

**Positronenemissionstomographie
(PET) und PET/CT zur
Rezidivdiagnostik bei Gliomen mit
hohem Malignitätsgrad
(III und IV)**

Vorläufiger Berichtsplan

Auftrag D06-01D
Version 1.0
Stand: 09.11.2009

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Positronenemissionstomographie (PET) und PET/CT zur Rezidivdiagnostik bei Gliomen mit hohem Malignitätsgrad (III und IV)

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

21.12.2006

Interne Auftragsnummer:

D06-01D

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Dillenburger Str. 27
51105 Köln

Tel.: 0221/35685-0

Fax: 0221/35685-1

Berichte@iqwig.de

www.iqwig.de

Bei dem vorliegenden Berichtsplan handelt es sich um eine vorläufige Version. Zu diesem Berichtsplan können Stellungnahmen abgegeben werden, die zu einer Ergänzung und / oder Überarbeitung des Berichtsplans führen können. Die Frist für den Eingang der Stellungnahmen befindet sich auf der Website des IQWiG (www.iqwig.de), ebenso wie die dafür notwendigen Formblätter und ein Leitfaden.

Schlagwörter: Positronenemissionstomographie, Computertomographie, Gliome, Rezidiv, Behandlungsresiduen, systematische Übersicht.

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	vi
Abkürzungsverzeichnis	vii
1 Hintergrund	8
1.1 Definition des Krankheitsbildes	8
1.2 Epidemiologie und individuelle Krankheitslast	8
1.3 Ursache der Erkrankung	9
1.4 Verlauf der Erkrankung	9
1.5 Klassifikation	10
1.6 Diagnostische Verfahren	10
1.7 Konventionelle Diagnostik: CT und MRT	11
1.8 PET und PET/CT	11
1.9 Empfehlungen zum Einsatz der PET bei Gliomen	13
1.10 Derzeit übliche Therapiepraxis	15
2 Ziele der Untersuchung	16
2.1 Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT	16
2.2 Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT	17
3 Projektbearbeitung	18
3.1 Zeitlicher Verlauf des Projekts	18
4 Methoden	19
4.1 Ermittlung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT	20
4.1.1 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT anhand von systematischen Übersichten	21
4.1.2 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT anhand von Primärstudien	21
4.1.2.1 Population	21
4.1.2.2 Prüf- und Vergleichsintervention	21

4.1.2.3	Zielgrößen	22
4.1.2.4	Studientypen.....	23
4.1.2.5	Sonstige Studiencharakteristika	24
4.1.3	Tabellarische Darstellung der Einschlusskriterien („Nutzenstudien“).....	24
4.2	Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT.....	24
4.2.1	Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Güte anhand von systematischen Übersichten	25
4.2.1.1	Population.....	25
4.2.1.2	Prüftechnologie (Indextest I), Vergleichstechnologie (Indextest II) und Referenztest	25
4.2.1.3	Zielgrößen	26
4.2.1.4	Den systematischen Übersichten zugrunde liegende Studientypen	26
4.2.1.5	Sonstige Charakteristika der systematischen Übersichten	27
4.2.1.6	Tabellarische Darstellung der Einschlusskriterien (Sekundärliteratur zur diagnostischen und prognostischen Güte).....	27
4.2.2	Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Güte anhand von Primärstudien („Ergänzungsrecherche“).....	28
4.2.2.1	Population.....	28
4.2.2.2	Prüftechnologie (Indextest I), Vergleichstechnologie (Indextest II) und Referenztest	28
4.2.2.3	Zielgrößen	28
4.2.2.4	Studientypen.....	29
4.2.2.5	Sonstige Studiencharakteristika	29
4.2.2.6	Tabellarische Darstellung der Einschlusskriterien („Ergänzungsrecherche“ nach Primärstudien zur diagnostischen- und prognostischen Güte).....	29
4.3	Informationsbeschaffung.....	29
4.3.1	Bibliografische Literaturrecherche	29
4.3.2	Suche nach weiteren publizierten und nicht publizierten Studien.....	30
4.3.3	Selektion relevanter Studien	30
4.4	Informationsbewertung	30
4.4.1	Bewertung von Studien zum patientenrelevanten Nutzen und Schaden	31
4.4.2	Bewertung von systematischen Übersichten zur diagnostischen und prognostischen Güte	32

4.4.3	Bewertung von Primärstudien zur diagnostischen und prognostischen Güte.....	32
4.5	Informationssynthese und -analyse	33
4.5.1	Gegenüberstellung der Ergebnisse	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“).....	34
4.5.1.4	Gegenüberstellung der Ergebnisse aus systematischen Übersichten und der Ergänzungsrecherche („Robustheitsprüfung“).....	35
4.5.2	Meta-Analysen.....	35
4.5.3	Sensitivitätsanalysen.....	36
4.5.4	Subgruppenmerkmale und andere Effektmodifikatoren.....	36
5	Literaturverzeichnis	38

Tabellenverzeichnis

Tabelle 1: Hochgradige Gliome mit WHO-Grad und zugehöriger ICD-Kodierung	10
Tabelle 2: Empfehlungen der DGN und internationaler Fachgesellschaften zum Einsatz der PET bei Gliomen	14
Tabelle 3: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren	19
Tabelle 4: Evidenzklassifizierung des G-BA für Unterlagen zu diagnostischen Methoden	20
Tabelle 5: Tabellarische Darstellung der Einschlusskriterien für Primärstudien zur Nutzenbewertung der PET bzw. PET/CT	24
Tabelle 6: Tabellarische Darstellung der Einschlusskriterien für die Sekundärliteratur zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT	28
Tabelle 7: Tabellarische Darstellung der Einschlusskriterien für die Primärliteratur zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT (Ergänzungsrecherche)	29

Abkürzungsverzeichnis

Abkürzung	Bedeutung
CONSORT	Consolidated Standards of Reporting Trials
CCT	Controlled Clinical Trial
CT	Computertomographie
ECOG	Eastern Cooperative Oncology Group
EEG	Elektroenzephalographie
FDG	Fluor-Deoxy-Glukose
G-BA	Gemeinsamer Bundesausschuss
HTA	Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention-to-Treat
MRS	Magnetresonanztomographie
MRT	Magnetresonanztomographie
PET	Positronenemissionstomographie
PET/CT	Integration von PET (Positronenemissionstomographie) und CT (Computertomographie) in einem Untersuchungsgerät
RCT	Randomised Controlled Trial (randomisierte kontrollierte Studie)
SPECT	Single Photon Emission Computed Tomography
SUV	Standardized Uptake Value
SROC	Summary-Receiver-Operating-Characteristic
VEGF	Vascular endothelial growth factor
VOPT	Verification of Only Positive Testers

1 Hintergrund

1.1 Definition des Krankheitsbildes

Als Gliome bezeichnet man eine Gruppe von Tumoren des Gehirns und des Rückenmarks. Sie entstehen aus glialen Zellen und neuronalen Stammzellen. Die Namensgebung verschiedener Gliome basiert auf dem Zelltyp mit der höchsten histologischen Ähnlichkeit. Entsprechend erfolgte die Benennung von Astrozytomen, die Astrozyten (Sternzellen) ähneln, sowie Oligodendrogliomen (ähneln Oligodendrozyten) und Ependymomen, die Ependymzellen ähneln, welche Hirnventrikel und Rückenmark auskleiden [1].

Nach der aktuellen Klassifikation der WHO werden Gliome in 4 verschiedene Malignitätsgrade eingeteilt [2]. Im Vergleich zu niedriggradigen Gliomen (Grad I und II nach WHO) wachsen höhergradige Gliome (Grad III und IV nach WHO) üblicherweise schneller und infiltrativer, die Abgrenzung zum umliegenden Gewebe ist meist schwerer und sie neigen eher zur Metastasenbildung und Rezidivierung. Glioblastome stellen 60 bis 70 % der höhergradigen Gliome, während anaplastische Astrozytome für 10 bis 15 %, anaplastische Oligodendrogliome und anaplastische Oligoastrozytome für 10 % und anaplastische Ependymome und Gangliogliome für den Rest der Fälle verantwortlich sind [3].

Für diesen Bericht soll in erster Linie die Rolle der Positronenemissionstomographie (PET) bzw. von Kombinationsgeräten (PET/CT) in der Diagnostik, Prognostik und Therapieplanung bzw. -bewertung von Rezidiven von Gliomen mit hohem Malignitätsgrad (III/IV, vgl. Abschnitt 1.5 Klassifikation) betrachtet werden.

