

**Positronenemissions-
tomographie (PET) und
PET/CT zur Lokalisation
epileptogener Zonen für die
chirurgische Behandlung der
Epilepsie**

Vorläufiger Berichtsplan

Auftrag: D06-01M
Version: 1.0
Stand: 12.09.2011

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Positronenemissions-tomographie (PET) und PET/CT zur Lokalisation epileptogener Zonen für die chirurgische Behandlung der Epilepsie

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

19.12.2006

Interne Auftragsnummer:

D06-01M

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Dillenburger Str. 27

51105 Köln

Tel.: +49 221 35685-0

Fax: +49 221 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: Epilepsie; Positronen-Emissions-Tomographie, Computertomographie, systematische Übersicht

Keywords: Epilepsy; Positron-Emission Tomography; Tomography, X-Ray Computed; Systematic Review

Inhaltsverzeichnis

	Seite
Impressum.....	i
Tabellenverzeichnis	vi
Abkürzungsverzeichnis.....	vii
1 Hintergrund.....	1
1.1 Definition des Krankheitsbildes.....	1
1.2 Epidemiologie und individuelle Krankheitslast	1
1.3 Ursache der Erkrankung.....	1
1.4 Klassifikationsschemata	2
1.5 Verlauf der Erkrankung	5
1.6 Derzeit übliche epilepsiechirurgische Verfahren	6
1.7 Prächirurgische diagnostische Verfahren	9
1.8 Empfehlungen wissenschaftlicher Fachgesellschaften zum prächirurgischen Einsatz der PET bei Epilepsie.....	10
2 Ziele der Untersuchung	11
3 Projektbearbeitung.....	13
3.1 Zeitlicher Verlauf des Projekts	13
4 Methoden.....	14
4.1 Ermittlung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT	16
4.1.1 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT anhand von systematischen Übersichten.....	17
4.1.2 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT anhand von Primärstudien.....	17
4.1.2.1 Populationen	17
4.1.2.2 Prüf- und Vergleichsinterventionen.....	17
4.1.2.3 Patientenrelevante Endpunkte.....	18
4.1.2.4 Studientypen	18
4.1.2.5 Studiendauer	19
4.1.2.6 Sonstige Studiencharakteristika.....	19
4.1.3 Tabellarische Darstellung der Einschlusskriterien („Nutzenstudien“).....	19
4.2 Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT	20
4.2.1 Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Güte anhand von systematischen Übersichten	21

4.2.1.1	Populationen	21
4.2.1.2	Prüftechnologien, Vergleichstechnologien („Komparatoren“) und Referenztests	21
4.2.1.3	Zielgrößen	22
4.2.1.4	Den systematischen Übersichten zugrunde liegende Studientypen.....	22
4.2.1.5	Sonstige Charakteristika der systematischen Übersichten.....	23
4.2.1.6	Tabellarische Darstellung der Einschlusskriterien (systematische Übersichten zur diagnostischen und prognostischen Güte)	23
4.2.2	Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Güte anhand von Primärstudien („Ergänzungsrecherche“).....	24
4.2.2.1	Populationen	24
4.2.2.2	Prüftechnologien (Indextest I), Vergleichstechnologien (Indextest II) und Referenztests	24
4.2.2.3	Zielgrößen	24
4.2.2.4	Studientypen	25
4.2.2.5	Sonstige Studiencharakteristika.....	25
4.2.2.6	Tabellarische Darstellung der Einschlusskriterien („Ergänzungsrecherche“ nach Primärstudien zur diagnostischen und prognostischen Güte)	25
4.2.3	Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen.....	25
4.3	Informationsbeschaffung.....	26
4.3.1	Bibliografische Literaturrecherche	26
4.3.2	Suche nach weiteren publizierten und nicht publizierten Studien und systematischen Übersichten.....	26
4.3.3	Selektion relevanter Studien und systematischer Übersichten	26
4.4	Informationsbewertung	27
4.4.1	Bewertung von Studien zum patientenrelevanten Nutzen und Schaden	27
4.4.2	Bewertung von systematischen Übersichten zur diagnostischen und prognostischen Güte	28
4.4.3	Bewertung von Primärstudien zur diagnostischen und prognostischen Güte	28
4.5	Informationssynthese und -analyse	29
4.5.1	Gegenüberstellung der Ergebnisse der Einzelstudien	30
4.5.1.1	Gegenüberstellung der Ergebnisse von (Primär-)Studien zur Bewertung des patientenrelevanten Nutzens und Schadens.....	30
4.5.1.2	Gegenüberstellung der Ergebnisse von systematischen Übersichten zur diagnostischen und prognostischen Güte.....	31
4.5.1.3	Gegenüberstellung der Ergebnisse von Primärstudien zur diagnostischen und prognostischen Güte („Ergänzungsrecherche“).....	31

4.5.1.4	Gegenüberstellung der Ergebnisse aus systematischen Übersichten und der Ergänzungsrecherche („Robustheitsprüfung“)	32
4.5.2	Meta-Analysen	32
4.5.3	Sensitivitätsanalyse	33
4.5.4	Subgruppenmerkmale und andere Effektmodifikatoren	34
5	Literaturverzeichnis	35

Tabellenverzeichnis

	Seite
Tabelle 1: Geschätzter Anteil mutmaßlicher Ursachen von Epilepsien in bevölkerungsbezogenen Inzidenzstudien	2
Tabelle 2: Klassifikation von epileptischen Anfällen	3
Tabelle 3: Elektroklinische Syndrome und andere Epilepsien.....	4
Tabelle 4: Langzeit-Anfallsfreiheit: Gesamtanteil nach chirurgischer Maßnahme	6
Tabelle 5: Engel-Klassifikation der postoperativen Ergebnisse.....	8
Tabelle 6: ILAE-Vorschlag einer neuen Outcome-Klassifikation hinsichtlich epileptischer Anfälle	9
Tabelle 7: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren.....	15
Tabelle 8: Evidenzklassifizierung des G-BA für Unterlagen zu diagnostischen Methoden....	16
Tabelle 9: Übersicht der Einschlusskriterien für Primärstudien zur Nutzenbewertung der PET bzw. PET/CT.....	20
Tabelle 10: Übersicht der Einschlusskriterien für systematische Übersichten zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT	24
Tabelle 11: Übersicht der Einschlusskriterien für die Primärliteratur zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT (Ergänzungsrecherche).....	25

Abkürzungsverzeichnis

Abkürzung	Bedeutung
CCT	controlled clinical trial (kontrollierte klinische Studie)
CT	Computertomographie
DGfE	Deutsche Gesellschaft für Epileptologie e. V.
DGN	Deutsche Gesellschaft für Nuklearmedizin e. V.
ECoG	Elektrokortikogramm
EEG	Elektroenzephalografie
EFNS	European Federation of Neurological Societies (Europäische Vereinigung Neurologischer Gesellschaften)
fMRT	funktionelle Magnetresonanztomographie
G-BA	Gemeinsamer Bundesausschuss
HSROC	hierarchical summary receiver operating curve (Methode zur Bewertung und Optimierung von Analyse-Strategien)
HTA	Health Technology Assessment
IBE	International Bureau for Epilepsy (Internationales Büro für Epilepsie)
ILAE	International League Against Epilepsy (Internationale Liga gegen Epilepsie)
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
LR	Likelihood Ratio
MID	Minimal Important Difference
MRT	Magnetresonanztomographie
PET	Positronenemissionstomographie
PET/CT	Integration von PET und CT in einem Untersuchungsgerät
PICO	Patient, Intervention, Comparison (Vergleichsintervention), Outcome (patientenrelevanter Endpunkt)
SF-36	Krankheitsübergreifendes Messinstrument zur Erfassung der gesundheitsbezogenen Lebensqualität von Patienten
SMD	standardized mean differences (Standardisierte Mittelwertdifferenzen)
SMR	standardized mortality ratio (Standardisiertes Sterblichkeitsverhältnis)

