte der PET möglichst unverzerrt bestimmen zu können, soll eine Gruppe von Patienten, die zu einem bestimmten Zeitpunkt

rekrutiert und mit der PET untersucht wurde, im zeitlichen Verlauf untersucht werden (Longitudinalstudie). Dabei sind zum einen ein konsekutiver Einschluss der Patienten und die Dokumentation aller fehlenden Werte notwendig. Zum anderen muss die betreffende untersuchte Kohorte vollständig (unabhängig vom PET-Ergebnis und von der Durchführung einer Intervention) nachbeobachtet werden. Historische oder indirekte Vergleiche liefern keine verlässlichen Daten. Im Idealfall soll die weitere Behandlung dieser Patienten ohne das Wissen um die Ergebnisse der PET erfolgen (Verblindung).

Für die Interpretation der zuvor genannten prognostischen Zielgrößen ist ein Follow-up von durchschnittlich 18 Monaten entscheidend. Für die Beurteilung medizinischer Interventionen bei Alzheimer-Demenz, die die Krankheitsprogression reduzieren oder verlangsamen sollen, wird sowohl in einer Richtlinie der European Medicines Agency (EMA) [76] als auch in einem Konsensbeschluss einer europäischen Arbeitsgruppe [77] ein Studienzeitraum von 18 Monaten für notwendig erachtet. Auch anhand von diversen Langzeitstudien zum natürlichen Verlauf der AD [78-81] erscheint eine Nachbeobachtung von 18 Monaten sinnvoll, um die Krankheitsprogression im Rahmen von diagnostischen oder prognostischen Studien zu erfassen (siehe auch 4.2.1.2). Systematische Übersichten, die sowohl prospektive als auch retrospektive Studien enthalten, werden nur dann berücksichtigt, wenn eine separate Auswertung der in ihnen enthaltenen prospektiv geplanten Studien möglich ist.

- **Diskordanzstudien und VOPT-Design**

Ist die Datenlage aus Studien, welche die diagnostischen Vorgehensweisen mit dem Referenzstandard direkt vergleichen, unzureichend, können systematische Übersichten der o. g. Studientypen herangezogen werden, die lediglich die diskordanten Befunde zwischen den alternativen diagnostischen Vorgehensweisen untersuchen. Diskordanzstudien müssen so konzipiert sein, dass alle diskordanten Befunde analog zu den Studien zur diagnostischen und prognostischen Güte mit dem gleichen, validen Referenzstandard untersucht wurden (siehe Abschnitt 4.2.1.2). Eine Abwandlung der Diskordanzstudien stellt das VOPT-Design (Verification of Only Positive Testers) dar, in welchem alle Ergebnisse, die in mindestens einem der beiden Indextests positiv sind, mit dem Referenzstandard untersucht werden [82].

4.2.1.5 Sonstige Charakteristika der systematischen Übersichten

Neben diesen genannten Charakteristika müssen die systematischen Übersichten folgende Kriterien erfüllen:

- Es muss sich um Übersichten handeln, die auf systematische, reproduzierbare und transparente Weise Ergebnisse von Primärstudien zusammenfassend dargestellt und bewertet haben. Daher müssen die Übersichten auf dem Qualitätsindex von Oxman und Guyatt mindestens 5 von 7 möglichen Punkten erreichen [83,84].

- Die jeweilige Forschungsfrage muss im Format einer sog. PICO-Frage formuliert sein, die Population, Indextest I, ggf. Indextest II, Referenzstandard und Endpunkte explizit benennt, oder die PICO-Fragestellung lässt sich der Publikation eindeutig entnehmen.
- Ein möglicherweise in den Primärstudien zu konstatierender Verifikations- bzw. Inkorporationsbias (vgl. Abschnitt 4.2.1.2) soll adressiert sein.
- Die Ergebnisse der Qualitätsbewertung, die Kerninformationen über die Population und die Ergebnisse einer jeden Studie sollen transparent und vergleichbar abgebildet sein und die Daten sollen eindeutig mit den zugehörigen Studien verknüpft werden können.
- Des Weiteren werden zur Bewertung der diagnostischen und prognostischen Güte nur Publikationen in den folgenden Sprachen eingeschlossen: Englisch, Deutsch, Französisch.

4.2.1.6 Tabellarische Darstellung der Einschlusskriterien (systematische Übersichten zur diagnostischen und prognostischen Güte)

Es werden alle systematischen Übersichten einbezogen, die alle nachfolgenden Einschlusskriterien erfüllen (siehe Tabelle 7).

Tabelle 7: Übersicht der Einschlusskriterien für systematische Übersichten zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT

Einschlusskriterien	
E1b	Mindestens 10 Patienten mit erhöhtem Risiko für bzw. Verdacht auf Alzheimer-Demenz oder Patienten mit einer diagnostizierten und nicht näher differenzierten Demenz (Einsatz der PET zur Differenzialdiagnostik; siehe auch Abschnitt 4.1.2.1)
E2b	Prüftechnologie (Vollring-PET bzw. PET/CT), Vergleichstechnologie und Referenzstandard wie in Abschnitt 4.2.1.2 beschrieben
E3b	Geeignete statistische Maßzahlen (patientenbezogene Daten zu Indextests und Referenzstandard müssen eindeutig ableitbar sein; siehe auch Abschnitt 4.2.1.3)
E4b	Systematische Übersichten aus Querschnitt- oder prospektiven Kohortenstudien wie in den Abschnitten 4.2.1.4 und 4.2.1.5 beschrieben
E5b	Keine Mehrfachpublikationen ohne relevante Zusatzinformationen
E6b	Vollpublikation verfügbar ^a
a: Als Vollpublikation gilt in diesem Zusammenhang auch die nicht vertrauliche Weitergabe eines Studienberichts an das Institut oder die nicht vertrauliche Bereitstellung eines Berichts über die Studien, der den Kriterien des PRISMA-Statements [85] oder des MOOSE-Statements [86] genügt und eine Bewertung der Studie ermöglicht. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis; MOOSE: Meta-analysis Of Observational Studies in Epidemiology	

4.2.2 Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Güte anhand von Primärstudien („Ergänzungsrecherche“)

Im Rahmen einer Ergänzungsrecherche wird in Ergänzung der systematischen Übersichten Primärliteratur identifiziert und anhand folgender Kriterien eingeschlossen:

4.2.2.1 Populationen

Die Primärstudien müssen Angaben zu mindestens 10 Patienten mit erhöhtem Risiko für bzw. Verdacht auf eine Alzheimer-Demenz oder Patienten mit einer diagnostizierten und nicht näher differenzierten Demenz (Einsatz der PET zur Differenzialdiagnostik) machen.

4.2.2.2 Prüftechnologien (Indextest I), Vergleichstechnologien (Indextest II) und Referenzstandard

Prüftechnologie ist die Vollring-PET bzw. PET/CT. Vergleichstechnologie, Referenzstandard und Evaluierung entsprechen den in Abschnitt 4.2.1.2 beschriebenen Vorgaben.

4.2.2.3 Zielgrößen

Für die Zielgrößen der Primärstudien aus der Ergänzungsrecherche gelten dieselben Einschlusskriterien wie für die Zielgrößen der Primärstudien in den systematischen Übersichten (siehe Abschnitt 4.2.1.3).