1.2 Epidemiologie und individuelle Krankheitslast

Maligne Neubildungen des Nervensystems machen mehr als 2 % aller malignen Neubildungen aus [4]. In Deutschland erkranken etwa 3700 Männer und etwa 3400 Frauen jährlich an bösartigen Neubildungen des Nervensystems. Etwa 10 % der Fälle treten bereits im Kindes- und Jugendalter auf. Mehr als die Hälfte der Fälle bei Männern und gut ein Drittel der Fälle bei Frauen treten vor dem 60. Lebensjahr auf [4].

Hochgradige Gliome zählen zu den bösartigsten Krebsformen, die mit massiven Einschränkungen neurologischer Funktionen einhergehen können und in den meisten Fällen eine infauste Prognose aufweisen [5].

In Deutschland beträgt die Inzidenz der Gliome etwa 5–6/100 000 pro Jahr [6]. In den USA machen Gliome mit hohem Malignitätsgrad (WHO III-IV) rund 70 % der jährlichen 22 500 Neuerkrankungen mit primären Hirntumoren aus. Die jährliche Inzidenz ist in den USA mit 5/100 000 vergleichbar mit derjenigen in Deutschland [3]. Nach Angaben des schwedischen Krebsregisters machen hochgradige Gliome 30,5 % aller primären Hirntumoren aus [7]. Das

Glioblastom (Glioblastoma multiforme) ist das häufigste Gliom und zugleich der häufigste bösartige Hirntumor des Erwachsenen [4].

Weltweit betrachtet beträgt die altersstandardisierte Mortalität für primäre bösartige Neubildungen des Gehirns 2,8/100 000 Personenjahre für Männer und 2,0/100 000 Personenjahre für Frauen. Im Vergleich zu weniger entwickelten Ländern wird eine höhere Mortalität in entwickelten Ländern beobachtet, mit 4,1/100 000 Personenjahre für Männer bzw. 2,7/100 000 Personenjahre für Frauen [7]. Unterschiedliche Inzidenzen in verschiedenen ethnischen Gruppen werden diskutiert [8].

1.3 Ursache der Erkrankung

Die Entstehung von Gliomen konnte bisher noch nicht vollständig geklärt werden und ist gegenwärtig Gegenstand aktueller Forschungsbemühungen.

Es gibt starke Hinweise darauf, dass Gliome nach Strahlentherapie wegen anderer Tumorerkrankungen gehäuft auftreten [3,4,7]. Ob ein Zusammenhang zwischen dem Auftreten von Gliomen und einer vorausgegangenen Hochdosis-Chemotherapie besteht, wird diskutiert [7]. Genetische Faktoren scheinen einen Einfluss auf die Entstehung von Gliomen nach einer Exposition mit äußeren Faktoren zu haben [8,9].

Ein Zusammenhang zwischen der Benutzung von Mobiltelefonen und der Entstehung von Gliomen scheint nicht zu bestehen [7,10]. Ein ebenfalls diskutierter Einfluss von elektromagnetischen Feldern, Nitrosamin in der Nahrung oder Kopfverletzungen konnte ebenfalls nicht belegt werden [3].

1.4 Verlauf der Erkrankung

Nach der Diagnose ist meist eine operative Entfernung die Therapie der ersten Wahl [11]. Bei hochgradigen Gliomen ist eine vollständige chirurgische Entfernung meistens nicht möglich. Das Ziel der Operation ist primär die Eingrenzung der Raumforderung und der mit ihr einhergehenden neurologischen Symptome. Die operative Entfernung wird durch Chemo- bzw. Strahlentherapie ergänzt [6].

Nach Daten der American Cancer Society (ACS) beträgt die 5-Jahres- und die 10-Jahres-Überlebensrate für Hirntumoren in den USA insgesamt 29,1 bzw. 25,3 % [7]. Diese Zahlen sind jedoch erheblich vom Allgemeinzustand (Karnofsky-Index) und dem Alter des Patienten sowie der Histopathologie des Tumors abhängig [3,7]. So hat das Glioblastoma multiforme eine 5-Jahres-Überlebensrate von 3,3 %, während die entsprechende Rate bei niedriggradigen Gliomen bei über 70 % liegt [7,12].

Bei Rezidiven von anaplastischen Astrozytomen können je nach Therapieansatz progressionsfreie Intervalle zwischen 13 und 23 Wochen bzw. ein progressionsfreies Überleben nach 6 Monaten von 31 bis 46 % der Patienten erzielt werden [13,14]. Bei

rezidivierenden Glioblastomen kommt für etwa 30 % aller Patienten eine erneute operative Entfernung in Betracht [6]. In Kombination mit unterschiedlichen Chemotherapieprotokollen können mediane progressionsfreie Intervalle zwischen 9 und 25 Wochen bzw. ein progressionsfreies Überleben nach 6 Monaten von 15 bis 44 % aller Patienten mit Rezidiv erzielt werden [6].

1.5 Klassifikation

Nach ihrer histologischen Erscheinung werden Gliome entsprechend der WHO-Klassifikation von 2007 in 4 Grade unterteilt [15]. Der vorliegende Bericht beschränkt sich auf Rezidive der in Tabelle 1 dargestellten hochgradigen Tumoren (Grad III und IV nach WHO). Als hochgradig werden dabei Gliome eingestuft, die aggressiv-infiltrativ wachsen. Dabei wird mit der Einteilung in Grad III die Erwartung einer Reduktion und in den Grad IV einer deutlichen Reduktion der Lebenszeit durch die Tumorerkrankung zum Ausdruck gebracht [15].

Tabelle 1: Hochgradige Gliome mit WHO-Grad und zugehöriger ICD-Kodierung

Tumor	WHO-Grad [2]	ICD-O [16]
Astrozytome		
Anaplastisches Astrozytom	III	9401/3
Glioblastom	IV	9440/3
Riesenzellglioblastom	IV	9441/3
Gliosarkom	IV	9442/3
Gliomatosis cerebri	III	9381/3
Oligodendrogliome		
Anaplastisches Oligodendrogliom	III	9451/3
Tumore des Plexus choroideus		
Plexuskarzinom	III	9390/3
Ependymome		
Anaplastisches Ependymom	III	9392/3
Gemischte Tumoren		
Anaplastisches Oligoastrozytom	III	9382/3

1.6 Diagnostische Verfahren

Neurologische Symptome wie Kopfschmerzen, Krampfanfälle, fokale Ausfälle, Verwirrung, Gedächtnisstörungen und Persönlichkeitsveränderungen sind häufig die ersten Anzeichen bei Patienten mit Gliomen, welche zu einer Vorstellung bei einem Arzt führen [3]. Bei der Diagnose von Gliomen können entsprechend den Leitlinien der Deutschen Gesellschaft für Neurologie folgende Schritte zum Einsatz kommen [6]:

- Anamneseerhebung
- Klinische Untersuchung
- Neuroradiologische Diagnostik
- Elektroenzephalographie (EEG)
- Biopsie oder Resektion
- Zytologische Diagnostik und Grading
- Ggf. Liquordiagnostik (z. B. bei Verdacht auf ein Ependymom)

1.7 Konventionelle Diagnostik: CT und MRT

In der 2008 veröffentlichten Leitlinie „Gliome“ der Deutschen Gesellschaft für Neurologie wird die Magnetresonanztomographie (MRT) ohne und mit Kontrastmittel als Methode der Wahl bei klinischem Verdacht auf einen Hirntumor bzw. bei (Verdacht auf) Rezidiv genannt. MRT-Kontrollen nach 3 (Glioblastom) bzw. 4 Monaten (anaplastisches Astrozytom, anaplastisches Oligodendrogliom, Oligoastrozytom) werden empfohlen, wenn keine klinischen Hinweise auf eine Verschlechterung (Progression bzw. Rezidiv) bestehen [6].

Bei bestimmten Fragestellungen, wie z. B. in der Differenzialdiagnose von Oligodendrogliomen, kann die Computertomographie (CT) durch die bessere Möglichkeit, Verkalkungen darzustellen, ergänzende Informationen liefern [6]. Neben der konventionellen Diagnostik mit CT und MRT werden in der Diagnostik von Hirntumoren unter anderem die Single Photon Emission Computed Tomography (SPECT), die Positronenemissionstomographie (PET), die funktionelle MRT und die Magnetresonanzspektroskopie (MRS) eingesetzt [3,6].

1.8 PET und PET/CT

Die PET ist ein minimalinvasives diagnostisches Verfahren, das freigesetzte Photonen registriert. 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 angeordnete Photonen-Detektoren werden die dabei 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 so genannte Nicht-Vollringsysteme 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 Detektor-Ring, bei modernen Systemen werden heute

mehrere Ringe nebeneinander angeordnet, wodurch die Sensitivität des Gesamtsystems noch weiter gesteigert wurde [6].