Abkürzung	Bedeutung
SN	Sensitivität
SP	Spezifität
RCT	randomized controlled trial (randomisierte kontrollierte Studie)
ROC	Receiver Operating Characteristic (Methode zur Bewertung und Optimierung von Analyse-Strategien)
SPECT	Single Photon Emission Computed Tomography (Einzelphotonenemissions-Computertomographie, Schnittbildverfahren der Nuklearmedizin)
SUDEP	Sudden unexpected death in epilepsy (plötzlich auftretender, ungeklärter Tod bei Epilepsie)
TLE	Temporallappenepilepsie
TLR	Temporallappenresektion
VOPT	Verification of Only Positive Testers
ZNS	zentrales Nervensystem

1 Hintergrund

1.1 Definition des Krankheitsbildes

Die Internationale Liga gegen Epilepsie (ILAE) und das Internationale Büro für Epilepsie (IBE) bezeichnen als *epileptischen Anfall* ein vorübergehendes Auftreten von Anzeichen und / oder Symptomen aufgrund von abnormalen exzessiven oder synchronen neurologischen Aktivitäten im Gehirn [1].

Epilepsie ist der ILAE und dem IBE zufolge eine Erkrankung des Gehirns, die durch eine andauernde Prädisposition zur Generierung epileptischer Anfälle charakterisiert ist. Ebenso ist sie durch neurobiologische, kognitive, psychologische und soziale Konsequenzen dieses Zustands gekennzeichnet. Die Definition von Epilepsie erfordert darüber hinaus das Auftreten von mindestens einem epileptischen Anfall.

Der Zeitraum während eines epileptischen Anfalls wird als iktal bezeichnet. In der Phase zwischen zwei Anfällen wird von einem interiktalen Zustand gesprochen [2].

1.2 Epidemiologie und individuelle Krankheitslast

Mit einer geschätzten Prävalenz von 0,6 % [3] zählen Epilepsien zu den häufigsten schwerwiegenden neurologischen Erkrankungen. Dabei unterscheidet sich die Auftretenshäufigkeit zwischen den verschiedenen Altersstufen. Für Kinder und Jugendliche in Europa schätzen Forsgren et al. [3] die Prävalenz einer aktiven Epilepsie auf 0,45 bis 0,5 %. In der Altersgruppe der 20- bis 64-Jährigen liegt der Prozentsatz bei 0,6 % und bei den über 65-Jährigen bei 0,7 %.

Die Inzidenz wird in Europa auf 50 bis 55 Personen pro 100 000 Einwohner pro Jahr geschätzt. In einer altersabhängigen Betrachtung weist dieses Krankheitsbild einen zweigipfeligen Verlauf auf: Die Anzahl neuer Fälle pro Jahr wird bei Kindern und Jugendlichen auf 70/100 000, bei 20- bis 64-Jährigen auf 30/100 000 und bei über 65-Jährigen auf 100/100 000 geschätzt [3].

1.3 Ursache der Erkrankung

Epilepsien scheinen eine Vielzahl von Ursachen zugrunde zu liegen. Unterteilt werden sie nach der revidierten Terminologie der ILAE von 2010 in *genetische, strukturell-metabolische* und *unbekannte* Ursachen [4,5]. Diese Begriffe ersetzen die veralteten Konzepte der idiopathischen, symptomatischen und kryptogenen Epilepsie. Die neuen Konzepte verlangen den definitiven Nachweis des genetischen Einflusses auf die jeweilige Epilepsieform bzw. des deutlich erhöhten Risikos struktureller und metabolischer Zustände oder Krankheiten, eine Epilepsie zu entwickeln, anhand von angemessenen Studien.

Im Jahr 2005 haben Forsgren et al. [3] eine systematische Übersicht publiziert und darin bevölkerungsbezogene Inzidenzstudien zusammengetragen, die sich den Ursachen von

Epilepsien gewidmet haben. Die in diesen Studien ermittelten Faktoren mit mutmaßlich ätiologischer Bedeutung sind in Tabelle 1 dargestellt. Mit Ausnahme der Ergebnisse aus der estnischen Studie sind die Ursachen der Epilepsien bei der Mehrzahl der Patienten unbekannt. Zu den häufigsten bekannten Ursachen zählen zerebrovaskuläre Erkrankungen, insbesondere ischämische Infarkte. Traumata, Hirntumoren, Infektionen des zentralen Nervensystems (ZNS) und kongenitale Veränderungen sind andere bekannte Ursachen. Auch bei Personen mit degenerativen Erkrankungen treten Epilepsien häufiger auf – insbesondere bei Personen mit Alzheimer oder vaskulärer Demenz [3].

Tabelle 1: Geschätzter Anteil mutmaßlicher Ursachen von Epilepsien in bevölkerungsbezogenen Inzidenzstudien (in %) [3]

	Land				Range
	Island [6]	Schweden [7,8]	Vereinigtes Königreich [9]	Estland [10]	
Gering oder progredient symptomatisch	31	46	39	56	31–56
Vaskulär	14	21	15	20	14–21
Ischämie		18		16	16–18
Hämorrhagie		3		4	3–4
Trauma	0	2	3	16	2–16
Hirntumor	7	7	6	10	6–10
Infektion des zentralen Nervensystems	2	0	2	1	0–2
Degenerative Erkrankung	2	5		1	1–5
Kongenitale Veränderung	5	7		4	4–7
Andere	0	4	13	4	0–13
Unbekannt	69	54	61	44	44–69

1.4 Klassifikationsschemata

Epileptische Anfälle werden in zwei Klassen unterteilt – in die Klassen der generalisierten und fokalen Anfälle (siehe auch Tabelle 2). Unter generalisierten Anfällen werden der ILAE zufolge Anfälle subsumiert, die „in einem bilateral verteilten Netzwerk auftreten und sich dort rasch ausbreiten“ [5, S. 101]. Fokale Anfälle hingegen treten in einem auf eine Großhirnhemisphäre beschränkten neuronalen Netzwerk auf und können dabei sowohl eng umschrieben als auch weiter ausgedehnt sein. Da eine natürliche Klassifikation letzterer Anfälle fehlt, sollten diese anhand ihrer Manifestationen dargestellt werden (z. B. dyskognitiv

oder fokal-motorisch) [5]. Darüber hinaus gibt es Anfälle, die keiner dieser Kategorien zugeordnet werden können.

Tabelle 2: Klassifikation von epileptischen Anfällen [4,5]

Generalisierte Anfälle^a

Tonisch-klonisch (in jeder Kombination)

Absence

Typisch

Atypisch

Absence mit speziellen Merkmalen

Myoklonische Absence

Lidmyoklonien mit Absence

Myoklonisch

Myoklonisch

Myoklonisch-atonisch

Myoklonisch-tonisch

Klonisch

Tonisch

Atonisch

Fokale Anfälle

Unbekannt

Epileptische Spasmen

a: Ein Anfall, der nicht ohne Weiteres in eine der vorgestellten Kategorien eingeordnet werden kann, sollte als „nicht klassifiziert“ betrachtet werden, bis weitere Informationen seine genaue Diagnose erlauben. Dies wird jedoch nicht als eine Klassifikationskategorie aufgefasst.