4.2.2.4 Studientypen

Für die Primärstudien der Ergänzungsrecherche gelten dieselben Einschlusskriterien hinsichtlich der Studientypen wie für die in den systematischen Übersichten enthaltenen Primärstudien (siehe Abschnitt 4.2.1.4).

4.2.2.5 Sonstige Studiencharakteristika

Zur Bewertung der diagnostischen und prognostischen Güte werden nur Publikationen in den folgenden Sprachen eingeschlossen: Englisch, Deutsch, Französisch.

4.2.2.6 Tabellarische Darstellung der Einschlusskriterien („Ergänzungsrecherche“ nach Primärstudien zur diagnostischen und prognostischen Güte)

Es werden alle diagnostischen und prognostischen Studien im Rahmen der „Ergänzungsrecherche“ einbezogen, die alle nachfolgenden Einschlusskriterien erfüllen (siehe Tabelle 8).

Tabelle 8: Übersicht der Einschlusskriterien für die Primärliteratur zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT (Ergänzungsrecherche)

Einschlusskriterien	
E1c	Mindestens 10 Patienten mit erhöhtem Risiko für bzw. Verdacht auf Alzheimer-Demenz oder Patienten mit einer diagnostizierten und nicht näher differenzierten Demenz (Einsatz der PET zur Differenzialdiagnostik; siehe auch Abschnitt 4.1.2.1)
E2c	Prüftechnologie (Vollring-PET bzw. PET/CT), Vergleichstechnologie und Referenzstandard wie in Abschnitt 4.2.1.2 beschrieben
E3c	Zielgrößen wie in Abschnitt 4.2.1.3 formuliert
E4c	Prospektiv geplante Querschnitts- oder Kohortenstudien wie in den Abschnitten 4.2.1.4 und 4.2.1.5 beschrieben
E5c	Keine Mehrfachpublikationen ohne relevante Zusatzinformationen
E6c	Vollpublikation verfügbar ^a
E7c	Publikationszeitraum oder Fragestellung durch Literatursuche der eingeschlossenen Evidenzsynthesen nicht abgedeckt.
a: Als Vollpublikation gilt in diesem Zusammenhang auch die nicht vertrauliche Weitergabe eines Studienberichts an das Institut oder die nicht vertrauliche Bereitstellung eines Berichts über die Studien, der den Kriterien des STARD-Statements [87] oder des STROBE-Statements [88] genügt und eine Bewertung der Studie ermöglicht. STARD: Standards for the Reporting of Diagnostic Accuracy Studies; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology	

4.2.3 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen

Für das Einschlusskriterium E1 (Population) reicht es aus, wenn bei mindestens 80 % der eingeschlossenen Patienten dieses Kriterium erfüllt ist. Liegen für solche Studien entsprechende Subgruppenanalysen vor, wird auf diese Analysen zurückgegriffen. Studien, bei denen das Einschlusskriterium E1 bei weniger als 80 % erfüllt ist, werden nur dann eingeschlossen, wenn entsprechende Subgruppenanalysen vorliegen.

Ebenfalls eingeschlossen werden Studien, die zu mindestens 80 % das Einschlusskriterium E2 für die Prüf- und Vergleichsintervention erfüllen.

4.3 Informationsbeschaffung

4.3.1 Bibliografische Literaturrecherche

Die systematische Literaturrecherche nach relevanten Publikationen soll in folgenden Quellen durchgeführt werden:

- Suche nach Primärstudien in den bibliografischen Datenbanken MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials (Clinical Trials)

- Suche nach systematischen Übersichten: Identifizierung mittels Suche in den Datenbanken MEDLINE und EMBASE parallel zur Suche nach relevanter Primärliteratur sowie mittels Suche in den Datenbanken Cochrane Database of Systematic Reviews (Cochrane Reviews), Database of Abstracts of Reviews of Effects (Other Reviews) und Health Technology Assessment Database (Technology Assessments)

Sollten relevante HTA-Berichte gefunden werden, welche die Einschlusskriterien für systematische Übersichten erfüllen, werden die jeweiligen HTA-Organisationen zu eventuellen Aktualisierungen befragt.

4.3.2 Suche nach weiteren publizierten und nicht publizierten Studien

Zusätzlich zur Suche in bibliografischen Datenbanken sollen folgende Quellen zur Identifizierung publizierter und nicht publizierter Primärstudien und systematischer Übersichten herangezogen werden:

- Systematische Übersichten
- durch den G-BA übermittelte Unterlagen
- öffentlich zugängliche Studienregister
- im Rahmen der Anhörung zum vorläufigen Berichtsplan und zum Vorbericht eingereichte Informationen

4.3.3 Selektion relevanter Studien und systematischer Übersichten

Die Selektion relevanter Primärstudien und systematischer Übersichten erfolgt durch 2 Reviewer unabhängig voneinander. Dazu wird das Ergebnis der Recherche in den oben genannten Quellen herangezogen.

Informationen aus den folgenden Suchquellen werden von 2 Reviewern unabhängig voneinander hinsichtlich ihrer Relevanz bewertet:

- bibliografische Literaturrecherche
- öffentlich zugängliche Studienregister
- vom G-BA übermittelte Unterlagen

Informationen aus den folgenden Suchquellen werden von einem Reviewer auf Studien gesichtet, der diese dann hinsichtlich ihrer Relevanz bewertet; ein zweiter Reviewer überprüft den gesamten Prozess inklusive der Bewertungen:

- im Rahmen der Anhörung zum vorläufigen Berichtsplan und zum Vorbericht eingereichte Informationen

Die identifizierten relevanten systematischen Übersichten werden nach weiteren potenziell relevanten Studien durchsucht, deren Relevanz von 2 Reviewern unabhängig voneinander geprüft wird.

Sofern in einem der genannten Selektionsschritte Diskrepanzen auftreten, werden diese jeweils durch Diskussion zwischen den beiden Reviewern aufgelöst.

Die Fragestellung 1 wird, wie in Abschnitt 4.1 beschrieben, auf Basis der Primärliteratur bearbeitet.

Die Fragestellung 2 zur diagnostischen und prognostischen Güte wird primär auf Basis von systematischen Übersichten beantwortet. Zunächst werden daher gezielt systematische Übersichten gesucht und die gefundenen Arbeiten mit dem Qualitätsindex von Oxman und Guyatt bewertet (vgl. Abschnitt 4.2.1.5). Anschließend wird für den Zeitraum und für Teilfragestellungen, die durch den Recherchezeitraum der systematischen Übersichten nicht abgedeckt werden, eine zusätzliche Recherche nach Primärliteratur durchgeführt (Ergänzungsrecherche; vgl. Abschnitt 4.2.2). Sollten für eine Fragestellung nicht ausreichend belastbare Daten aus systematischen Übersichten vorliegen, wird sie ausschließlich auf Basis der Primärliteratur beantwortet.