Im Gegensatz zu den klassischen 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 von Gliomen werden meist Fluor-18-2-fluoro-2-deoxy-D-Glukose (FDG) oder Aminosäure-basierte Tracer eingesetzt [17-19].

Zur Befundung wird unter anderem die Stoffwechselaktivität eines bestimmten Areals mit der des umliegenden „normalen“ Gewebes verglichen. Aufgrund von Unterschieden in der physiologischen Aufnahme des Tracers in verschiedenen Körperregionen kann die Zuordnung einer bestimmten Region als potenziell pathologischer sogenannter „erhöhter Uptake“ schwierig sein und erfordert erhebliche Expertise des Untersuchers. Um bei der Interpretation eine gewisse Standardisierung zu erreichen, kann der Metabolismus als Standardized Uptake Value (SUV) wiedergegeben werden. Allerdings ist ein erhöhter Umsatz nicht malignom-spezifisch, vielmehr können sich auch andere Prozesse mit erhöhtem Umsatz dahinter verbergen (z. B. Entzündungen).

Die Aktivitätsanreicherungen bei PET lassen sich anatomisch jedoch nicht immer präzise lokalisieren, da in PET-Bildern in erster Linie Stoffwechselprozesse abgebildet werden; hinzu kommt die begrenzte Ortsauflösung von etwa 4 bis 6 mm. Eine technische Weiterentwicklung stellt eine Kombination von PET und CT in einem Untersuchungsgerät (PET/CT) dar. Hier wird der Patient in einem Untersuchungsgang durch die beiden Detektorringssysteme von CT und PET gefahren. Die entstehenden Bilder werden im Computer fusioniert. Üblicherweise wird die CT-Information in Graustufen abgebildet und von der PET-Information farbig überlagert. Mit dieser Methode wird die hohe Auflösung der CT (Ortsauflösung < 1 mm) mit den hochsensitiven Stoffwechselfinformationen der PET verknüpft. Die kombinierten PET/CT-Geräte sind seit dem Jahr 2001 auf dem Markt und verdrängen seitdem reine PET-Scanner bei den Neuanschaffungen in deutschen Krankenhäusern [20].

Mittels moderner Software lässt sich eine Überlagerung (sogenannte Soft-Fusion bzw. Koregistrierung) von CT-Bildern und PET-Daten auch aus separaten Untersuchungsschritten berechnen. Dies geschieht mithilfe gemeinsamer Referenzpunkte wie etwa Knochenstrukturen oder anhand externer Positionsdaten. Die Bildqualität eines durch eine solche Rekonstruktion von PET- und CT-Daten entstehenden Bildes ist allerdings i. d. R. schlechter als die einer kombinierten PET/CT.

Erste Studien zur PET in der Diagnostik von Hirntumoren wurden bereits in den frühen 1980er-Jahren veröffentlicht [21]. Mögliche Einsatzgebiete der PET bzw. von kombinierten PET/CT-Geräten sind die Diagnose und das Staging bzw. Restaging von Gliomen, das Erkennen von Rezidiven sowie die Therapieplanung bzw. -bewertung [22,23].

1.9 Empfehlungen zum Einsatz der PET bei Gliomen

Die 2008 veröffentlichte Leitlinie der Deutschen Gesellschaft für Neurologie verweist darauf, dass „modernere bildgebende Methoden (Single photon emission computed tomography [SPECT], Positronenemissionstomographie [PET], Magnetresonanztomographie [MRS], funktionelle MRT) für die klinische Routine Gegenstand aktueller Untersuchungen“ sind [6].

Eine Arbeitsgruppe der Deutschen Gesellschaft für Nuklearmedizin (DGN), hat im Jahr 1995 in einer interdisziplinären Konferenz mit Experten aus 9 Fachgebieten eine Bewertung des Standes der PET in der Onkologie erarbeitet [24]. Diese Arbeitsgruppe tagte im Juli 1997 zum 2. Mal, um eine Indikationsliste zum klinischen Einsatz der PET bei neurologischen und psychiatrischen Erkrankungen zu erstellen [25]. Mittlerweile wurden bereits die Ergebnisse der 3. Konsensuskonferenz zur PET-Diagnostik bei onkologischen Erkrankungen publiziert [26].

Im Folgenden sind die von der DGN und 3 internationalen Fachgesellschaften bisher empfohlenen Indikationen für eine PET-Diagnostik von malignen Gliomen tabellarisch aufgelistet:

Tabelle 2: Empfehlungen der DGN und internationaler Fachgesellschaften zum Einsatz der PET bei Gliomen

Indikation	Deutsche Gesellschaft für Nuklearmedizin (DGN) 2001 Onko-PET [26]	Schweizerische Gesellschaft für Nuklearmedizin (SGNM) 2008 [27]³	American College of Radiology (ACR) 2007 [22]	European Association of Nuclear Medicine (EANM) 2006 [19]
Tracer: Fluor-18-2-fluoro-2-deoxy-D-Glukose (FDG)				
1 Differenzierung zwischen Rezidiv und Strahlennekrose bei malignen Gliomen	empfohlen 1a ¹	empfohlen	empfohlen	— ²
2 Erkennung der malignen Entdifferenzierung eines Gliomrezidivs (bzw. -reliktes)	empfohlen 1a ¹	empfohlen	empfohlen	— ²
3 Bestimmung des Biopsieortes bei Verdacht auf Gliom	empfohlen 1a ¹	empfohlen	empfohlen	— ²
4 Beurteilung der biologischen Aggressivität von Hirntumoren (Raumforderungen)	empfohlen 1b ¹	empfohlen	empfohlen	— ²
5 Nachweis von Resttumor bei malignen Gliomen nach Operation	empfohlen 1b ¹	empfohlen	empfohlen	— ²
6 Differenzierung zwischen zerebralem Lymphom und Toxoplasmose	empfohlen 1b ¹	— ²	— ²	— ²
Tracer: markierte Aminosäure				
7 Differenzierung zwischen Rezidiv und Strahlennekrose bei malignen Gliomen	— ²	— ²	— ²	empfohlen
8 Abgrenzung des Tumors	— ²	— ²	— ²	empfohlen
9 Bestimmung des besten Entnahmeortes als Vorbereitung einer Biopsie	— ²	— ²	— ²	empfohlen
10 Nicht-invasives Grading von Hirntumoren	— ²	— ²	— ²	empfohlen
11 Therapieplanung bzw. -bewertung	— ²	— ²	— ²	empfohlen
1: 1a: „Angemessen, klinischer Nutzen ist erwiesen“, 1b: „Akzeptabel, Ergebnisse deuten auf klinischen Wert hin“				
2: Keine Angabe				
3: Nach Angaben der Autoren an die Empfehlungen der ACR angelehnt				

Dieser Bericht bewertet lediglich den Einsatz der PET bzw. PET/CT im Rahmen der Rezidivdiagnostik (vgl. Kapitel 2). Daher sind nicht alle in Tabelle 2 angeführten

Indikationen für diesen Bericht relevant. Außerdem könnten bei der Sichtung neuerer Arbeiten weitere Indikationen identifiziert und im Rahmen der Rezidivdiagnostik betrachtet werden.

1.10 Derzeit übliche Therapiepraxis

Nach der Erstdiagnose von hochgradigen Gliomen wird i. d. R. eine chirurgische Entfernung des Tumors angestrebt. Aufgrund ihres diffus-infiltrativen Wachstums sind hochgradige Gliome meist nicht vollständig zu entfernen. Daher wird meist eine postoperative oder bei Inoperabilität primär palliative Bestrahlung durchgeführt, die oft mit einer Chemotherapie kombiniert wird [3,6]. Es wird davon ausgegangen, dass eine Strahlentherapie, die zusätzlich zu einer chirurgischen Tumorentfernung durchgeführt wird, das mediane Überleben von Patienten mit Glioblastomen um 3 bis 4 Monate auf eine Spanne von 7 bis 12 Monaten verlängert [28,29]. Eine Chemotherapie zusätzlich zur postoperativen Strahlentherapie kann die 1-Jahres-Überlebensrate absolut um bis zu 10 % steigern [30]. Die 1-Jahres-Überlebensrate kann so auf 46 % angehoben werden [5].