Epilepsien können vor dem Hintergrund des aktuellen Forschungsstandes und der darauf basierenden revidierten Terminologie der ILAE von 2010 in mindestens drei oder vier Gruppierungen unterteilt werden [4,5] (siehe auch Tabelle 3). In der ersten Gruppe werden klinische Entitäten zusammengefasst, die zuverlässig „durch ein Cluster elektroklinischer Merkmale identifiziert werden können“ [5, S. 107] und ausgeprägte entwicklungsbezogene oder genetische Komponenten aufweisen. Eine weitere Gruppe von Krankheitsentitäten wird unter dem Begriff der unverwechselbaren Konstellationen subsumiert. Ihnen liegen spezifische Läsionen oder andere eindeutige Ursachen zugrunde. Sie stellen diagnostisch bedeutsame Formen dar und haben Implikationen auf die Therapie und im Speziellen auf epilepsiechirurgische Interventionen. In der nächsten Gruppe werden Epilepsien aufgrund von spezifischen strukturellen oder metabolischen Läsionen oder Zuständen zusammengefasst. Epilepsien, deren Ursachen bislang unbekannt sind und in veralteten Konzepten als kryptogen charakterisiert wurden, werden als Epilepsien unbekannter Ursache bezeichnet. Mindestens ein Drittel aller Epilepsieformen fallen in diese Kategorie. Die Einteilung der Epilepsien basiert auf dem Forschungsstand von 2009 und kann sich in Zukunft ändern [4,5].

Tabelle 3: Elektroklinische Syndrome und andere Epilepsien [4,5]

<p>Elektroklinische Syndrome – nach Manifestationsalter geordnet^a</p> <p>Neugeborenenzeit</p> <ul style="list-style-type: none"> Benigne familiäre neonatale Epilepsie (BFNE) Frühe myoklonische Enzephalopathie (FME) Ohtahara-Syndrom (OS) <p>Kleinkindalter</p> <ul style="list-style-type: none"> Epilepsie der frühen Kindheit mit migratorischen fokalen Anfällen West-Syndrom (WS) Myoklonische Epilepsie der frühen Kindheit (MEI) Benigne frühkindliche Epilepsie (BFE) Benigne familiäre frühkindliche Epilepsie (BFFE) Dravet-Syndrom (DS) Myoklonische Enzephalopathie bei nicht-progredienten Störungen <p>Kindheit</p> <ul style="list-style-type: none"> Fiebergebundene Anfälle plus (FA+; „Fieberkrämpfe“ plus; können in der frühen Kindheit bzw. im Kleinkindalter beginnen) Panayiotopoulos-Syndrom Epilepsie mit myoklonisch-atonischen (früher astatischen) Anfällen Benigne Epilepsie mit zentrotemporalen Spikes (BEZTS; Rolando-Epilepsie) Autosomal-dominante nächtliche Frontallappenepilepsie (ADNFLE) Spät beginnende kindliche Okzipitallappenepilepsie (Gastaut-Typ) Epilepsie mit myoklonischen Absencen Lennox-Gastaut-Syndrom (LGS) Epileptische Enzephalopathie mit kontinuierlichen Spike-und-wave-Entladungen im Schlaf (CSWS)^b Landau-Kleffner-Syndrom (LKS) Kindliche Absencenepilepsie (KAE) <p>Adoleszenz – Erwachsenenalter</p> <ul style="list-style-type: none"> Juvenile Absencenepilepsie (JAE) Juvenile myoklonische Epilepsie (JME) Epilepsie mit nur generalisierten tonisch-klonischen Anfällen Progressive Myoklonusepilepsien (PME) Autosomal-dominante fokale Epilepsie mit akustischen Merkmalen (ADFEAM) Andere familiäre Temporallappenepilepsien 	<p>Weniger spezifische Altersbeziehung</p> <ul style="list-style-type: none"> Familiäre fokale Epilepsie mit variablen Herden (Kindheit bis Erwachsenenalter) Reflexepilepsien <p>Unverwechselbare Konstellationen</p> <ul style="list-style-type: none"> Mesiale Temporallappenepilepsie mit Hippokampussklerose (MTLE mit HS) Rasmussen-Syndrom Gelastische Anfälle bei hypothalamischen Hamartomen Hemikonvulsions-Hemiplegie-Epilepsie(-Syndrom) <p>Epilepsien, die <i>nicht</i> in diese diagnostischen Kategorien passen, können zunächst auf der Basis des Vorhandenseins oder Fehlens einer bekannten strukturellen oder metabolischen Störung (vermutliche Ursache) und dann auf der Basis des primären Anfallsbeginns (generalisiert versus fokal) unterschieden werden.</p> <p>Epilepsien aufgrund von und eingeteilt nach strukturell-metabolischen Ursachen</p> <ul style="list-style-type: none"> Malformationen der kortikalen Entwicklung (Hemimegalenzephalie, Heterotopien etc.) Neurokutane Syndrome (Tuberöse-Sklerose-Komplex, Sturge-Weber-Syndrom etc.) Tumoren Infektionen Traumen <p>Angiome</p> <ul style="list-style-type: none"> Perinatale Insulte Schlaganfälle etc. <p>Epilepsien unbekannter Ursache</p> <p>Zustände mit epileptischen Anfällen, die traditionell nicht als eine Epilepsieform per se betrachtet werden</p> <ul style="list-style-type: none"> Benigne neonatale Anfälle (BNA) Fiebergebundene Anfälle (FA, „Fieberkrämpfe“) <p>a: Diese Einordnung erfolgt ohne Bezug zur Ätiologie.</p> <p>b: Diese Epilepsieform wird manchmal auch als Epilepsie mit Status Epileptici im Schlaf (ESES) bezeichnet.</p>
---	--

1.5 Verlauf der Erkrankung

Der natürliche Verlauf von Epilepsien ist weitgehend unbekannt, da seit über 100 Jahren effektive Behandlungsformen verfügbar sind [11]. Sogenannte „Indizienbeweise“, die vornehmlich aus ressourcenarmen Ländern stammen, in denen Antiepileptika nicht leicht verfügbar sind, deuten Kwan und Sander [11] zufolge darauf hin, dass Spontanremissionen in bis zu 30 % der Fälle auftreten. Beobachtungen aus diesen ergänzenden Quellen legen nahe, die Prognose neu diagnostizierter Epilepsien auf Bevölkerungsebene allgemein in drei Gruppen zu kategorisieren [11]: Die erste Gruppe hat eine *exzellente Prognose* – 20–30 % der Patienten können dieser Gruppe zugeordnet werden. Sie werden – vermutlich ohne antiepileptische Behandlung – eine Langzeitremission erreichen. Die ILAE fasst die dieser Gruppe von Patienten zugrunde liegenden Epilepsieformen als „selbst limitierend“ zusammen [4,5]. Der zweiten Gruppe sind 20–30 % der Patienten zuzuordnen. Bei ihnen tritt eine *Remission unter kontinuierlicher Einnahme antiepileptischer Medikamente* ein. Die ILAE bezeichnet die in dieser Gruppe auftretenden Syndrome als „pharmakoresponsiv“. Die dritte Gruppe besteht aus Patienten, die trotz medikamentöser Behandlung Anfälle erleiden. 30–40 % sind von dieser *medikamentösen Therapieresistenz* betroffen. Generell hängt die Prognose stark von der Klassifikation der Epilepsie bzw. des Epilepsiesyndroms sowie der zugrunde liegenden Ätiologie ab ([12-14] zitiert nach [11]).