4.4 Informationsbewertung

Die Dokumentation und Bewertung der Studiencharakteristika und -ergebnisse der relevanten Publikationen erfolgt anhand von standardisierten Datenextraktionsbögen und Evidenztabellen. Ein Reviewer führt die Datenextraktion unter Verwendung des Extraktionsbogens durch. Ein zweiter Reviewer überprüft die Extraktion. Etwaige Diskrepanzen in der Bewertung werden durch Diskussion zwischen den Reviewern, ggf. unter Hinzuziehung einer 3. Person, aufgelöst.

4.4.1 Bewertung von Studien zum patientenrelevanten Nutzen und Schaden

Die Bewertung der Informationen der eingeschlossenen Studien hängt stark von den verfügbaren Angaben und der Qualität der jeweiligen Publikationen und weiterer Informationsquellen ab. Alle für die Nutzenbewertung relevanten Ergebnisse werden hinsichtlich ihrer Ergebnissicherheit, bestehend aus dem Verzerrungspotenzial und der Präzision der Ergebnisse, überprüft.

Bewertung des Verzerrungspotenzials der Ergebnisse

Das Verzerrungspotenzial der Ergebnisse wird für jede in die Nutzenbewertung eingeschlossene Studie bewertet, und zwar separat für jeden patientenrelevanten Endpunkt. Dazu werden insbesondere folgende endpunktübergreifende (A) und endpunktspezifische (B) Aspekte, die das Verzerrungspotenzial beeinflussen, systematisch extrahiert und bewertet:

A: Aspekte des Verzerrungspotenzials der Ergebnisse auf Studienebene

- Erzeugung der Randomisierungssequenz (bei randomisierten Studien)

- Verdeckung der Gruppenzuteilung (bei randomisierten Studien)
- zeitliche Parallelität der Gruppen (bei nicht randomisierten kontrollierten Studien)
- Vergleichbarkeit der Gruppen bzw. Berücksichtigung prognostisch relevanter Faktoren (bei nicht randomisierten kontrollierten Studien)
- Verblindung des Patienten sowie des Behandlers (bei randomisierten Studien)
- ergebnisgesteuerte Berichterstattung

B: Aspekte des Verzerrungspotenzials der Ergebnisse auf Endpunktebene

- Verblindung der Endpunkterheber
- Umsetzung des ITT-Prinzips
- ergebnisgesteuerte Berichterstattung

Das Verzerrungspotenzial wird als „niedrig“ oder „hoch“ eingestuft. Ein niedriges Verzerrungspotenzial liegt dann vor, wenn mit großer Wahrscheinlichkeit ausgeschlossen werden kann, dass die Ergebnisse relevant verzerrt sind. Unter einer relevanten Verzerrung ist zu verstehen, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

Für die Bewertung eines Endpunkts wird zunächst das Verzerrungspotenzial endpunktübergreifend anhand der unter A aufgeführten Aspekte als „niedrig“ oder „hoch“ eingestuft. Falls diese Einstufung als „hoch“ erfolgt, wird das Verzerrungspotenzial für den Endpunkt in der Regel auch als „hoch“ bewertet. Ansonsten finden die unter B genannten endpunkt-spezifischen Aspekte Berücksichtigung.

Eine Einstufung des Verzerrungspotenzials des Ergebnisses für einen Endpunkt als „hoch“ führt nicht zum Ausschluss aus der Nutzenbewertung. Die Klassifizierung dient vielmehr der Diskussion heterogener Studienergebnisse und beeinflusst die Sicherheit der Aussage.

4.4.2 Bewertung von systematischen Übersichten zur diagnostischen und prognostischen Güte

Es wird eine Übersicht der in den systematischen Übersichten jeweils eingeschlossenen Studien mit den wesentlichen Studiencharakteristika in tabellarischer Form erstellt. Die Ergebnisse der zugrunde liegenden Primärstudien werden dabei nach den Technologien Indextest I, ggf. Indextest II und Referenzstandard und den Studientypen aufgeteilt.

Die Bewertung des Verzerrungspotenzials einer systematischen Übersicht erfolgt anhand der Kriterien von Oxman und Guyatt [84,89]. Diese Kriterien werden um die Angabe von Interessenkonflikten ergänzt. Die Bewertung erfolgt regelhaft für die gesamte systematische Übersicht und ist daher für mehrere Fragestellungen bzw. Indikationen einer Übersicht immer gleich.

4.4.3 Bewertung von Primärstudien zur diagnostischen und prognostischen Güte

Die Studien werden zunächst nach Indikationen, Technologievergleichen und Studientypen aufgeteilt und separat dargestellt. Die Angaben zu Studienpopulationen, verglichenen Technologien, methodischen Aspekten und Ergebnissen einzelner Studien werden tabellarisch dargestellt.

Die Bewertung des Verzerrungspotenzials der Primärstudien zur diagnostischen Güte wird mithilfe des QUADAS-2-Instruments durchgeführt [90]. Für die Bewertung des Verzerrungspotenzials prognostischer Studien wurde auf Grundlage der Arbeiten von Hayden et al. [91] und Altman et al. [92] ein neues Instrument entwickelt [93].

Das Verzerrungspotenzial der Primärstudien zur diagnostischen und prognostischen Güte wird als „niedrig“ oder „hoch“ eingestuft.

Eine Einstufung des Verzerrungspotenzials des Ergebnisses für einen Endpunkt als „hoch“ führt nicht zum Ausschluss aus der Bewertung der diagnostischen und prognostischen Güte. Die Klassifizierung dient vielmehr der Diskussion heterogener Studienergebnisse und beeinflusst die Sicherheit der Aussage.

4.5 Informationssynthese und -analyse

Die Informationen werden einer Informationssynthese und -analyse unterzogen. Wenn möglich werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die unten beschriebenen Werkzeuge (Meta-Analyse, Sensitivitäts- und Subgruppenanalyse) eingesetzt. Eine abschließende zusammenfassende Bewertung der Informationen erfolgt darüber hinaus in jedem Fall. Die Ergebnisse der systematischen Übersichten werden entsprechend den Darstellungen in den Publikationen berichtet. Auch hier erfolgt eine Darstellung der Ergebnisse jeder darin enthaltenen Primärstudie. Zusammenfassende Ergebnisse der systematischen Übersichten werden ebenfalls berichtet.

Bei der Betrachtung patientenrelevanter Endpunkte, die mithilfe von (komplexen) Skalen operationalisiert werden, ist es in besonderer Weise notwendig, neben der statistischen Signifikanz der Effekte die Relevanz der beobachteten Wirkungen der untersuchten Interventionen zu bewerten, da die Komplexität der Skalen häufig eine sinnvolle Interpretation geringfügiger Unterschiede erschwert. Die Relevanzbewertung kann auf Basis von Mittelwertdifferenzen und Responderanalysen vorgenommen werden. Welches dieser Verfahren gewählt werden kann, hängt insbesondere von der Verfügbarkeit der Daten aus den Primärstudien ab.