Mögliche Erfolge der Therapie sind meist auf eine Lebensverlängerung beschränkt; eine Heilung ist nur in den seltensten Fällen möglich [4]. Das weitere therapeutische Vorgehen wird vom Alter und Zustand des Patienten sowie klinischen Faktoren, wie z. B. die erreichbare Radikalität der Resektion oder der Tumorgrad und die Tumorlokalisation, beeinflusst [6]. Viele therapeutische Maßnahmen erfolgen dabei beschwerdeorientiert und auch zur Behandlung der häufig mit hochgradigen Gliomen einhergehenden Begleitkomplikationen wie Krampfanfälle, Ödeme im Bereich der Neubildung, venöse Thrombosen oder Erschöpfungszustände [3,6].

Im Falle eines Rezidivs kann eine erneute Operation empfehlenswert oder erforderlich sein. Strahlen- und Chemotherapie kommt bei verschiedenen Gliom-Subtypen zum Einsatz [6]. Die Indikationen werden derzeit noch diskutiert und sind u. a. abhängig von verschiedenen klinischen Faktoren sowie dem Ursprungsgewebe des Tumors [3]. Die Ergebnisse aus PET bzw. PET/CT-Untersuchungen könnten dabei einen unmittelbaren Einfluss auf die weitere Behandlung, z. B. auf die Wahl des Zielgebietes für eine Strahlentherapie haben [6].

2 Ziele der Untersuchung

Die vorliegende Untersuchung soll als Erstes prüfen, welcher patientenrelevante Nutzen und Schaden mit der bildgebenden Diagnostik mittels PET bzw. PET/CT im Vergleich oder in Ergänzung zur konventionellen Diagnostik bei der Rezidiverkennung von Gliomen mit hohem Malignitätsgrad (WHO III und IV) verbunden ist.

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 von PET bzw. PET/CT die prognostische bzw. diagnostische Güte der Diagnosestrategie verbessert.

Eine präzise Abgrenzung der Definitionen für die Primärtherapie, Progression und Rezidiverkennung wird bei malignen Gliomen als schwierig beschrieben [31]. Zum einen ist durch das infiltrative Wachstum eine vollständige Entfernung des Tumors im Rahmen der Primärtherapie in den meisten Fällen nicht möglich, sodass nur äußerst selten von einer kompletten Remission ausgegangen werden kann. Daher werden die Begriffe Rezidiv und Progression häufig synonym verwendet [31]. Zum anderen existiert eine große Zahl unterschiedlicher Definitionen. Diese reichen von dem Wechsel zwischen einem tumorfreien Intervall und einer Phase des unkontrollierbaren Wachstums [31] bis hin zu einer bestimmten prozentualen Zunahme des Volumens (oder Durchmessers) im Vergleich zum Residuum (z. B. 25 % [32]). Andere Definitionen legen eine prozentuale Vergrößerung der Aufnahmewerte zugrunde (z. B. 50 % [33]). Aus diesen Gründen wird im Rahmen dieses Berichts jeweils die Definition der eingeschlossenen Studien verwendet.

Alle Befunde werden im Rahmen der deutschen Strahlenschutzgesetzgebung erörtert.

2.1 Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT

Das 1. Ziel dieses Berichts ist die Bewertung des patientenrelevanten Nutzens und Schadens der Bildgebung mit PET bzw. PET/CT bei der Behandlung von Patienten mit malignen Gliomen.

Als zu untersuchende Indikation für den Einsatz des Verfahrens bei dieser Patientenpopulation wird gemäß Auftragskonkretisierung durch die Arbeitsgruppe des G-BA der Nachweis eines Rezidivs betrachtet.

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 werden in diesem Bericht primär die Computertomographie (CT) und die Magnetresonanztomographie (MRT) verstanden. Es sollen aber darüber hinaus alle in den gefundenen Studien angestellten diagnostischen Vergleiche dargestellt werden.

Bei allen Vergleichen werden Studien, in denen der Tracer FDG verwendet wurde, von solchen, in denen Aminosäure-Tracer oder andere zur Anwendung kamen, getrennt ausgewertet. Die unterschiedliche Biologie der Tracer spricht für unterschiedliche diagnostische Methoden und steht daher u. U. einer aggregierten Auswertung dieser Studien entgegen.

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 (darunter auch das Auftreten erneuter Rezidive) 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 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, einer Scheinintervention oder keiner Intervention [34].

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

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. Das impliziert folgende Fragestellungen:

- Verbessert sich durch den (zusätzlichen) Einsatz der PET bzw. PET/CT die diagnostische Güte bei Erkennung von Gliom-Rezidiven (III und IV WHO)?
- Lassen sich mittels PET und PET/CT zuverlässigere prognostische Aussagen über die Wahrscheinlichkeit für schwerwiegende Ereignisse (Hospitalisierung, Notwendigkeit einer Operation, Tod) machen, als es mit den konventionellen diagnostischen Verfahren möglich ist?

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

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 Positronen-emissionstomographie (PET), auch als Integrationsgerät von PET und Computertomographie (PET/CT), bei 14 verschiedenen Krankheitsbildern bzw. insgesamt 21 Indikationen beauftragt. Der vorliegende vorläufige Berichtsplan bezieht sich auf den Teilauftrag zu der Indikation Rezidiverkennung bei Gliomen mit hohem Malignitätsgrad (WHO III und IV).

Die endgültigen Konkretisierungsunterlagen wurden dem Institut mit dem Schreiben des G-BA vom 11.03.2008 übersandt. In die Bearbeitung des Projekts sind externe Sachverständige eingebunden.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Zur Erstellung des Berichtsplans wurde am 04.09.2009 ein Patientenvertreter der Deutschen Hirntumorhilfe e. V. konsultiert.

Der vorliegende vorläufige Berichtsplan (Version 1.0) wird zur Anhörung gestellt. Hierzu können schriftliche Stellungnahmen eingereicht werden. Das Ende der Stellungnahmefrist wird auf der Website des IQWiG (www.iqwig.de) bekannt gegeben. Stellungnahmen können von allen interessierten Personen, Institutionen und Gesellschaften abgegeben werden. Die Stellungnahmen müssen bestimmten formalen Anforderungen genügen, die ebenfalls auf den Internetseiten des Instituts in einem entsprechenden Leitfaden dargelegt sind. Gegebenenfalls wird eine wissenschaftliche Erörterung zur Klärung unklarer Aspekte aus den schriftlichen Stellungnahmen durchgeführt. Die Anhörung kann zu Änderungen und / oder Ergänzungen des Berichtsplans führen. Im Anschluss an diese Anhörung wird der dann gültige Berichtsplan publiziert.

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. seine Fähigkeit, kranke von gesunden Personen zu unterscheiden [35]. 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 [36,37]. Fryback und Thornbury haben dazu in ihrer Publikation „The efficacy of diagnostic imaging“ von 1991 [36] 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 vierte Stufe und den patientenrelevanten Auswirkungen („patient outcome efficacy“) als fünfte Stufe reichen. Als oberste Stufe betrachten sie schließlich die gesellschaftlichen Konsequenzen („societal efficacy“) der diagnostischen Technologie (s. auch Tabelle 3).

Tabelle 3: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren

	Köbberling et al. [37]	Fryback & Thornbury [36]	Evidenzklassifizierung G-BA [38]
Patientenrelevante Zielgrößen	Phase 4: Wirksamkeit	6. Stufe: Auswirkung auf Systemebene	Evidenzstufe I
		5. Stufe: Auswirkung im Hinblick auf patientenrelevante Endpunkte	
Surrogatzielgröße (Management-änderung)		4. Stufe: Auswirkung im Hinblick auf das therapeutische Denken des Behandelnden	Evidenzstufe II
		3. Stufe: Auswirkung im Hinblick auf das (differenzial-) diagnostische Denken („Nachtstwahrscheinlichkeit“)	
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 von 2006 bei der Bewertung diagnostischer Tests (§ 18 Abs. 7) dieser Sichtweise angeschlossen [38]. In Tabelle 3 sind verschiedene Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren einander gegenübergestellt. Neben der Einteilung von Fryback und Thornbury und der Evidenzklassifizierung des G-BA (s. Tabelle 4) wurde auch das Phasenschema nach Köbberling et al. [37] mit aufgenommen.