Für ca. die Hälfte der Patienten mit medikamentöser Therapieresistenz stellt ein epilepsiechirurgischer Eingriff eine Behandlungsoption dar [15]. Ebenso sollte bei Patienten mit Läsionen ein epilepsiechirurgischer Eingriff in Erwägung gezogen werden [2]. Bei Temporallappeneingriffen wird im Median in 66 % der Fälle Langzeit-Anfallsfreiheit erreicht. Bei anderen epilepsiechirurgischen Konstellationen ist die mediane Anfallsfreiheitsrate geringer [16] (siehe auch Tabelle 4), aber immer noch günstiger als bei medikamentöser Therapie [17]. Der vorteilhafte Einfluss langfristiger Anfallsfreiheit auf eine günstigere gesundheitsbezogene Lebensqualität ist entscheidend höher als der einer bloßen Anfallsminderung [18].

Tabelle 4: Langzeit-Anfallsfreiheit: Gesamtanteil nach chirurgischer Maßnahme [16]

Chirurgische Maßnahme (Anzahl der Studien)	Anzahl der Patienten	% anfallsfrei (gepoolt bei ≥ 2 Studien)	95 %-KI
Temporal (n = 40)	3895	66	62, 70
Hemisphärektomie (n = 2)	169	61	54, 68
Temporal und extratemporal gruppiert (n = 25)	2334	59	56, 62
Parietal (n = 1)	82	46	35, 57
Okzipital (n = 1)	35	46	29, 63
Kallosotomie ^a (n = 3)	99	35	26, 44
Extratemporal gruppiert (n = 2)	169	34	28, 40
Frontal (n = 7)	486	27	23, 30
Multiple subpiale Transektion (n = 2)	74	16	8, 24

a: Freiheit bezogen auf die stark beeinträchtigenden Anfallstypen wie z. B. Drop attacks

Die Mortalität von Personen mit Epilepsie ist im Vergleich zur Allgemeinbevölkerung um das bis zu Dreifache erhöht. In bevölkerungsbezogenen Studien zu Epilepsie liegt der Bereich des standardisierten Sterblichkeitsverhältnisses (standardized mortality ratio [SMR]) zwischen 1,6 und 3,0 [19]. Die relative Überlebensrate 5 Jahre nach Diagnose einer Epilepsie liegt bei 91 %, nach 10 Jahren bei 85 % und nach 15 Jahren bei 83 %. In Populationen, in denen die Epilepsie erstmals im Kindesalter aufgetreten ist, liegt die relative Überlebensrate nach 10 Jahren bei 94 % und nach 20 Jahren bei 88 %. Die Höhe des Mortalitätsrisikos wird dabei durch mehrere Faktoren beeinflusst, z. B. durch die Epilepsieform, das Alter oder die Grunderkrankung. Die häufigste direkte anfallsbedingte Todesursache im Jugend- und jungen Erwachsenenalter ist der „sudden unexpected death in epilepsy“ (SUDEP). Dieser tritt in dieser Gruppe 24-mal häufiger auf als in der Allgemeinbevölkerung [19].

1.6 Derzeit übliche epilepsiechirurgische Verfahren

Unter dem Begriff Epilepsiechirurgie werden alle neurochirurgischen Interventionen subsumiert, die eine Linderung bei medikamentenresistenten Epilepsien zum primären Ziel haben [20,21]. Unterteilt werden die verschiedenen Interventionsmaßnahmen in resektive, d. h. strukturentfernende, und nichtresektive Verfahren.

Das Ziel eines *resektiven* Einsatzes ist in der Regel das Erreichen von Anfallsfreiheit bei gleichzeitiger Minimierung von Nebeneffekten und Verbesserung der Lebensqualität [20]. Um dies anhand einer Entfernung des epileptogenen Herdes erreichen zu können, ist es notwendig, diesen Herd so präzise wie möglich zu identifizieren. Eine direkte Erfassung ist derzeit mit keinem diagnostischen Verfahren möglich. Es handelt sich beim epileptogenen Herd somit um ein theoretisches Konstrukt [22]. Ist eine präzise Identifikation nicht möglich, können alternative chirurgische Maßnahmen zum Einsatz kommen. Den häufigsten resektiven Eingriff stellt die Temporallappenresektion (TLR) dar, die je nach Lokalisation der

verschiedenen anatomischen Komponenten weiter unterteilt wird (z. B. vordere Temporallappenresektion, anteromediale temporale Resektion oder selektive Amygdala-Hippokampektomie) [23]. Resektive Eingriffe bei Epilepsien außerhalb des Temporallappens werden als extratemporale Resektionen bezeichnet. Auch hier erfolgt eine Unterteilung je nach Lage des Anfallsherds, Ausbreitungsvolumen etc. (z. B. Lobektomien, Teillappenresektionen, Kortikektomien, Mehrlappenresektionen) [2]. Ferner gibt es den resektiven Eingriff der Hemisphärektomie, bei dem „der Einfluss einer insgesamt als epileptogen begriffenen Großhirnhemisphäre auf die Gegenseite und auf tiefere zerebrospinale Strukturen ausgeschaltet“ wird [23, S. 409]. Als Läsionektomie werden chirurgische Eingriffe bezeichnet, bei denen umschriebene intrazerebrale Läsionen (und umgebender Kortex) reseziert werden, die als epileptogen nachgewiesen wurden.

Zu den *nichtresektiven* chirurgischen Verfahren zählen die Kallosotomie, multiple subpiale Transsektionen und die Vagus-Nerv-Stimulation [21,23]. Alle Verfahren gelten als palliativ, da mit ihnen nur selten Anfallsfreiheit erreicht wird. Mit ihrem Einsatz wird vielmehr eine Reduktion der Anfallsfrequenz oder eine Verbesserung der Anfallskontrolle verfolgt. Sie kommen erst in Betracht, wenn die Möglichkeit eines potenziell kurativen resektiven Eingriffs nicht gegeben ist. Während Kallosotomie und multiple subpiale Transsektion diskonnektive Verfahren darstellen, bei denen das Ziel verfolgt wird, die Ausbreitung epileptischer Aktivitäten über bestimmte Bahnsysteme zu unterbrechen, ist die Vagus-Nerv-Stimulation ein Stimulationsverfahren. Durch dieses soll mittels Stimulation afferenter oder intrazerebraler Strukturen die Anfallsfrequenz beeinflusst werden [23].

Für die Beurteilung der postoperativen Anfallskontrolle wird in Studien mehrheitlich die sogenannte „Engel-Klassifikation“ herangezogen [24], die in Tabelle 5 dargestellt ist. Unter anderem wegen der Mehrdeutigkeit der Klasse III der Engel-Klassifikation („lohnenswerte Verbesserung“) wurde von der ILAE im Jahr 2001 ein Vorschlag für eine neue Klassifikation der postoperativen Ergebnisse gemacht [25]. In Tabelle 6 ist diese Klassifikation dargestellt.