Um skalen- und indikationsspezifischen Besonderheiten gerecht zu werden, soll die Relevanzbewertung primär auf Basis validierter (bzw. etablierter) Relevanzkriterien für die jeweilige Skala durchgeführt werden. Für die Betrachtung von Mittelwertdifferenzen ist dies z. B. eine validierte bzw. etablierte Irrelevanzschwelle für den Gruppenunterschied: Liegt das zum beobachteten Effekt korrespondierende Konfidenzintervall vollständig oberhalb dieser

Irrelevanzschwelle, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Bei der Betrachtung von Responderanalysen ist es erforderlich, dass bei diesen Analysen ein validiertes bzw. etabliertes Responsekriterium angewendet wurde (im Sinne einer individuellen Minimal Important Difference [MID]). Liegt bei einer solchen Auswertung ein statistisch signifikanter Unterschied der Anteile an Respondern zwischen den Gruppen vor, wird dies als relevanter Effekt angesehen, da die Response-Definition bereits eine Schwelle der Irrelevanz (nämlich die MID) beinhaltet.

Falls skalenspezifisch validierte bzw. etablierte Relevanzkriterien nicht vorliegen, muss auf ein allgemeines statistisches Maß zur Relevanzbewertung zurückgegriffen werden. In diesem Fall werden standardisierte Mittelwertdifferenzen (SMD in Form von Hedges' g) betrachtet. Als Irrelevanzschwelle wird dann 0,2 verwendet: Liegt das zum beobachteten Effekt korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Dies soll gewährleisten, dass der beobachtete Effekt hinreichend sicher mindestens als „klein“ angesehen werden kann [94,95].

4.5.1 Gegenüberstellung der Ergebnisse der Einzelstudien

Die Ergebnisse zu den in den Studien berichteten Endpunkten werden im Bericht vergleichend beschrieben. Weiterhin erfolgt eine Abschätzung der Heterogenität der Ergebnisse innerhalb der jeweiligen Fragestellung.

In bestimmten Fällen werden einzelne Ergebnisse aus den Studien zu einem Endpunkt nicht dargestellt bzw. nicht in die Nutzenbewertung oder Bewertung der diagnostischen und prognostischen Güte einbezogen. Dies trifft insbesondere zu, wenn viele Patienten nicht in der Auswertung enthalten sind. Ergebnisse fließen in der Regel nicht in die Bewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patienten basieren, das heißt, wenn der Anteil der Patienten ohne jegliche Berücksichtigung in der Auswertung (Nichtberücksichtigungsanteil) größer als 30 % ist. In der Literatur werden zum Teil bereits Nichtberücksichtigungsanteile größer als 20 Prozentpunkte als nicht mehr aussagekräftig betrachtet [96].

Ausnahmen von dieser Regel können zum Beispiel dann gemacht werden, wenn aus logistischen Gründen für ganze Zentren (ganze Randomisierungsblöcke) keine Daten erhoben wurden und dies bereits bei der Studienplanung vorgesehen war [97].

Die Ergebnisse werden auch dann nicht in die Nutzenbewertung oder Bewertung der diagnostischen und prognostischen Güte einbezogen, wenn der Unterschied der Nichtberücksichtigungsanteile zwischen den Gruppen größer als 15 Prozentpunkte ist.

4.5.1.1 Gegenüberstellung der Ergebnisse von (Primär-)Studien zur Bewertung des patientenrelevanten Nutzens und Schadens

Im Bericht werden die Studien anhand des Designs (Studiendesign, Studiendauer, Zahl der randomisierten / zugeteilten Patienten, Zuteilungsmechanismus, Ort und Zeitraum der Studiendurchführung, relevante Zielkriterien) beschrieben. Darüber hinaus werden die in der jeweiligen Studie verglichenen diagnostischen Strategien (z. B. eingesetzte Technologien, eventuelle Effekte der Vorselektion der Patienten etc.) dargestellt. Die Studienpopulationen werden weiterhin durch demografische Daten (Alter, Geschlecht), durch Charakteristika der betrachteten Krankheitsentität und durch prognostische Angaben beschrieben. Zusätzlich wird die Zahl der Studienabbrecher dargestellt.

Die Ergebnisse der jeweiligen Behandlungsgruppen zu den in den Studien berichteten patientenrelevanten Endpunkten werden im Bericht vergleichend beschrieben. Dazu werden zunächst die in den Studien berichteten jeweiligen Endpunkte den im Berichtsplan definierten patientenrelevanten Endpunkten (siehe Abschnitt 4.1.2.3) zugeordnet. Endpunkte aus Studien, die keiner vordefinierten Zielgröße gemäß Abschnitt 4.1.2.3 zugeordnet werden können, sind als nicht patientenrelevant anzusehen und werden daher in der Nutzenbewertung nicht berücksichtigt.

4.5.1.2 Gegenüberstellung der Ergebnisse von systematischen Übersichten zur diagnostischen und prognostischen Güte

Bei der Beurteilung der diagnostischen und prognostischen Güte werden zunächst die zu der Fragestellung vorhandenen Ergebnisse der systematischen Übersichten abgebildet. Die Ergebnisse aus den zugrunde liegenden Primärstudien werden dabei – soweit notwendig und möglich – zunächst nach den Fragestellungen und Ordnungskriterien „Indextest I“ (z. B. PET), „Indextest II“ (z. B. PET/CT) und „Referenzstandard“ klassifiziert und qualitativ zusammengefasst. Zudem erfolgt eine Abschätzung der Heterogenität der Ergebnisse innerhalb dieser Klassen. Beim Vorliegen inkonsistenter Ergebnisse zu einer Fragestellung werden mögliche Erklärungen für die beobachtete Heterogenität (z. B. Unterschiede bei der Berücksichtigung eines möglichen Verifikations- oder Inkorporationsbias) beschrieben [98]. Sollten die eingeschlossenen systematischen Übersichten Ergebnisse zu Managementänderungen berichten, werden diese ebenfalls extrahiert und tabellarisch zusammengefasst.

Sollten durch eine Ergänzungsrecherche für den Zeitraum nach Abschluss des Recherchezeitraums der eingeschlossenen systematischen Übersichten noch weitere relevante Primärstudien identifiziert werden, so werden diese separat und analog zu den (Primär-)Studien zur Bewertung des patientenrelevanten Nutzens und Schadens präsentiert (siehe Abschnitt 4.5.1.3).

4.5.1.3 Gegenüberstellung der Ergebnisse von Primärstudien zur diagnostischen und prognostischen Güte („Ergänzungsrecherche“)

Im Bericht werden die Studien anhand der zugrunde liegenden Methodik (Studiendesign, Studiendauer, Ort und Zeitraum der Studiendurchführung, relevante Endpunkte) beschrieben. Die Studienpopulationen werden weiterhin durch demografische Daten (Alter, Geschlecht), durch Charakteristika der betrachteten Krankheitsentität und durch prognostische Angaben beschrieben. Zusätzlich sollte die Zahl der Studienabbrecher dargestellt werden.

Bei der Beurteilung der diagnostischen und prognostischen Güte werden die zu der Fragestellung vorhandenen Ergebnisse der Primärstudien abgebildet. Die Ergebnisse werden dabei – soweit notwendig und möglich – zunächst nach den Fragestellungen und Ordnungskriterien „Indextest I“, „Indextest II“ und „Referenzstandard“ klassifiziert und qualitativ zusammengefasst. Bei prognostischen Primärstudien werden – soweit untersucht – multi-faktorielle Analysen und der Einfluss von möglichen Confoundern auf die prognostische Güte präsentiert. Sollten die eingeschlossenen Primärstudien Ergebnisse zu Managementänderungen berichten, werden diese ebenfalls extrahiert und tabellarisch zusammengefasst.