Tabelle 4: Evidenzklassifizierung des G-BA für Unterlagen zu diagnostischen Methoden

Evidenzklasse nach G-BA [38]	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 Behandlung von Patienten mit Gliomen zu bewerten, genügt es nicht festzustellen, inwieweit die PET bzw. die PET/CT den gängigen diagnostischen Methoden hinsichtlich einer korrekten Identifizierung des Rezidivs (bzw. seiner Ausbreitung im Körper) 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 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 „Management-Studien“). Aber auch andere Studiendesigns kommen infrage [39]. Wie sich die Behandlungsoptionen (z. B. aggressiveres oder weniger aggressives Vorgehen, Chemo- oder Strahlentherapie oder beides) verändern und welche Auswirkungen die unterschiedlichen Strategien schließlich auf patientenrelevante Endpunkte haben, lässt sich am besten durch Studien mit ausreichender Nachbeobachtungszeit beantworten (s. auch Abschnitt 4.2.1.3).

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

Vorrecherchen haben ergeben, dass vermutlich keine qualitativ ausreichende systematische Übersicht zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bei Gliomen vorliegt. Daher soll der patientenrelevante Nutzen in dem zu erstellenden Bericht nicht auf Basis systematischer Übersichten sondern anhand von Primärstudien (sogenannten „Nutzen-Studien“) 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 Population

Studien zu folgenden Patientenpopulationen werden eingeschlossen:

- Patienten mit nachgewiesenen Gliomen mit hohem Malignitätsgrad (WHO III und IV) gemäß WHO-Klassifikation nach Primärtherapie, bei denen aufgrund eines begründeten Verdachts oder auch im Rahmen routinemäßiger Nachuntersuchungen ein Rezidiv erkannt bzw. ausgeschlossen werden soll.

Zu den Gliomen mit hohem Malignitätsgrad zählen im Rahmen dieses Berichts alle in Tabelle 1 (WHO-Klassifikation) angeführten Tumorentitäten.

Sofern in einzelnen Studien die Zahl der Patienten mit anderen histologischen Typen maligner Gliome, die nicht WHO-Grad III oder IV entsprechen, einen Anteil von 20 % der ursprünglich eingeschlossenen Patienten nicht überschreitet und die Ergebnisse für die eingeschlossene Population nicht separat dargestellt sind, werden die entsprechenden Arbeiten ebenfalls berücksichtigt.

4.1.2.2 Prüf- und Vergleichsintervention

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 Prüflintervention 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 [39].

4.1.2.3 Zielgrößen

Als Zielgrößen werden die nachfolgend angeführten Endpunkte verwendet, die eine Beurteilung patientenrelevanter Interventionsziele ermöglichen:

- Mortalität
- Ereignisfreies Überleben
- Morbidität (z. B. unerwünschte Ereignisse durch aggressivere / weniger aggressive Behandlung)
- Funktionalität (z. B. Sprechen, Bewegen, Karnofsky- oder ECOG-Index)
- Gesundheitsbezogene Lebensqualität (z. B. SF-36)

Da Gliome mit hohem Malignitätsgrad die Lebenserwartung in den meisten Fällen deutlich reduzieren, erfolgt die Bewertung der Zielgrößen ereignisfreies Überleben, Morbidität und Lebensqualität in Abhängigkeit von der Zielgröße Mortalität.

Eine Änderung lediglich im Patientenmanagement (ohne Verknüpfung zu den oben genannten Zielgrößen) stellt keine patientenrelevante Zielgröße im Sinne dieses Berichts dar. Änderungen im klinischen Management der Patienten, die sich unmittelbar auf patientenrelevante Endpunkte auswirken (wie z. B. Vermeidung von nebenwirkungsreichen Therapien, Vermeidung von unnötigen, schmerzhaften diagnostischen Verfahren, Zeitdauer bis zur Sicherung der Diagnose) werden ergänzend dargestellt.

Neben den hier aufgelisteten Zielen wurde bei dem Gespräch mit Patientenvertretern am 04.09.2009 der Informationszugewinn als ein weiteres Behandlungsziel genannt. Je nach

Indikation und Tumorentität könne valide Information aus Patientensicht sehr hilfreich sein und auch die gesundheitsbezogene Lebensqualität unmittelbar verbessern (z. B. der Nicht-nachweis eines Rezidivs in einer Kontrolluntersuchung). Außerdem wurde genannt, dass alleine das Wissen der Patienten darum, dass alle diagnostischen Verfahren ausgeschöpft wurden, einen beruhigenden Effekt haben könne. Andererseits könne – als negative Konsequenz – die frühzeitige Diagnose eines nicht kurativ behandelbaren Rezidivs oder die Feststellung nicht behandelbarer Metastasen auch die sorgenfreie Lebenszeit verkürzen. Außerdem müsse auch für potenziell heilbare Tumoren, die durch eine verbesserte Diagnostik zusätzlich gefunden würden, zunächst geprüft werden, ob vorhandene Therapien auch für diese Entitäten effektiv anwendbar sind. Damit erweist sich der Zugewinn an Information als intermediärer Endpunkt, wie er in der Fragestellung 2 dieses Berichts untersucht wird. Der diagnostische Informationszugewinn müsste daher, sofern er in relevanten Änderungen des therapeutischen Managements oder der Lebensführung des Patienten mündet, in letzter Konsequenz einen Effekt auf die oben angeführten patientenrelevanten Endpunkte haben und entsprechend gemessen werden können. Aus den genannten Gründen und aufgrund der Probleme mit der Operationalisierbarkeit wird der Endpunkt „Informationszugewinn“ nicht in die Bewertung aufgenommen.

4.1.2.4 Studientypen

Randomisierte kontrollierte Studien (RCTs) liefern für die Bewertung des Nutzens und Schadens einer medizinischen Intervention die zuverlässigsten Ergebnisse, weil sie, sofern methodisch adäquat und der jeweiligen Fragestellung angemessen durchgeführt, mit der geringsten Ergebnisunsicherheit behaftet sind. 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 prospektiv geplante, 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 prospektiv geplante, 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 erlauben möglicherweise nicht randomisierte kontrollierte Interventionsstudien – wie oben skizziert – Hinweise auf 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.

Tabelle 5: Tabellarische Darstellung der Einschlusskriterien für Primärstudien zur Nutzenbewertung der PET bzw. PET/CT

Einschlusskriterien für Primärstudien zur Evaluierung des patientenrelevanten Nutzens und Schadens	
E1a	Patienten im Zustand nach Primärtherapie von Gliomen mit hohem Malignitätsgrad (WHO III und IV) wie in Abschnitt 4.1.2.1 beschrieben
E2a	Vergleich von 2 diagnostischen Strategien wie unter 4.1.2.2 beschrieben
E3a	Erhebung von patientenrelevanten Endpunkten wie unter 4.1.2.3 beschrieben
E4a	Randomisierte kontrollierte Studien (Bei nicht hinreichender 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 Zusatzinformation
E6a	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 CONSORT- [40], STROBE- [41] oder MOOSE-Statements [42] genügt und eine Bewertung der Studie ermöglicht.	

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

In den orientierenden Vorrecherchen fanden sich 7 HTA-Berichte, die Teilaspekte der Fragestellungen zur diagnostischen Güte der PET oder PET/CT bei Gliomen 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 HTA-Berichten und systematischen Übersichten erfolgen, sofern sichergestellt ist, dass sie bestimmte Mindestanforderungen erfüllen. Findet sich keine geeignete systematische Übersicht bzw. kein geeigneter HTA-Bericht zu einer Fragestellung, wird die diagnostische und prognostische Güte anhand von Primärstudien beurteilt. Ist die Arbeit nicht mehr aktuell (Suchzeitraum vor Februar 2009), wird für den Zeitraum, den sie nicht abdeckt, relevante Primärliteratur recherchiert und im Bericht berücksichtigt (Ergänzungsrecherche).

Nachfolgend werden zunächst die Kriterien für einen Einschluss von systematischen Übersichten / HTA-Berichten 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 Population

Die systematischen Übersichten müssen Angaben enthalten zu Patienten im Zustand nach Primärtherapie von Gliomen mit hohem Malignitätsgrad (WHO III und IV; vgl. Abschnitt 4.1.2.1).

4.2.1.2 Prüftechnologie (Indextest I), Vergleichstechnologie (Indextest II) und Referenztest

Die in den systematischen Übersichten zugrundeliegenden 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 2.1).

Indextest I und Indextest II müssen zu einem Referenztest (= der beste zur Verfügung stehende Test) in Beziehung gebracht worden sein, der für eine adäquate Interpretation der Ergebnisse unabhängig von Index- und Vergleichstest erhoben worden ist, da es anderenfalls zu Verzerrungen der Güte- und/oder Variabilitätsschätzer kommen kann (sogenannter Verifikations-Bias oder Inkorporations-Bias) [43-47].