Tabelle 5: Engel-Klassifikation der postoperativen Ergebnisse [24]

Klasse I	Frei von behindernden Anfällen^a
A.	Komplett anfallsfrei seit der Operation
B.	Ausschließlich nicht behindernde, einfach partielle Anfälle seit der Operation
C.	Einige behindernde postoperative Anfälle, danach mindestens 2 Jahre keine behindernden Anfälle
D.	Ausschließlich generalisierte Anfälle bei Absetzen der antiepileptischen Therapie
Klasse II	Seltene behindernde Anfälle („nahezu anfallsfrei“)
A.	Anfänglich anfallsfrei, danach seltene Anfälle
B.	Seltene behindernde Anfälle seit der Operation
C.	Mehr als seltene behindernde Anfälle postoperativ, aber seltene Anfälle in den letzten 2 Jahren
D.	Ausschließlich nächtliche Anfälle
Klasse III	Lohnenswerte Verbesserung^b
A.	Lohnenswerte Anfallsreduktion
B.	Verlängerte anfallsfreie Intervalle, zusammen mehr als die Hälfte der Follow-up-Periode, jedoch nicht unter 2 Jahren
Klasse IV	Keine lohnenswerte Verbesserung^b
A.	Signifikante Anfallsreduktion
B.	Keine wesentliche Veränderung
C.	Verschlechterung der Anfälle

a: Frühe postoperative Anfälle ausgenommen (die ersten Wochen)

b: Die Ermittlung einer „lohnenswerten Verbesserung“ wird quantitative Analysen von zusätzlichen Daten erfordern, wie prozentuale Anfallsreduktionen, kognitive Funktionen und Lebensqualität

Tabelle 6: ILAE-Vorschlag einer neuen Outcome-Klassifikation hinsichtlich epileptischer Anfälle [25]

Outcome Klassifikation	Definition
1	Völlig anfallsfrei; keine Auren ^a
2	Ausschließlich Auren; keine anderen Anfälle
3	1 bis 3 Anfallstage pro Jahr; ± Auren
4	Mindestens 4 Anfallstage pro Jahr, bis zu 50 %ige Reduktion der präoperativen („Baseline“) Anfallstage; ± Auren
5	Weniger als 50 %ige Reduktion der präoperativen („Baseline“) Anfallstage bis zu 100 %-iger Zunahme der präoperativen („Baseline“) Anfallstage; ± Auren
6	Mehr als 100 %ige Zunahme der präoperativen („Baseline“) Anfallstage; ± Auren

a: Eine Aura ist ein fokaler Anfall, der von Zeugen nicht beobachtbar oder beurteilbar ist. Es handelt sich um ein subjektives Erleben, das die Funktion des Patienten nicht beeinflusst. (Hierzu zählen beispielsweise Schwindel, Kribbeln, Taubheit, visuelle Halluzinationen, akustische Phänomene, Geruchs- und Geschmackswahrnehmungen [26].)

1.7 Prächirurgische diagnostische Verfahren

Dem epilepsiechirurgischen Eingriff voraus gehen zahlreiche diagnostische Untersuchungen, deren primäres Ziel es ist, geeignete Patienten für resektive Eingriffe zu identifizieren [20]. Die ILAE und die European Federation of Neurological Societies (EFNS) zählen folgende Aspekte auf, die im Vorfeld einer Operation mittels verschiedener diagnostischer Prozeduren abgeklärt werden sollten [20,21], deren insgesamtes Ziel darin besteht, einerseits die epileptogene Region zu lokalisieren und andererseits zu ermitteln, welche spezifischen Risiken mit deren operativer Entfernung einhergehen:

- 1) Anamnese inklusive Erfassen interiktaler/iktaler Symptome und Anzeichen, die nachfolgende Untersuchungen beeinflussen und durch sie abgeklärt werden können
- 2) Typ, anatomische Lokalisation und Ausmaß der morphologischen/strukturellen Läsion. Hochauflösende Bildgebung ist hierfür unverzichtbar. Das bevorzugte Verfahren stellt die Magnetresonanztomographie (MRT) dar.
- 3) Dokumentation jeglicher funktioneller Defizite. Eine oder mehrere der folgenden Techniken können hierfür verwendet werden: spezialisierte neuropsychologische Testung, iktale Einzelphotonenemissions-Computertomographie (SPECT), PET und funktionelle MRT (fMRT).
- 4) Bei einigen Patienten ist es nützlich, durch vorübergehende lokale Narkose Hirnareale medikamentös zu inaktivieren (Wada-Test).
- 5) Elektrophysiologische Lokalisation der epileptogenen Zone – einschließlich Aufzeichnung typischer iktaler Zustände und, wenn nötig, Nutzung von intrakranieller Elektroenzephalografie (EEG).

- 6) In den meisten Fällen ist ein Video-EEG mit Oberflächenelektroden notwendig, um die Anfallsursprungszone zu lokalisieren. Es soll eine ausreichende Anzahl an Anfällen aufgezeichnet werden, die mit dem üblichen Anfallsmuster des Patienten übereinstimmen.
- 7) Zusätzliche neurophysiologische Techniken, die vor oder während der Operation zum Einsatz kommen: evozierte Potenziale, funktionelle Kartierung von Kortexabschnitten durch elektrische Reizung (cerebral function mapping), Magnetstimulation und Elektrokortikogramm (ECoG).
- 8) Sorgfältige Untersuchung der psychiatrischen Verfassung unter Verwendung anerkannter psychiatrischer Beurteilungsskalen.

Die Empfehlungen der ILAE für eine prächirurgische Evaluation bei pädiatrischen Eingriffen umspannen dieselben diagnostischen Komponenten [27].

1.8 Empfehlungen wissenschaftlicher Fachgesellschaften zum prächirurgischen Einsatz der PET bei Epilepsie

Die Deutsche Gesellschaft für Epileptologie e. V. (DGfE) und die AG für prächirurgische Diagnostik und operative Epilepsitherapie haben im Jahr 2010 eine Definition von Epilepsiezentren herausgegeben. In dieser wird für Zentren mit Spezialisierung auf Epilepsiechirurgie als ein Element der apparativen Ausstattung die PET vorausgesetzt [28].

Auch die ILAE und die EFNS erwähnen in ihren Empfehlungen für prächirurgische Abklärungen, wie bereits im vorangehenden Abschnitt dargestellt, die PET [20,21,27]. Allerdings wird sie nicht alleine angeführt, sondern als eine optionale Technik funktioneller Bildgebung.

Die Deutsche Gesellschaft für Nuklearmedizin e. V. (DGN) bewertet den Einsatz der PET je nach Epilepsieform bzw. -syndrom unterschiedlich [29]. Zur Lokalisation des epileptogenen Fokus im Rahmen der prächirurgischen Diagnostik bei Temporallappenepilepsien sieht die DGN den Einsatz der PET (mit F-18-FDG) als angemessen und den klinischen Nutzen als erwiesen an. Bei den folgenden Formen betrachtet die DGN den Einsatz als akzeptabel, da Ergebnisse auf den klinischen Wert hindeuten: (1) Lokalisation des epileptogenen Fokus im Rahmen der prächirurgischen Diagnostik bei extratemporalen Epilepsien (mit F-18-FDG); (2) Lokalisation temporaler und extratemporaler Foci im Rahmen der prächirurgischen Diagnostik (mit C-11-Flumazenil); (3) Diagnostik kortikaler Dysplasien und Hamartome (mit F-18-FDG und Rezeptorliganden). Als „möglicherweise hilfreich“ stuft die DGN den Einsatz der PET beim Lennox-Gastaut-Syndrom, beim Sturge-Weber-Syndrom und beim Landau-Kleffner-Syndrom (jeweils mit F-18-FDG) ein. Der Nutzen bei diesen Syndromen ist der DGN zufolge noch nicht ausreichend belegt.

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 Lokalisation epileptogener Zonen für die chirurgische Behandlung der Epilepsie 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 der PET bzw. PET/CT die prognostische bzw. diagnostische Güte der epileptologischen prächirurgischen Diagnosestrategie verbessert.

Alle Fragestellungen werden auch im Rahmen der deutschen Strahlenschutzgesetzgebung erörtert.