4.5.1.4 Gegenüberstellung der Ergebnisse aus systematischen Übersichten und der Ergänzungsrecherche („Robustheitsprüfung“)

Finden sich sowohl relevante systematische Übersichten als auch relevante Primärstudien in der Ergänzungsrecherche, wird geprüft, ob die (quantitativ zusammengefassten) Ergebnisse der aktuelleren Primärstudien die Ergebnisse der systematischen Übersichten infrage stellen könnten („Robustheitsprüfung“). Nur wenn dies zu erwarten ist, werden zusammenfassende Analysen über alle Primärstudien – sowohl jene, die in den systematischen Übersichten vorhanden waren, als auch jene aus der Ergänzungsrecherche – in Erwägung gezogen.

4.5.2 Meta-Analysen

4.5.2.1 Meta-Analysen für vergleichende Interventionsstudien

Sofern die Primärstudien hinsichtlich der Fragestellung und relevanter Charakteristika vergleichbar sind, sollen die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammengefasst werden. Für die statistische Auswertung vergleichender Studien sollen primär die Ergebnisse aus Intention-to-treat-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet werden.

Die Meta-Analysen sollen auf Basis von Modellen mit zufälligen Effekten erfolgen (engl.: random-effects model) [99]. In begründeten Ausnahmefällen sollen Modelle mit festen Effekten (engl.: fixed-effects model) eingesetzt werden. Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, sollen diese nach Möglichkeit anhand der vorhandenen Informationen eigenständig berechnet bzw. näherungsweise bestimmt werden. Bei Studien zum patientenrelevanten Nutzen und Schaden

wird für kontinuierliche Variablen die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges' g , als Effektmaß eingesetzt. Bei binären Variablen sollen Meta-Analysen primär anhand des Odds Ratio und beim Vergleich von Überlebenszeiten anhand des Hazard Ratio durchgeführt werden. In begründeten Ausnahmefällen können auch andere Effektmaße zum Einsatz kommen. Die Effektschätzer und Konfidenzintervalle aus den Studien sollen mittels Forest Plots zusammenfassend dargestellt werden. Anschließend soll die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand des Maßes I^2 und des statistischen Tests auf Vorliegen von Heterogenität erfolgen [100]. In den Fällen, in denen die Heterogenität der Studienergebnisse nicht bedeutsam ist ($p > 0,2$ für Heterogenitätstest), soll der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt werden. Bei bedeutsamer Heterogenität sollen die Ergebnisse nur in begründeten Ausnahmefällen gepoolt werden. Außerdem soll untersucht werden, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu sollten methodische Faktoren (siehe Abschnitt 4.5.3) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.5.4), berücksichtigt werden.

4.5.2.2 Meta-Analysen für Prognosestudien

Die Meta-Analysen von Prognosestudien sollen entsprechend dem Vorgehen bei der Bewertung des patientenrelevanten Nutzens erstellt werden.

4.5.2.3 Meta-Analysen für Diagnosestudien

Die Punktschätzer und die dazugehörigen univariaten 95 %-Konfidenzintervalle [101] aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Außerdem wird für die Testgütekriterien Sensitivität und Spezifität eine bivariate Meta-Analyse durchgeführt [102]. Die Schätzung der Modellparameter erfolgt über ein generalisiertes, lineares, gemischtes Modell [103,104]. Der Algorithmus zum Schätzen der Parameter im bivariaten Modell kann zu unpräzisen Schätzern führen, das heißt zu Schätzern mit zu großen Standardfehlern und entsprechenden Konfidenzregionen. Auch kann der Algorithmus ggf. keine Schätzer liefern, wenn das Maximum-Likelihood-Verfahren nicht konvergiert. Die Gründe für das Fehlen von brauchbaren Schätzern können beispielsweise sein, dass zu wenige Studien vorliegen oder dass einzelne Studien extreme Werte aufweisen. In diesen Fällen werden die Ergebnisse der bivariaten Meta-Analysen in der Regel nicht dargestellt. Wenn zu wenige vergleichende Studien für eine bestimmte Technologie vorliegen, können Technologien, zum Beispiel die PET und PET/CT, in den bivariaten Meta-Analysen zusammengefasst werden, falls die Analyse eine sinnvolle Interpretation zulässt.

Zusätzlich werden bei diagnostischen Studien die beobachteten Paare aus Sensitivität und Spezifität zweidimensional grafisch dargestellt. Studien, in denen Technologien direkt verglichen wurden, werden gekennzeichnet. Des Weiteren werden die aus der bivariaten Meta-Analyse gewonnenen Schätzer für die Erwartungswerte als gepoolte Paare der Sensitivität und der Spezifität mit den dazugehörigen 95 %-Konfidenzregionen dargestellt [105].

Falls die bivariate Meta-Analyse präzise Schätzer liefert, wird auf einen Unterschied in der diagnostischen Güte zwischen 2 Technologien mit dem Likelihood-Ratio-Test (LRT) getestet. Ist der Test statistisch signifikant, so wird die Differenzierung zwischen einer „höheren diagnostischen Güte“ und einer „vermutlich höheren diagnostischen Güte“ anhand der grafischen Darstellungen der direkten Vergleiche vorgenommen.

Liefert die bivariate Meta-Analyse unpräzise Schätzer, die zu sehr großen Konfidenz- und Prädiktionsregionen führen, werden keine Aussagen über einen Unterschied in der diagnostischen Güte zwischen 2 Technologien abgeleitet. In begründeten Ausnahmefällen, wie beispielsweise beim Vorliegen von mehreren großen Studien mit niedrigem Verzerrungspotenzial, werden gegebenenfalls die Ergebnisse geeigneter univariater statistischer Tests, das heißt für die Sensitivität und Spezifität getrennt, dargestellt. Unter Berücksichtigung der grafischen Darstellungen sollten gegebenenfalls Aussagen zu einem Unterschied in der diagnostischen Güte zwischen 2 Technologien abgeleitet werden.

Das Vorliegen von Heterogenität wird anhand von Sensitivitätsanalysen und Subgruppenanalysen untersucht.

4.5.3 Sensitivitätsanalysen

Zur Einschätzung der Robustheit der Ergebnisse sind Sensitivitätsanalysen hinsichtlich methodischer Faktoren geplant. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ und „niedrig“ soll hierfür verwendet werden. Dies soll sowohl Primärstudien zum patientenrelevanten Nutzen und Schaden als auch jene zur diagnostischen und prognostischen Güte betreffen. Beim Einschluss von mehreren Evidenzsynthesen sollen, wenn möglich, Sensitivitätsanalysen anhand der Beurteilung mit dem Index von Oxman und Guyatt [84,89] durchgeführt werden (Stufe 5 versus 6 versus 7; alle < 5 werden ausgeschlossen). Weitere methodische Faktoren können aus den im Rahmen der Informationsbewertung aufgefallenen Merkmalen abgeleitet werden, zum Beispiel die Festlegung von Cut-off-Werten und Erhebungszeitpunkten oder die Wahl des Effektmaßes.