Als Referenztest (= bester zur Verfügung stehender Test) gilt bei der Bewertung der diagnostischen Güte der histopathologische Befund ggf. in Kombination mit dem

individuellen Krankheitsverlauf und bei der Bewertung der prognostischen Güte der tatsächlich beobachtete individuelle Krankheitsverlauf im Rahmen der Nachbeobachtung.

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 Genauigkeit mittels Sensitivität / Spezifität (ggf. ROC-Kurven) oder positivem / negativem Likelihood-Quotienten, daraus abgeleitet ggf. prädiktive Aussagen mittels z. B. positivem / negativem prädiktiven Wert. Dabei müssen die Vierfeldertafel-Daten zur Berechnung der diagnostischen Güte der Primärstudien ableitbar sein.
- Prognostische Aussagen (Wahrscheinlichkeit) hinsichtlich zukünftiger klinischer Ereignisse (z. B. Überleben, Rezidiv) anhand geeigneter statistischer Maßzahlen (z. B. aus Kaplan-Meier-Schätzungen)

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

Querschnittstudien, ggf. mit klinischer Nachbeobachtung, durchgeführt in der Anwendungssituation (gemäß Phase 3 nach Köbberling et al. [37])

- Prognostische Güte

Prospektiv geplante Kohortenstudien: Um die prognostische Güte der PET möglichst unverzerrt bestimmen zu können, sollte eine Gruppe von Patienten, die zu einem bestimmten Zeitpunkt rekrutiert und mit PET untersucht wurde, im zeitlichen Verlauf untersucht werden (Longitudinalstudie). Dabei ist zum einen ein konsekutiver Einschluss der Patienten und die Dokumentation aller fehlenden Werte notwendig. Zum Zweiten sollten die PET-Positiven und die PET-Negativen zeitlich parallel, prospektiv und im direkten Vergleich beobachtet werden (historische Vergleiche oder der Vergleich positiver und negativer Ergebnisse aus unterschiedlichen Patientenkollektiven liefern keine verlässlichen Daten). Im Idealfall sollte die weitere Behandlung dieser Patienten ohne das Wissen um die Ergebnisse der PET erfolgen (Verblindung).

Für die Interpretation des Überlebens ist ein Follow-up von mindestens 6 Monaten entscheidend. 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 Referenztest 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 Referenztest untersucht wurden (s. 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 waren, mit dem Referenztests untersucht werden [48].

4.2.1.5 Sonstige Charakteristika der systematischen Übersichten

Neben diesen genannten Charakteristika müssen die 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 Publikationen auf dem Qualitätsindex von Oxman und Guyatt mindestens 5 von 7 möglichen Punkten erreichen [49-51].
- Die jeweilige Forschungsfrage wurde im Format einer sogenannten PICO-Frage formuliert, die Population, Indextest I, ggf. Indextest II, Referenztest 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) wurde adressiert.
- Die Ergebnisse der Qualitätsbewertung, die Kerninformationen über die Population und die Ergebnisse einer jeden Studie sind transparent und vergleichbar abgebildet, und die Daten können eindeutig mit den zugehörigen Studien verknüpft werden.

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

Es werden alle systematischen Übersichten einbezogen, die alle nachfolgenden Einschlusskriterien erfüllen.

Tabelle 6: Tabellarische Darstellung der Einschlusskriterien für die Sekundärliteratur zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT

Einschlusskriterien zur Evaluierung der diagnostischen und prognostischen Güte anhand von systematischen Übersichten	
E1b	Patienten im Zustand nach Primärtherapie von Gliomen mit hohem Malignitätsgrad (WHO III und IV) wie in Abschnitt 4.1.2.1 beschrieben
E2b	Prüftechnologie: Vollring-PET bzw. PET/CT, Evaluierung wie in Abschnitt 4.2.1.2 beschrieben
E3b	Zielgrößen wie in Abschnitt 4.2.1.3 beschrieben
E4b	Systematische Übersichten aus prospektiv geplanten Querschnitt- oder Kohortenstudien wie in den Abschnitten 4.2.1.4 und 4.2.1.5 beschrieben
E5b	Keine Mehrfachpublikationen ohne relevante Zusatzinformation
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 [52] genügt und eine Bewertung der Studie ermöglicht.	

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 Population

Patienten im Zustand nach Primärtherapie von Gliomen mit hohem Malignitätsgrad (WHO III und IV) wie in Abschnitt 4.1.2.1 beschrieben.

4.2.2.2 Prüftechnologie (Indextest I), Vergleichstechnologie (Indextest II) und Referenztest

Prüftechnologie ist die Vollring-PET bzw. -PET/CT. Vergleichstechnologie, Referenztest und Evaluierung wie in Abschnitt 4.2.1.2 beschrieben.

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 (vgl. Abschnitt 4.2.1.3).

4.2.2.4 Studientypen

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

4.2.2.5 Sonstige Studiencharakteristika

Es erfolgt keine weitere Einschränkung.

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

Es werden alle Studien einbezogen, die alle nachfolgenden Einschlusskriterien erfüllen.

Tabelle 7: Tabellarische Darstellung der Einschlusskriterien für die Primärliteratur zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT (Ergänzungsrecherche)

Einschlusskriterien für Primärstudien zur Evaluierung der diagnostischen und prognostischen Güte	
E1c	Patienten im Zustand nach Primärtherapie von Gliomen mit hohem Malignitätsgrad (WHO III und IV) wie in Abschnitt 4.1.2.1 beschrieben
E2c	Prüftechnologie: Vollring-PET bzw. PET/CT, Evaluierung wie in Abschnitt 4.1.2.2 beschrieben
E3c	Zielgrößen wie in Abschnitt 4.2.1.3 beschrieben
E4c	Querschnitt- oder prospektiv geplante Kohortenstudien wie in Abschnitt 4.2.1.4 beschrieben
E5c	Keine Mehrfachpublikationen ohne relevante Zusatzinformation
E6c	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 STARD-Statements [53], STROBE-Statements [41] oder MOOSE-Statements [42] genügt und eine Bewertung der Studie ermöglicht.	

4.3 Informationsbeschaffung

4.3.1 Bibliografische Literaturrecherche

Die systematische Suche nach relevanten Studien 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 relevanten Sekundärpublikationen: 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 (CDSR), Database of Abstracts of Reviews of Effects (DARE) und Health Technology Assessment Database (HTA)

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 Studien herangezogen werden:

- Durch den G-BA übermittelte Unterlagen
- Öffentlich zugängliche Studienregister
- Gegebenenfalls Informationen von Autoren einzelner Publikationen, z. B. zur Frage nach nicht publizierten Teilaspekten
- Gegebenenfalls Informationen von Sachverständigen / Experten / Fachgesellschaften
- Im Rahmen der Anhörung zum vorläufigen Berichtsplan und zum Vorbericht eingereichte Informationen

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

4.3.3 Selektion relevanter Studien

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

Die Fragestellung zur diagnostischen und prognostischen Güte wird primär auf Basis systematischer Übersichten und HTA-Berichte beantwortet. Zunächst wird daher systematisch nach Sekundärliteratur gesucht und die gefundenen Arbeiten mit dem Qualitätsindex von Oxman und Guyatt bewertet (vgl. Abschnitt 4.2.1.5). Anschließend wird für den Zeitraum, der durch den Recherchezeitraum der systematischen Übersichten nicht abgedeckt ist, eine zusätzliche Recherche für Primärliteratur durchgeführt (Ergänzungsrecherche; vgl. Abschnitt 4.2.2). Sollten für die 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

Evidenztabelle. Ein Gutachter führt die Datenextraktion unter Verwendung des Extraktionsbogens durch. Ein zweiter Gutachter überprüft die Extraktion. Etwaige Diskrepanzen in der Bewertung werden durch Diskussion zwischen den Gutachtern, 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 weiteren Informationsquellen ab. Alle für die Nutzenbewertung relevanten, mit validierten Messinstrumenten erhobene Ergebnisse werden hinsichtlich ihrer Ergebnissicherheit, bestehend aus dem Verzerrungspotenzial und der Präzision, 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
- Ergebnigesteuerte Berichterstattung

B: Aspekte des Verzerrungspotenzials der Ergebnisse auf Endpunktebene

- Verblindung der Endpunkterheber
- Umsetzung des ITT-Prinzips
- Ergebnigesteuerte 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 i. d. R. auch als „hoch“ bewertet. Ansonsten finden die unter B genannten endpunktspezifischen 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 bzw. HTA-Berichten 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 Referenztest und den Studientypen aufgeteilt.