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 Anamnese, (Video-)EEG und MRT verstanden. 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.

Allgemeine Voraussetzungen für Studien zur Nutzenbewertung („Nutzenstudien“)

Unter „patientenrelevanter 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 [30].

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 Identifikation der epileptogenen Region verstanden, unter „prognostischer Güte“ eine korrekte Vorhersage der patientenrelevanten Endpunkte (hier: postoperative Anfallsfreiheit).

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

Im Folgenden wird die laut Konkretisierung des G-BA festgelegte Indikation 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 Indikation untersucht werden:

- Lokalisation der epileptogenen Zonen im Rahmen der prächirurgischen Epilepsiediagnostik

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 Lokalisation epileptogener Zonen für die chirurgische Behandlung der Epilepsie.

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

Während der Erstellung des Berichtsplans wurden am 13.05.2011 Patientenvertreter der Deutschen Epilepsievereinigung e. V. zur Festlegung patientenrelevanter Zielgrößen konsultiert.

Der vorliegende vorläufige Berichtsplan (Version 1.0) wird zur Anhörung gestellt. Hierzu können schriftlich 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 der Website des IQWiG 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. seiner Fähigkeit, kranke von gesunden Personen zu unterscheiden [31]. 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 [32,33]. Fryback und Thornbury haben dazu in ihrer Publikation „The efficacy of diagnostic imaging“ von 1991 [32] 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 7).

Tabelle 7: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren

	Köbberling et al. [33]	Fryback & Thornbury [32]	Evidenzklassifizierung G-BA [34]
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 bei der Bewertung diagnostischer Tests (§ 11 Abs. 2) dieser Sichtweise angeschlossen [34]. In Tabelle 7 sind verschiedene Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren einander gegenübergestellt. Neben der Einteilung von Fryback und Thornbury und der Evidenzklassifizierung des G-BA (Tabelle 6) wurde auch das Phasenschema nach Köbberling et al. [33] mit aufgenommen.

Tabelle 8: Evidenzklassifizierung des G-BA für Unterlagen zu diagnostischen Methoden

Evidenzklasse nach G-BA [34]	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 chirurgische Behandlung von Patienten mit Epilepsie zu bewerten, genügt es nicht festzustellen, inwieweit die PET bzw. die PET/CT den gängigen diagnostischen Methoden hinsichtlich einer korrekten Lokalisation beziehungsweise Detektion epileptogener Zonen 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 [35]. 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.

Vorabrecherchen haben ergeben, dass zur Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT bei der Lokalisation epileptogener Zonen für die chirurgische Behandlung der Epilepsie vermutlich keine qualitativ ausreichende systematische Übersicht 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

Studien zu mindestens 10 Patienten mit Epilepsie werden eingeschlossen, die potenziell geeignet sind für einen chirurgischen Eingriff.

4.1.2.2 Prüf- und Vergleichsinterventionen

Als Prüfindervention 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 Prüfindervention 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 [35].

4.1.2.3 Patientenrelevante Endpunkte

Für die Untersuchung werden folgende patientenrelevante Endpunkte verwendet:

- Mortalität
- Morbidität, wie z. B.
 - postoperative Anfallsfreiheit über mindestens 12 Monate
 - unerwünschte neurologische oder neuropsychologische Ereignisse
- Gesundheitsbezogene Lebensqualität

Eine Änderung im Patientenmanagement allein (ohne Verknüpfung zu den oben genannten patientenrelevanten Endpunkten) stellt keine Zielgröße im Sinne dieses Berichts dar und wird daher nur ergänzend betrachtet.

Subjektive Endpunkte (zum Beispiel gesundheitsbezogene Lebensqualität) werden nur dann berücksichtigt, wenn sie mit validierten Messinstrumenten (zum Beispiel SF-36) erfasst worden sind.

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 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 Studiendauer

Hinsichtlich der Studiendauer besteht keine Einschränkung.

4.1.2.6 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 9).

Tabelle 9: Übersicht der Einschlusskriterien für Primärstudien zur Nutzenbewertung der PET bzw. PET/CT

Einschlusskriterien	
E1a	Mindestens 10 Patienten mit Epilepsie, die potenziell geeignet sind für einen chirurgischen Eingriff (siehe auch Abschnitt 4.1.2.1)
E2a	Vergleich von 2 diagnostischen Strategien <ul style="list-style-type: none"> ▪ PET versus Strategie ohne PET ▪ PET/CT versus Strategie ohne PET/CT ▪ PET/CT versus PET (siehe auch Abschnitt 4.1.2.2)
E3a	Patientenrelevante Endpunkte wie z. B. Mortalität, Morbidität und gesundheitsbezogene Lebensqualität (siehe auch Abschnitt 4.1.2.3)
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
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 [36] oder des TREND-Statements [37] genügt und eine Bewertung der Studie ermöglicht.	

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

In den orientierenden Vorabrecherchen fanden sich systematische Übersichten, die die diagnostische Güte der PET oder PET/CT beim Krankheitsbild der Epilepsie untersuchten. Mit dem Ziel einer möglichst raschen Ergebnisergebnisgewinnung 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, werden die diagnostische und die prognostische Güte anhand von Primärstudien beurteilt. Sind die identifizierten systematischen Übersichten nicht mehr aktuell, wird für den Zeitraum, den 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

Studien zu mindestens 10 Patienten mit Epilepsie werden eingeschlossen, die potenziell geeignet sind für einen chirurgischen Eingriff (siehe Abschnitt 4.1.2.1).

4.2.1.2 Prüftechnologien, Vergleichstechnologien („Komparatoren“) und Referenztests

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 Referenztest (= dem besten zur Verfügung stehenden Test) in Beziehung gebracht worden sein, der für eine adäquate Interpretation der Ergebnisse unabhängig vom Index- und Vergleichstest erhoben worden ist, da es anderenfalls zu Verzerrungen der Güte- und/oder Variabilitätsschätzer kommen kann (sogenannter Verifikationsbias oder Inkorporationsbias) [38-40].

Als Referenztest (= bester zur Verfügung stehender Test) bzw. Referenzstandard gilt die postoperative Anfallsfreiheit als Kriterium des tatsächlich beobachteten individuellen Krankheitsverlaufs im Rahmen der Nachbeobachtung von mindestens einem Jahr nach der Operation, da bei postoperativer Anfallsfreiheit zu folgern ist, dass die epileptogene Zone in der Resektion enthalten war [22]. Ist die postoperative Anfallsfreiheit nicht erfassbar, da beispielsweise gar keine oder keine vollständige Operation durchgeführt wurde, wird auch das intrakranielle EEG als Referenztest akzeptiert. Beide Referenzstandards gehen mit systematischen Verzerrungen einher, stellen jedoch die besten zur Verfügung stehenden Optionen dar. Die postoperative Anfallsfreiheit, die viel mehr als patientenrelevanter Endpunkt als als Referenzstandard betrachtet werden muss, ist abhängig von Qualität, Umfang und Möglichkeit des chirurgischen Eingriffs ebenso wie von der postoperativen medikamentösen Behandlung (Behandlungsparadox [40]). Der Steuerung des chirurgischen Eingriffs gehen wiederum diagnostische Prozeduren (z. B. EEG, MRT und ggfs. PET) voraus, sodass das Ergebnis des chirurgischen Eingriffs – die postoperative Anfallsfreiheit – nur so gut sein kann wie das Ergebnis der vorangegangenen Diagnostik (Interdependenz). Das intrakranielle EEG hingegen wird zum einen in Abhängigkeit von vorangegangenen diagnostischen Prozeduren (z. B. EEG, MRT und ggfs. PET) durchgeführt (differenzielle Verifikation). Zum anderen wird es aus ethischen Gründen nur bei Patienten eingesetzt, bei denen die nicht invasive Evaluation diffizil war oder zu diskordanten Ergebnissen geführt hat.