Bei vergleichenden Interventionsstudien (Nutzenfragestellung) werden ab 4 Studien Prädiktionsintervalle berechnet, um einen Eindruck von der Heterogenität zu gewinnen. In bivariaten Meta-Analysen (diagnostische Güte) werden 95 %-Prädiktionsregionen für den Erwartungswert einer neuen Studie bestimmt, um Ausreißer zu identifizieren.

Sensitivitätsanalysen werden nach dem Leave-one-out-Prinzip durchgeführt, um den Einfluss jeder einzelnen Studie auf das Ergebnis zu bewerten.

Das Ergebnis der Sensitivitätsanalysen kann die Sicherheit der aus den beobachteten Effekten abgeleiteten Aussagen beeinflussen. Ein als nicht robust eingestufter Effekt kann zum Beispiel dazu führen, dass aus vergleichenden Interventionsstudien nur ein Hinweis auf anstelle eines Belegs für einen Nutzen oder Schaden abgeleitet wird. Bei Diagnose- und Prognosestudien würden nicht robuste Ergebnisse dazu führen, dass keine Aussage über den

Vergleich der diagnostischen oder prognostischen Güte zweier Indextests abgeleitet werden kann.

4.5.4 Subgruppenmerkmale und andere Effektmodifikatoren

Die Ergebnisse sollen hinsichtlich potenzieller Effektmodifikatoren, das heißt klinischer Faktoren, die die Effekte beeinflussen, untersucht werden. Dies können direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Behandlungen sein. Im Gegensatz zu den in Abschnitt 4.5.3 beschriebenen methodischen Faktoren für Sensitivitätsanalysen ist hier das Ziel, mögliche Effektunterschiede zwischen Patientengruppen und Behandlungsspezifika aufzudecken. Für den Nachweis unterschiedlicher Effekte sollte die auf einem Homogenitäts- beziehungsweise Interaktionstest basierende statistische Signifikanz Voraussetzung sein. In die Untersuchung von Effektmodifikatoren sollten die vorliegenden Ergebnisse aus Regressionsanalysen, die Interaktionsterme beinhalten, und aus Subgruppenanalysen einbezogen werden. Außerdem sollen eigene Analysen in Form von Meta-Regressionen oder Meta-Analysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren erfolgen. Es ist vorgesehen, folgende Faktoren bezüglich einer möglichen Effektmodifikation in die Analysen einzubeziehen:

- Geschlecht
- Alter
- Krankheitsstadium
- Behandlungsansatz (kurativ vs. palliativ)
- Setting (z. B. universitäres Zentrum vs. ambulante Praxis)
- verwendeter Referenzstandard

Ergeben sich aus den verfügbaren Informationen weitere relevante Subgruppen oder Effektmodifikatoren, können diese ebenfalls begründet einbezogen werden.

Bei Identifizierung möglicher Effektmodifikatoren erfolgt gegebenenfalls eine Präzisierung der aus den beobachteten Effekten abgeleiteten Aussagen. Beispielsweise kann der Beleg eines Zusatznutzens auf eine spezielle Subgruppe von Patienten eingeschränkt werden.

5 Literaturverzeichnis

1. Dilling H, Mombour W, Schmidt MH, Schulte-Markwort E. Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F); klinisch-diagnostische Leitlinien. Bern: Huber; 2005.
2. World Health Organization. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme: 10. Revision; German modification; Version 2011 [online]. 24.09.2010 [Zugriff: 10.08.2011]. URL: <http://www.dimdi.de/static/de/klassi/diagnosen/icd10/htmlgm2011/index.htm>.
3. Qiu C, De Ronchi D, Fratiglioni L. The epidemiology of the dementias: an update. *Curr Opin Psychiatry* 2007; 20(4): 380-385.
4. Weyerer S. Altersdemenz. Berlin: Robert Koch Institut; 2005. (Gesundheitsberichterstattung des Bundes; Band 28). URL: http://www.gbe-bund.de/gbe10/owards.prc_show_pdf?p_id=9663&p_sprache=d.
5. Bickel H. Demenzsyndrom und Alzheimer Krankheit: eine Schätzung des Krankenbestandes und der jährlichen Neuerkrankungen in Deutschland. *Gesundheitswesen* 2000; 62(4): 211-218.
6. Fratiglioni L, De Ronchi D, Agüero-Torres H. Worldwide prevalence and incidence of dementia. *Drugs Aging* 1999; 15(5): 365-375.
7. Prince M, Jackson J. World Alzheimer Report 2009. Alzheimer's Disease International (ADI).
8. National Collaborating Centre for Mental Health. Dementia: a NICE–SCIE guideline on supporting people with dementia and their carers in health and social care (commissioned by the Social Care Institute for Excellence National Institute for Health and Clinical Excellence). Leicester: British Psychological Society; 2007. (National Clinical Practice Guidelines; Band 42). URL: <http://www.nice.org.uk/nicemedia/pdf/CG42Dementiafinal.pdf>.
9. Andersen K, Launer LJ, Dewey ME, Letenneur L, Ott A, Copeland JR et al. Gender differences in the incidence of AD and vascular dementia: the EURODEM studies. *Neurology* 1999; 53(9): 1992-1997.
10. Jayadev S, Steinbart EJ, Chi YY, Kukull WA, Schellenberg GD, Bird TD. Conjugal Alzheimer disease: risk in children when both parents have Alzheimer disease. *Arch Neurol* 2008; 65(3): 373-378.
11. Van Duijn CM, Clayton D, Chandra V, Fratiglioni L, Graves AB, Heyman A et al. Familial aggregation of Alzheimer's disease and related disorders: a collaborative re-analysis of case-control studies. *Int J Epidemiol* 1991; 20(Suppl 2): S13-S20.
12. Morris JC. Dementia update 2005. *Alzheimer Dis Assoc Disord* 2005; 19(2): 100-117.