Die Bewertung der allgemeinen Qualität einer systematischen Übersicht bzw. eines HTA-Berichts erfolgt mindestens anhand der Kriterien von Oxman und Guyatt [50,51]. Diese Kriterien können bei Bedarf um weitere z. B. zum Umgang mit Verifikations- bzw. Inkorporationsbias, Angabe von Interessenkonflikten, ergänzt werden.

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 Qualitätsbewertung der Primärstudien zur diagnostischen Güte wird mithilfe des QUADAS-Instruments durchgeführt [45]. Dieses kann bei Bedarf an die Fragestellungen angepasst werden (z. B. für Konkordanzstudien). Für die Qualitätsbewertung von prognostischen Studien wird zusätzlich zu QUADAS die adäquate Berücksichtigung von möglichen Confoundern ein weiteres Qualitätskriterium sein [54].

Die Qualität 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 Nutzenbewertung. 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. In jedem Fall erfolgt eine Darstellung der Ergebnisse jeder darin enthaltenen Primärstudie. Zusammenfassende Ergebnisse der systematischen Übersichten werden ebenfalls berichtet.

4.5.1 Gegenüberstellung der Ergebnisse

Die Ergebnisse zu den in den Studien berichteten patientenrelevanten 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 Daten zu vielen Patienten nicht in der Auswertung berücksichtigt sind. Ergebnisse fließen i. d. R. nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patienten basieren, d. h. wenn der Anteil der fehlenden Werte größer als 30 % ist. In der Literatur werden zum Teil bereits Ausfallraten größer als 20 Prozentpunkte als nicht mehr aussagekräftig betrachtet [55]. Die Festlegung der Grenze auf 30 % kann daher als vergleichsweise liberal angesehen werden. Das beschriebene Vorgehen wird auch angewendet, wenn der Unterschied der fehlenden Werte zwischen den Gruppen größer als 15 Prozentpunkte ist.

Ausnahmen von dieser Regel können z. B. dann gemacht werden, wenn aus logistischen Gründen für ganze Zentren (ganze Randomisierungsblöcke) keine Daten erhoben wurden, obwohl dies bereits bei der Studienplanung vorgesehen war [56]. Das beschriebene Vorgehen wird auch angewendet, wenn der Unterschied der fehlenden Werte 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 Krankheitsentitäten und durch prognostische Angaben beschrieben. Zusätzlich soll die Zahl der Studienabbrecher dargestellt werden.

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 Zielgrößen (s. 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 „Referenztest“ 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 [57].

Sollten durch eine Ergänzungsrecherche für den Zeitraum nach Abschluss des Recherchezeitraums der eingeschlossenen systematischen Übersicht noch weitere relevante Primärstudien identifiziert werden, so werden diese separat und analog zu den (Primär-)Studien zur Bewertung des patientenrelevanten Nutzens und Schadens präsentiert (s. 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 zugrundeliegenden Methodik (Studiendesign, Studiendauer, Ort und Zeitraum der Studiendurchführung, relevante Endpunkte) beschrieben.

Die Studienpopulationen werden weiterhin durch demografische Daten (Alter, Geschlecht), durch Charakteristika der betrachteten Krankheitsentitäten und durch prognostische Angaben beschrieben. Zusätzlich soll 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 „Referenztest“ klassifiziert und qualitativ zusammengefasst. Bei prognostischen Primärstudien werden – soweit untersucht – multifaktorielle Analysen und der Einfluss von möglichen Confoundern auf die prognostische Güte präsentiert.

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 Studien – sowohl jene, die in den systematischen Übersichten vorhanden waren, als auch jene aus der Ergänzungsrecherche – in Erwägung gezogen.

4.5.2 Meta-Analysen

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

Die Meta-Analysen erfolgen i. d. R. auf Basis von Modellen mit zufälligen Effekten [58]. In begründeten Ausnahmefällen werden Modelle mit festen Effekten eingesetzt. Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, werden diese nach Möglichkeit anhand der vorhandenen Informationen eigenständig berechnet bzw. näherungsweise bestimmt.

Bei Studien zum patientenrelevanten Nutzen und Schaden wird für kontinuierliche Variablen die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedge's g , als Effektmaß eingesetzt. Bei binären Variablen werden Meta-Analysen primär anhand des Odds Ratio und beim Vergleich von Überlebenszeiten anhand des Hazard Ratio durchgeführt. Bei Diagnostikstudien wird, sofern die dafür nötigen Anforderungen als erfüllt gelten können, für die Testgütekriterien eine Meta-Analyse anhand der Sensitivität und Spezifität in einer

bivariaten Analyse durchgeführt [59]. In begründeten Ausnahmefällen kommen auch andere Effektmaße zum Einsatz.

Die Effektschätzer und Konfidenzintervalle aus den Studien werden mittels Forest Plots zusammenfassend dargestellt. Anschließend erfolgt die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand des Maßes I^2 und des statistischen Tests auf Vorliegen von Heterogenität [60]. Ist die Heterogenität der Studienergebnisse nicht bedeutsam ($p > 0,2$ für Heterogenitätstest), wird der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt. Bei bedeutsamer Heterogenität werden die Ergebnisse nur in begründeten Ausnahmefällen gepoolt. Außerdem wird untersucht, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt 4.5.3) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.5.4).

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“ oder „niedrig“ wird hierfür verwendet. Dies betrifft sowohl Primärstudien zum patientenrelevanten Nutzen und Schaden als auch jene zur diagnostischen und prognostischen Güte. Bei Einschluss von mehreren systematischen Übersichten kann eine Sensitivitätsanalyse anhand der Beurteilung mit dem Index von Oxman und Guyatt [50,51] 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, z. B. die Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder die Wahl des Effektmaßes.

Das Ergebnis der Sensitivitätsanalysen kann die Sicherheit der aus den beobachteten Effekten abgeleiteten Aussagen beeinflussen. Ein als nicht robust eingestufteffekt kann z. B. dazu führen, dass nur ein Hinweis auf anstelle eines Belegs für einen Nutzen oder Schaden attestiert wird.

4.5.4 Subgruppenmerkmale und andere Effektmodifikatoren

Die Ergebnisse werden hinsichtlich potenzieller Effektmodifikatoren, d. h. klinischer Faktoren, die die Effekte beeinflussen, untersucht. Dies können direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Behandlungen sein. Im Gegensatz zu den in Abschnitt 4.5.3 beschriebenen methodischen Faktoren für Sensitivitätsanalysen besteht hier das Ziel, mögliche Effektunterschiede zwischen Patientengruppen und Behandlungsspezifika aufzudecken. Für einen Nachweis unterschiedlicher Effekte ist die auf einem Homogenitäts- bzw. Interaktionstest basierende statistische Signifikanz Voraussetzung. In die Untersuchung von Effektmodifikatoren werden die vorliegenden Ergebnisse aus Regressionsanalysen, die Interaktionsterme beinhalten, und aus Subgruppenanalysen

einbezogen. Außerdem erfolgen eigene Analysen in Form von Meta-Regressionen oder Meta-Analysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren. Es ist vorgesehen, folgende Faktoren bezüglich einer möglichen Effektmodifikation in die Analysen einzubeziehen:

- Geschlecht
- Alter (insbesondere Kinder)
- Tumorentitäten
- Vorangegangene Therapie: Die diagnostische Güte und die Präzision der PET oder PET/CT nach einer Strahlentherapie sollen mit der diagnostischen Güte nach einer Chemotherapie verglichen werden, da es z. B. denkbar ist, dass eine vorherige Strahlentherapie durch eine stärkere Entzündungsreaktion vermehrt zu falsch positiven Befunden führt
- Zeitpunkt der PET-Diagnostik nach Abschluss der Primärtherapie

Sollten in den verfügbaren Informationen Anhaltspunkte für weitere mögliche Effektmodifikatoren vorliegen, können diese ebenfalls begründet einbezogen werden.