Allein das Erfordernis eines intrakraniellen EEG ist somit mit einer größeren Wahrscheinlichkeit für eine schlechte Prognose assoziiert ([41,42] zitiert nach [43]).

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 [Receiver Operating Characteristic]) oder positiven / negativen Likelihood-Quotienten, daraus abgeleitet ggf. prädiktive Aussagen mittels z. B. positiven / negativen prädiktiven Werts.
- Prognostische Aussagen (Wahrscheinlichkeit) hinsichtlich zukünftiger klinischer Ereignisse (z. B. postoperative Anfallsfreiheit) anhand geeigneter statistischer Maßzahlen. 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
Querschnittstudien, ggf. mit klinischer Nachbeobachtung, durchgeführt in der Anwendungssituation (gemäß Phase 3 nach Köbberling et al. [33]).
- Prognostische Güte
Prospektiv geplante 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 Zweiten sollen 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 soll die weitere Behandlung dieser Patienten ohne das Wissen um die Ergebnisse der PET erfolgen (Verblindung).

Für die Interpretation der postoperativen Anfallsfreiheit ist ein Follow-up von mindestens 12 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 sind, mit dem Referenztest untersucht werden [44].

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 Publikationen auf dem Qualitätsindex von Oxman und Guyatt mindestens 5 von 7 möglichen Punkten erreichen [45-47].
- Die jeweilige Forschungsfrage muss im Format einer sog. PICO-Frage formuliert sein, 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) 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.

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 10).

Tabelle 10: Übersicht der Einschlusskriterien für systematische Übersichten 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	Mindestens 10 Patienten mit Epilepsie, die potenziell geeignet sind für einen chirurgischen Eingriff (siehe auch Abschnitt 4.1.2.1)
E2b	Prüftechnologie (Vollring-PET bzw. PET/CT), Vergleichstechnologie und Referenztests, Evaluierung wie in Abschnitt 4.2.1.2 beschrieben
E3b	Zielgrößen wie z. B. Sensitivität / Spezifität oder positiver / negativer Likelihood-Quotient (siehe auch Abschnitt 4.2.1.3)
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 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 [48] oder des MOOSE-Statements [49] 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 Populationen

Studien zu mindestens 10 Patienten mit Epilepsie werden eingeschlossen, die potenziell geeignet sind für einen chirurgischen Eingriff (siehe Abschnitt 4.1.2.1).

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

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

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

Es erfolgt keine weitere Einschränkung.

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

Tabelle 11: Übersicht 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	Mindestens 10 Patienten mit Epilepsie, die potenziell geeignet sind für einen chirurgischen Eingriff (siehe auch Abschnitt 4.1.2.1)
E2c	Prüftechnologie (Vollring-PET bzw. PET/CT), Vergleichstechnologie und Referenztests, Evaluierung wie in Abschnitt 4.2.1.2 beschrieben
E3c	Zielgrößen wie z. B. Sensitivität / Spezifität oder positiver / negativer Likelihood-Quotient (siehe auch Abschnitt 4.2.1.3)
E4c	Querschnitt- oder prospektiv geplante Kohortenstudien wie in Abschnitt 4.2.1.4 beschrieben
E5c	Keine Mehrfachpublikationen ohne relevante Zusatzinformationen
E6c	Vollpublikation verfügbar ^a
E7c	Publikationszeitraum durch Literatursuche der eingeschlossenen systematischen Übersichten 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 [50] oder des STROBE-Statements [51] genügt und eine Bewertung der Studie ermöglicht.	

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 erfüllen (Prüf- und Vergleichsintervention bzw. Indextests und Referenztests).

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 relevanten 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)

4.3.2 Suche nach weiteren publizierten und nicht publizierten Studien und systematischen Übersichten

Zusätzlich zur Suche in bibliografischen Datenbanken sollen folgende Quellen zur Identifizierung publizierter und nicht publizierter Studien 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.

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 wird daher gezielt nach systematischen 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, der durch den Recherchezeitraum der systematischen Übersichten nicht abgedeckt wird, eine zusätzliche

Recherche nach Primärliteratur durchgeführt (Ergänzungsrecherche; vgl. Abschnitt 4.2.2). Sollten 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 weiterer Informationsquellen ab. Alle für die Nutzenbewertung relevanten, mit validierten Messinstrumenten erhobene 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 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 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 des Verzerrungspotenzials einer systematischen Übersicht erfolgt anhand der Kriterien von Oxman und Guyatt [46,47]. Diese Kriterien werden um die Angabe von Interessenkonflikten ergänzt. Die Bewertung erfolgt immer 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 QUADAS2-Instruments durchgeführt [52].

Für die Bewertung des Verzerrungspotenzials von prognostischen Studien ist zusätzlich zu QUADAS2 die adäquate Berücksichtigung von möglichen Confoundern als ein weiteres Qualitätskriterium vorgesehen. Auf Grundlage der Arbeiten von Hayden et al. [53] und Altman et al. [54] wurde hierfür ein neues Instrument entwickelt.

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 Responderdefinition 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 [55,56].

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 [57].

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 [58].

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 „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 [59].

Sollten durch eine Ergänzungsrecherche für den Zeitraum nach Abschluss des Recherchzeitraums 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 „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 Studien hinsichtlich der Fragestellung und relevanter Charakteristika vergleichbar sind, werden die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammengefasst. Für die statistische Auswertung vergleichender Studien werden primär die Ergebnisse aus Intention-to-Treat-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet. Die Meta-Analysen erfolgen in der Regel auf Basis von Modellen mit zufälligen Effekten [60]. 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 aus den vorhandenen Informationen eigenständig berechnet beziehungsweise näherungsweise bestimmt.

Studien zum patientenrelevanten Nutzen

Für kontinuierliche Variablen wird die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges' 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 kategorialen Variablen wird ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und den verfügbaren Daten verwendet [61].

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 [62]. 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).

Diagnosestudien

Die Effektschätzer und Konfidenzintervalle aus den Studien werden ebenfalls mittels Forest Plots zusammenfassend dargestellt. Außerdem 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 [63]. Die Schätzung der Modellparameter (Sensitivität, Spezifität und Parameter für die „hierarchical summary receiver operating curve“, HSROC) erfolgt über einen Maximum-Likelihood-Ansatz in einem nicht linearen gemischten Modell. In begründeten Ausnahmefällen kommen auch andere Effektmaße zum Einsatz.

Zusätzlich werden bei diagnostischen Studien die beobachteten Paare aus Sensitivität und Spezifität zweidimensional grafisch dargestellt. Studien, in denen Technologien wie PET und MRT 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 grafisch dargestellt. Schließlich werden die geschätzten HSROC-Kurven abgebildet. Die grafischen Darstellungen von Sensitivität und Spezifität und der HSROC-Kurve können Tendenzen zu einer besseren Testgüte in den vorliegenden Studien aufzeigen. Schneiden sich die Konfidenzregionen zweier Technologien nicht, so wird davon ausgegangen, dass in den vorliegenden Studien ein Unterschied in der diagnostischen Güte besteht. Die Richtung des Unterschieds wird ggf. grafisch bestimmt. Überlappen sich zwei Konfidenzregionen oder schneiden sich zwei HSROC-Kurven im Bereich der Grafik, in dem auch Paare aus Sensitivität und Spezifität beobachtet wurden, wird davon ausgegangen, dass in den vorliegenden Studien kein Unterschied in der diagnostischen Güte besteht.