13. Alonso A, Jacobs DR Jr, Menotti A, Nissinen A, Dontas A, Kafatos A et al. Cardiovascular risk factors and dementia mortality: 40 years of follow-up in the Seven Countries Study. *J Neurol Sci* 2009; 280(1-2): 79-83.
14. Kivipelto M, Solomon A. Alzheimer's disease: the ways of prevention. *J Nutr Health Aging* 2008; 12(1): 89S-94S.
15. Lethaby A, Hogervorst E, Richards M, Yesufu A, Yaffe K. Hormone replacement therapy for cognitive function in postmenopausal women. *Cochrane Database Syst Rev* 2008; (1): CD003122.
16. Low LF, Anstey KJ. Hormone replacement therapy and cognitive performance in postmenopausal women: a review by cognitive domain. *Neurosci Biobehav Rev* 2006; 30(1): 66-84.
17. Shumaker SA, Legault C, Kuller L, Rapp SR, Thal L, Lane DS et al. Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *JAMA* 2004; 291(24): 2947-2958.
18. Feigin V, Ratnasabapathy Y, Anderson C. Does blood pressure lowering treatment prevents dementia or cognitive decline in patients with cardiovascular and cerebrovascular disease? *J Neurol Sci* 2005; 229-230: 151-155.
19. Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin. Leitlinie Demenz. Düsseldorf: Omikron Publishing; 2008. (DEGAM-Leitlinien; Band 12). URL: http://leitlinien.degam.de/uploads/media/LL-12_Langfassung_TJ_03_korr_01.pdf.
20. Verghese J, Lipton RB, Katz MJ, Hall CB, Derby CA, Kuslansky G et al. Leisure activities and the risk of dementia in the elderly. *N Engl J Med* 2003; 348(25): 2508-2516.
21. Boothby LA, Doering PL. Vitamin C and vitamin E for Alzheimer's disease. *Ann Pharmacother* 2005; 39(12): 2073-2080.
22. Tyas SL, Salazar JC, Snowdon DA, Desrosiers MF, Riley KP, Mendiondo MS et al. Transitions to mild cognitive impairments, dementia, and death: findings from the Nun Study. *Am J Epidemiol* 2007; 165(11): 1231-1238.
23. Butterfield DA, Castegna A, Lauderback CM, Drake J. Evidence that amyloid beta-peptide-induced lipid peroxidation and its sequelae in Alzheimer's disease brain contribute to neuronal death. *Neurobiol Aging* 2002; 23(5): 655-664.
24. Carter DB, Dunn E, McKinley DD, Stratman NC, Boyle TP, Kuiper SL et al. Human apolipoprotein E4 accelerates beta-amyloid deposition in APPsw transgenic mouse brain. *Ann Neurol* 2001; 50(4): 468-475.
25. Hsiao K, Chapman P, Nilsen S, Eckman C, Harigaya Y, Younkin S et al. Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice. *Science* 1996; 274(5284): 99-102.

26. Braak E, Braak H, Mandelkow EM. A sequence of cytoskeleton changes related to the formation of neurofibrillary tangles and neuropil threads. *Acta Neuropathol* 1994; 87(6): 554-567.
27. Cacabelos R, Takeda M, Winblad B. The glutamatergic system and neurodegeneration in dementia: preventive strategies in Alzheimer's disease. *Int J Geriatr Psychiatry* 1999; 14(1): 3-47.
28. 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 and the Alzheimer's Association Workgroup. *Alzheimers Dement* 2011; 7(3): 280-292.
29. Herrup K. Reimagining Alzheimer's disease: an age-based hypothesis. *J Neurosci* 2010; 30(50): 16755-16762.
30. Pimplikar SW, Nixon RA, Robakis NK, Shen J, Tsai LH. Amyloid-independent mechanisms in Alzheimer's disease pathogenesis. *J Neurosci* 2010; 30(45): 14946-14954.
31. Rymer S, Salloway S, Norton L, Malloy P, Correia S, Monast D. Impaired awareness, behavior disturbance, and caregiver burden in Alzheimer disease. *Alzheimer Dis Assoc Disord* 2002; 16(4): 248-253.
32. Wolfson C, Wolfson DB, Asgharian M, M'Lan CE, Ostbye T, Rockwood K et al. A reevaluation of the duration of survival after the onset of dementia. *N Engl J Med* 2001; 344(15): 1111-1116.
33. Jack CR Jr, 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(3): 257-262.
34. The three stages of Alzheimer's disease. *Lancet* 2011; 377(9776): 1465.
35. Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technol Assess* 2005; 9(10): iii-108.
36. Wolfson C, Moride Y, Perrault A, Momoli F, Demers L, Oremus M. Drug treatments for Alzheimer's disease; II: a review of outcome measures in clinical trials. Ottawa: Canadian Coordinating Office for Health Technology Assessment; 2000. (CCOHTA Technology Reports; Band 9). URL: http://cadth.ca/media/pdf/118_alzheimer2_tr_e.pdf.
37. Sands LP, Ferreira P, Stewart AL, Brod M, Yaffe K. What explains differences between dementia patients' and their caregivers' ratings of patients' quality of life? *Am J Geriatr Psychiatry* 2004; 12(3): 272-280.

38. Desmond DW, Erkinjuntti T, Sano M, Cummings JL, Bowler JV, Pasquier F et al. The cognitive syndrome of vascular dementia: implications for clinical trials. *Alzheimer Dis Assoc Disord* 1999; 13(Suppl 3): S21-S29.
39. Dilling H, Mombour W, Schmidt MH, Schulte-Markwort E. Internationale Klassifikation psychischer Störungen: ICD-10, Kapitel V (F); diagnostische Kriterien für Forschung und Praxis. Bern: Huber; 2004.
40. Saß H, Wittchen HU, Zaudig M, Houben I. Diagnostisches und statistisches Manual psychischer Störungen: Textrevision; DSM-IV-TR. Göttingen: Hogrefe; 2003.
41. 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.
42. 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(9): 1143-1153.
43. Varma AR, Snowden JS, Lloyd JJ, Talbot PR, Mann DM, Neary D. Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 1999; 66(2): 184-188.
44. 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(3): 270-279.
45. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, 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(3): 263-269.
46. Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement* 2012; 8(1): 1-13.
47. Lopez OL, McDade E, Riverol M, Becker JT. Evolution of the diagnostic criteria for degenerative and cognitive disorders. *Curr Opin Neurol* 2011; 24(6): 532-541.
48. 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(8): 734-746.
49. Jellinger KA. Con: can neuropathology really confirm the exact diagnosis? *Alzheimers Res Ther* 2010; 2(2): 11.

50. Scheltens P, Rockwood K. How golden is the gold standard of neuropathology in dementia? *Alzheimers Dement* 2011; 7(4): 486-489.
51. Tyas SL, Snowdon DA, Desrosiers MF, Riley KP, Markesbery WR. Healthy ageing in the Nun Study: definition and neuropathologic correlates. *Age Ageing* 2007; 36(6): 650-655.
52. Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C. Age, neuropathology, and dementia. *N Engl J Med* 2009; 360(22): 2302-2309.
53. Meyer PT, Hellwig S, Amtage F, Rottenburger C, Sahm U, Reuland P et al. Dual-biomarker imaging of regional cerebral amyloid load and neuronal activity in dementia with PET and ¹¹C-labeled Pittsburgh compound B. *J Nucl Med* 2011; 52(3): 393-400.
54. Vemuri P, Wiste HJ, Weigand SD, Shaw LM, Trojanowski JQ, Weiner MW et al. MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurology* 2009; 73(4): 294-301.
55. Yaffe K, Weston A, Graff-Radford NR, Satterfield S, Simonsick EM, Younkin SG et al. Association of plasma beta-amyloid level and cognitive reserve with subsequent cognitive decline. *JAMA* 2011; 305(3): 261-266.
56. Birks J, Flicker L. Donepezil for mild cognitive impairment. *Cochrane Database Syst Rev* 2006; (3): CD006104.
57. Clarfield AM. The decreasing prevalence of reversible dementias: an updated meta-analysis. *Arch Intern Med* 2003; 163(18): 2219-2229.
58. Kotzerke J, Oehme L, Lindner O, Hellwig D. Positron emission tomography 2008 in Germany: results of the query and current status. *Nuklearmedizin* 2010; 49(2): 58-64.
59. Deutsche Gesellschaft für Psychiatrie Psychotherapie und Nervenheilkunde, Deutsche Gesellschaft für Neurologie. S3-Leitlinie "Demenzen" - AWMF Register Nr. 038/013. 2009. URL: http://www.awmf.org/uploads/tx_szleitlinien/038-013_S3_Demenzen_lang_11-2009_11-2011.pdf.
60. Kantarci K, Senjem ML, Lowe VJ, Wiste HJ, Weigand SD, Kemp BJ et al. Effects of age on the glucose metabolic changes in mild cognitive impairment. *AJNR Am J Neuroradiol* 2010; 31(7): 1247-1253.
61. Mosconi L, Berti V, Glodzik L, Pupi A, De Santi S, De Leon MJ. Pre-clinical detection of Alzheimer's disease using FDG-PET, with or without amyloid imaging. *J Alzheimers Dis* 2010; 20(3): 843-854.
62. Quigley H, Colloby SJ, O'Brien JT. PET imaging of brain amyloid in dementia: a review. *Int J Geriatr Psychiatry* 2011; 26(10): 991-999.
63. Hort J, O'Brien JT, Gainotti G, Pirttila T, Popescu BO, Rektorova I et al. EFNS guidelines for the diagnosis and management of Alzheimer's disease. *Eur J Neurol* 2010; 17(10): 1236-1248.