Bei Identifizierung möglicher Effektmodifikatoren erfolgt ggf. 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. Gray F, De Girolami U, Poirier J. Escourolle and Poirier's manual of basic neuropathology. Oxford: Butterworth-Heinemann; 2004.
2. Rousseau A, Mokhtari K, Duyckaerts C. The 2007 WHO classification of tumors of the central nervous system: what has changed? *Curr Opin Neurol* 2008; 21(6): 720-727.
3. Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med* 2008; 359(5): 492-507.
4. Schön D, Bertz J, Görsch B, Haberland J, Ziegler H, Stegmaier C et al. Entwicklung der Überlebensraten von Krebspatienten in Deutschland: Gesundheitsberichterstattung für Deutschland; Schwerpunktbericht. Berlin: Robert Koch-Institut; 1999.
5. Stewart LA. Chemotherapy in adult high-grade glioma: a systematic review and meta-analysis of individual patient data from 12 randomised trials. *Lancet* 2002; 359(9311): 1011-1018.
6. Deutsche Gesellschaft für Neurologie. Gliome [Online]. 10.2008 [Zugriff am 18.06.2009]. URL: <http://www.uni-duesseldorf.de/awmf/II/030-099.htm>.
7. Bondy ML, Scheurer ME, Malmer B, Barnholtz-Sloan JS, Davis FG, Il'yasova D et al. Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium. *Cancer Supplement* 2008; 113(7): 1953-1968.
8. Hess KR, Broglio KR, Bondy ML. Adult glioma incidence trends in the United States, 1977-2000. *Cancer* 2004; 101(10): 2293-2299.
9. SNPLogic: interactive SNP selection, annotation and prioritization system [Online]. 2008 [Zugriff am 18.09.2009]. URL: <http://www.snplogic.org>.
10. Schüz J, Böhler E, Berg G, Schlehofer B, Hettinger I, Schläefer K et al. Cellular phones, cordless phones, and the risks of glioma and meningioma (Interphone Study Group, Germany). *Am J Epidemiol* 2006; 163(6): 512-520.
11. Südwestdeutsches Tumorzentrum. Primäre Hirntumoren und ZNS-Metastasen: Empfehlungen zur Diagnostik, Therapie und Nachsorge [Online]. 02.2006 [Zugriff am 30.10.2009]. URL: http://www.stuttgart-region.de/uktmedia/Einrichtungen/Zentren/Tumorzentrum/PDF_Archiv-p-765/Therapieempfehlungen/TLL_Hirntumoren.pdf
12. Davis FG, McCarthy BJ, Freels S, Kupelian V, Bondy ML. The conditional probability of survival of patients with primary malignant brain tumors: surveillance, epidemiology, and end results (SEER) data. *Cancer* 1999; 85(2): 485-491.
13. Wong ET, Hess KR, Gleason MJ, Jaeckle KA, Kyritsis AP, Prados MD et al. Outcomes and prognostic factors in recurrent glioma patients enrolled onto phase II clinical trials. *J Clin Oncol* 1999; 17(8): 2572-2578.

14. Yung WK, Prados MD, Yaya-Tur R, Rosenfeld SS, Brada M, Friedman HS et al. Multicenter phase II trial of temozolomide in patients with anaplastic astrocytoma or anaplastic oligoastrocytoma at first relapse. Temodal Brain Tumor Group. *J Clin Oncol* 1999; 17(9): 2762-2771.
15. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK. WHO classification of tumours of the central nervous system. Genf: WHO Press; 2007. (WHO/ IARC classification of tumours; Band 1).
16. Feiden S, Feiden W. WHO-Klassifikation der ZNS-Tumoren: revidierte Fassung von 2007 mit kritischen Anmerkungen zum "Typing" und "Grading" diffuser Gliome. *Pathologe* 2008; 29(6): 411-421.
17. Chen W. Clinical applications of PET in brain tumors. *J Nucl Med* 2007; 48(9): 1468-1481.
18. Ullrich RT, Kracht LW, Jacobs AH. Neuroimaging in patients with gliomas. *Semin Neurol* 2008; 28(4): 484-494.
19. Vander Borgh T, Asenbaum S, Bartenstein P, Halldin C, Kapucu O, Van Laere K et al. EANM procedure guidelines for brain tumour imaging using labelled amino acid analogues. *Eur J Nucl Med Mol Imaging* 2006; 33(11): 1374-1380.
20. Deutsche Forschungsgemeinschaft. PET-CT [Online]. 2009 [Zugriff am 29.10.2009]. URL: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/wgi/download/pet_ct_stellungnahme_050209.pdf.
21. Bergström M, Collins VP, Ehrin E, Ericson K, Eriksson L, Greitz T et al. Discrepancies in brain tumor extent as shown by computed tomography and positron emission tomography using [68Ga]EDTA, [11C]glucose, and [11C]methionine. *J Comput Assist Tomogr* 1983; 7(6): 1062-1066.
22. American College of Radiology. ACR practice guideline for performing FDG-PET/CT in oncology [Online]. 2007 [Zugriff am 30.01.2009]. URL: http://www.acr.org/EducationCenter/ACRFutureClassroom/fdg_pet_ct.aspx.
23. Krause BJ, Beyer T, Bokisch A, Delbeke D, Kotzerke J, Minkov V et al. Leitlinie FDG-PET/CT in der Onkologie. *Nuklearmedizin* 2007; 46(6): 291-301.
24. Reske SN, Bares R, Bull U, Guhlmann A, Moser E, Wannemacher MF. Klinische Wertigkeit der Positronen-Emissions-Tomographie (PET) bei onkologischen Fragestellungen: Ergebnisse einer interdisziplinären Konsensuskonferenz. *Nuklearmedizin* 1996; 35(2): 42-52.
25. Deutsche Gesellschaft für Nuklearmedizin. Konsensus: Neuro-PET [Online]. 2009 [Zugriff am 28.01.2009]. URL: http://www.nuklearmedizin.de/publikationen/k_papiere_neuro.php.

26. Reske SN, Kotzerke J. FDG-PET for clinical use: results of the 3rd German Interdisciplinary Consensus Conference, "Onko-PET III", 21 July and 19 September 2000. *Eur J Nucl Med* 2001; 11(28): 1707-1723.
27. Schweizerische Gesellschaft für Nuklearmedizin. Klinische Richtlinien der Schweizerischen Gesellschaft für Nuklearmedizin (SGNM) für PET-Untersuchungen [Online]. 07.04.2008 [Zugriff am 30.01.2009]. URL: http://www.nuklearmedizin.ch/download/rechtliches/Klin_PET_Richtlinien_7_4_2008.pdf.
28. Stupp R, Mason WP, Van den Bent MJ, Weller M, Fisher B, Taphoorn MJ et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 2005; 352(10): 987-996.
29. Stupp R, Hegi ME, Mason WP, Van den Bent MJ, Taphoorn MJ, Janzer RC et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* 2009; 10(5): 459-466.
30. Fine HA, Dear KB, Loeffler JS, Black PM, Canellos GP. Meta-analysis of radiation therapy with and without adjuvant chemotherapy for malignant gliomas in adults. *Cancer* 1993; 71(8): 2585-2597.
31. Hou LC, Veeravagu A, Hsu AR, Tse VC. Recurrent glioblastoma multiforme: a review of natural history and management options. *Neurosurg Focus* 2006; 20(4): E5.
32. Suh DY, Olson JJ. Management of recurrent malignant primary brain tumors; part 1: etiology, clinical diagnosis, and treatment options. *Contemp Neurosurg* 1998; 20: 1-6.
33. Barker FG, Chang SM, Gutin PH, Malec MK, McDermott MW, Prados MD et al. Survival and functional status after resection of recurrent glioblastoma multiforme. *Neurosurgery* 1998; 42(4): 709-720.
34. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 3.0 [Online]. 27.05.2008 [Zugriff am 28.01.2009]. URL: http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf.
35. How to read clinical journals; II: to learn about a diagnostic test. *Can Med Assoc J* 1981; 124(6): 703-710.
36. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2): 88-94.
37. Köbberling J, Trampisch HJ, Windeler J. Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem* 1990; 28(12): 873-879.
38. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [Online]. 18.04.2006 [Zugriff am 22.08.2007]. URL: http://www.g-ba.de/downloads/62-492-83/VerfO_2006-04-18.pdf.

39. Lijmer JG, Bossuyt PMM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009; 62(4): 364-373.
40. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001; 134(8): 657-662.
41. Von Elm E., Altman DG, Egger M, Pocock SJ, Gotsche 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.
42. 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.
43. Cheran SK, Herndon JE, Patz EF. Comparison of whole-body FDG-PET to bone scan for detection of bone metastases in patients with a new diagnosis of lung cancer. *Lung Cancer* 2004; 44(3): 317-325.
44. Kirby AM, George MN. The role of FDG PET in the management of lymphoma: practical guidelines. *Nucl Med Commun* 2007; 28(5): 355-357.
45. 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.
46. 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.
47. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004; 140(3): 189-202.
48. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2001; 2(3): 249-260.
49. Jadad AR, Murray WE. *Randomized controlled trials: questions, answers and musings*. Malden: Blackwell; 2007.
50. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44(11): 1271-1278.
51. 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.
52. 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.

53. 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.
54. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; 144(6): 427-437.
55. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
56. 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.
57. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156(10): 1411-1416.
58. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
59. 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.
60. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.