64. Silverman DH, Alavi A. PET imaging in the assessment of normal and impaired cognitive function. *Radiol Clin North Am* 2005; 43(1): 67-77.
65. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 4.0 [online]. 23.09.2011 [Zugriff: 07.12.2011]. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4_0.pdf.
66. How to read clinical journals; II: to learn about a diagnostic test. *Can Med Assoc J* 1981; 124(6): 703-710.
67. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2): 88-94.
68. Köbberling J, Trampisch HJ, Windeler J. Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem* 1990; 28(12): 873-879.
69. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. 21.02.2013 [Zugriff: 27.05.2013]. URL: http://www.g-ba.de/downloads/62-492-706/VerfO_2013-02-21.pdf.
70. Lijmer JG, Bossuyt PMM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009; 62(4): 364-373.
71. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869.
72. Whiting P, Rutjes AW, Dinnes J, Reitsma JB, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004; 8(25): 1-234.
73. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3: 25.
74. Reitsma JB, Rutjes AW, 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(8): 797-806.
75. Jellinger KA. Criteria for the neuropathological diagnosis of dementing disorders: routes out of the swamp? *Acta Neuropathol* 2009; 117(2): 101-110.
76. European Medicines Agency. Guideline on medicinal products for the treatment of Alzheimer's disease and other dementias [online]. 24.07.2008 [Zugriff: 07.12.2011]. URL: <http://www.emea.europa.eu/pdfs/human/ewp/055395en.pdf>.
77. Vellas B, Andrieu S, Sampaio C, Coley N, Wilcock G. Endpoints for trials in Alzheimer's disease: a European task force consensus. *Lancet Neurol* 2008; 7(5): 436-450.
78. Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH et al. Mild cognitive impairment represents early-stage Alzheimer disease. *Arch Neurol* 2001; 58(3): 397-405.

79. Hodges JR, Erzinclioglu S, Patterson K. Evolution of cognitive deficits and conversion to dementia in patients with mild cognitive impairment: a very-long-term follow-up study. *Dement Geriatr Cogn Disord* 2006; 21(5-6): 380-391.
80. Meyer J, Xu G, Thornby J, Chowdhury M, Quach M. Longitudinal analysis of abnormal domains comprising mild cognitive impairment (MCI) during aging. *J Neurol Sci* 2002; 201(1-2): 19-25.
81. Hampel H, Prvulovic D, Teipel S, Jessen F, Luckhaus C, Frölich L et al. The future of Alzheimer's disease: the next 10 years. *Prog Neurobiol* 2011; 95(4): 718-728.
82. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2001; 2(3): 249-260.
83. Jadad AR, Murray WE. *Randomized controlled trials: questions, answers and musings*. Malden: Blackwell; 2007.
84. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44(11): 1271-1278.
85. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151(4): 264-269, W64.
86. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000; 283(15): 2008-2012.
87. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138(1): 1-12.
88. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.
89. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; 44(1): 91-98.
90. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; 155(8): 529-536.
91. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; 144(6): 427-437.
92. Altman DG. Systematic reviews of evaluations of prognostic variables. *Br Med J* 2001; 323(7306): 224-228.

93. Wolff R, Westwood M, Scheibler F, Schröer-Günther M, Janßen I, Kleijnen J. Assessment of risk of bias in prognostic studies. *Cochrane Database Syst Rev* 2010; (Suppl): 23.
94. Cohen J (Ed). *Statistical power analysis for the behavioral sciences*. New York: Academic Press; 1977.
95. Fayers PM, Hays R (Ed). *Assessing quality of life in clinical trials: methods and practice*. Oxford: Oxford University Press; 2005.
96. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
97. Lange S. The all randomized/full analysis set (ICH E9): may patients be excluded from the analysis? *Drug Inf J* 2001; 35(3): 881-891.
98. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156(10): 1411-1416.
99. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
100. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
101. Leemis LM, Trivedi KS. A comparison of approximate interval estimators for the Bernoulli parameter. *Am Stat* 1996; 50(1): 63-68.
102. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005; 58(10): 982-990.
103. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006; 59(12): 1331-1332.
104. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010; 49(1): 54-64.
105. Hotelling H. The generalization of student's ratio. *Ann Math Statist* 1931; 2(3): 360-378.

Anhang A – Offenlegung potenzieller Interessenkonflikte von Patientenvertretern

Im Folgenden sind die potenziellen Interessenkonflikte der Patientenvertreter zusammenfassend dargestellt, die an der Konsultation zur Festlegung patientenrelevanter Zielgrößen teilgenommen haben. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“.

Potenzielle Interessenkonflikte der konsultierten Patientenvertreter

Organisation / Institution / Firma / privat	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Deutsche Alzheimer Gesellschaft e. V.	Jansen, Sabine	Nein	Nein	Nein	Nein	Ja	Nein
	Von Lützau-Hohlbein, Heike	Nein	Nein	Nein	Nein	Ja	Nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt (Version 12/2011):

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor angestellt bei einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere bei einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor ein Unternehmen, eine Institution oder einen Interessenverband im Gesundheitswesen, insbesondere ein pharmazeutisches Unternehmen, einen Hersteller von Medizinprodukten oder einen industriellen Interessenverband, direkt oder indirekt beraten?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband, Honorare erhalten für Vorträge, Stellungnahmen oder Artikel?

Frage 4: Haben Sie und / oder hat die Einrichtung¹, die Sie vertreten, abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband, finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten?

Frage 5: Haben Sie und / oder hat die Einrichtung¹, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren ohne wissenschaftliche Gegenleistung) erhalten von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile eines Unternehmens oder einer anderweitigen Institution, insbesondere von einem pharmazeutischen Unternehmen oder einem Hersteller von Medizinprodukten? Besitzen Sie Anteile eines „Branchenfonds“, der auf pharmazeutische Unternehmen oder Hersteller von Medizinprodukten ausgerichtet ist?

¹ Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.