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

4.5.3 Sensitivitätsanalyse

Zur Einschätzung der Robustheit der Ergebnisse sind Sensitivitätsanalysen hinsichtlich methodischer Faktoren geplant. Die methodischen Faktoren bilden sich aus den im Rahmen der Informationsbeschaffung und bewertung getroffenen Entscheidungen, zum Beispiel die Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder die Wahl des Effektmaßes. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ und „niedrig“ wird für Sensitivitätsanalysen 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 [46,47] durchgeführt werden (Stufe 5 versus 6 versus 7; alle < 5 werden ausgeschlossen).

Der Algorithmus zum Schätzen der Parameter im bivariaten Model zur diagnostischen Güte kann ggf. keinen Schätzer liefern, da das Maximum-Likelihood-Verfahren nicht konvergiert, oder zu unrealistischen Schätzern führen, d. h. zu Schätzern mit zu großen Standardfehlern und entsprechenden Konfidenzintervallen. Die Gründe hierfür können beispielsweise darin bestehen, dass zu wenige Studien vorliegen, als dass die Parameter geschätzt werden könnten.

Aber auch einzelne Studien mit extremen Werten können die Ursache sein. Um diese Studien zu identifizieren, werden Sensitivitätsanalysen durchgeführt.

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 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
- Feldstärke der MRT (Tesla)
- Temporallappenepilepsie (TLE) vs. extra-TLE
- Anzahl an Läsionen im MRT

Sollten sich aus den verfügbaren Informationen Anhaltspunkte für weitere mögliche Effektmodifikatoren ergeben, 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. Fisher RS, Van Emde Boas W, Blume W, Elger C, Genton P, Lee P et al. Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* 2005; 46(4): 470-472.
2. Schirmer M. Neurochirurgie. München: Urban und Fischer; 2004.
3. Forsgren L, Beghi E, Oun A, Sillanpaa M. The epidemiology of epilepsy in Europe: a systematic review. *Eur J Neurol* 2005; 12(4): 245-253.
4. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, Van Emde Boas W et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia* 2010; 51(4): 676-685.
5. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, Van Emde Boas W et al. Revidierte Terminologie und Konzepte zur Einteilung von epileptischen Anfällen und Epilepsien: Bericht der Klassifikations- und Terminologiekommission der Internationalen Liga gegen Epilepsie, 2005-2009. *Epileptologie* 2010; 27(2): 101-114.
6. Olafsson E, Hauser WA, Ludvigsson P, Gudmundsson G. Incidence of epilepsy in rural Iceland: a population-based study. *Epilepsia* 1996; 37(10): 951-955.
7. Forsgren L, Bucht G, Eriksson S, Bergmark L. Incidence and clinical characterization of unprovoked seizures in adults: a prospective population-based study. *Epilepsia* 1996; 37(3): 224-229.
8. Sidenvall R, Forsgren L, Blomquist HK, Heijbel J. A community-based prospective incidence study of epileptic seizures in children. *Acta Paediatr* 1993; 82(1): 60-65.
9. Sander JW, Hart YM, Johnson AL, Shorvon SD. National General Practice Study of Epilepsy: newly diagnosed epileptic seizures in a general population. *Lancet* 1990; 336(8726): 1267-1271.
10. Oun A, Haldre S, Magi M. Incidence of adult epilepsy in Estonia. *Acta Neurol Scand* 2003; 108(4): 245-251.
11. Kwan P, Sander JW. The natural history of epilepsy: an epidemiological view. *J Neurol Neurosurg Psychiatry* 2004; 75(10): 1376-1381.
12. Berg AT, Shinnar S. Do seizures beget seizures? An assessment of the clinical evidence in humans. *J Clin Neurophysiol* 1997; 14(2): 102-110.

13. Sander JW. Some aspects of prognosis in the epilepsies: a review. *Epilepsia* 1993; 34(6): 1007-1016.
14. Semah F, Picot MC, Adam C, Broglin D, Arzimanoglou A, Bazin B et al. Is the underlying cause of epilepsy a major prognostic factor for recurrence? *Neurology* 1998; 51(5): 1256-1262.
15. Baumgartner C, Patariaia E. Präoperative Epilepsiediagnostik und Epilepsiechirurgie. In: Baumgartner C (Ed). *Handbuch der Epilepsien: Klinik, Diagnostik, Therapie und psychosoziale Aspekte*. Wien: Springer; 2001. S. 388-403.
16. Tellez-Zenteno JF, Dhar R, Wiebe S. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. *Brain* 2005; 128(Pt 5): 1188-1198.
17. Schmidt D, Stavem K. Long-term seizure outcome of surgery versus no surgery for drug-resistant partial epilepsy: a review of controlled studies. *Epilepsia* 2009; 50(6): 1301-1309.
18. Seiam AHR, Dhaliwal H, Wiebe S. Determinants of quality of life after epilepsy surgery: systematic review and evidence summary. *Epilepsy Behav* 2011; 21(4): 441-445.
19. Forsgren L, Hauser WA, Olafsson E, Sander JWAS, Sillanpää M, Tomson T. Mortality of epilepsy in developed countries: a review. *Epilepsia* 2005; 46(Suppl 11): 18-27.
20. Binnie CD, Polkey CE. Commission on Neurosurgery of the International League Against Epilepsy (ILAE) 1993-1997: recommended standards. *Epilepsia* 2000; 41(10): 1346-1349.
21. European Federation of Neurological Societies Task Force. Pre-surgical evaluation for epilepsy surgery: European standards. *Eur J Neurol* 2000; 7(1): 119-122.
22. Klein KM, Rosenow F. The epileptogenic zone: general principles. In: Lüders HO (Ed). *Textbook of epilepsy surgery*. London: Informa; 2008. S. 899-901.
23. Czech T, Novak K. Epilepsiechirurgische Verfahren. In: Baumgartner C (Ed). *Handbuch der Epilepsien: Klinik, Diagnostik, Therapie und psychosoziale Aspekte*. Wien: Springer; 2001. S. 404-411.
24. Engel J Jr, Van Ness PC, Rasmussen TB, Ojemann LM. Outcome with respect to epileptic seizures. In: Engel J Jr (Ed). *Surgical treatment of the epilepsies*. New York: Raven Press; 1993. S. 609-621.
25. Wieser HG, Blume WT, Fish D, Goldensohn E, Hufnagel A, King D et al. ILAE Commission Report: proposal for a new classification of outcome with respect to epileptic seizures following epilepsy surgery. *Epilepsia* 2001; 42(2): 282-286.

26. Baumgartner C. Anfallsformen: klinische Symptomatik. In: Baumgartner C (Ed). Handbuch der Epilepsien: Klinik, Diagnostik, Therapie und psychosoziale Aspekte. Wien: Springer; 2001. S. 17-34.
27. Cross JH, Jayakar P, Nordli D, Delalande O, Duchowny M, Wieser HG et al. Proposed criteria for referral and evaluation of children for epilepsy surgery: recommendations of the Subcommittee for Pediatric Epilepsy Surgery. *Epilepsia* 2006; 47(6): 952-959.
28. Deutsche Gesellschaft für Epileptologie, AG für prächirurgische Diagnostik und operative Epilepsitherapie. Definition von Epilepsiezentren [online]. 12.11.2010 [Zugriff: 18.05.2011]. URL: <http://www.dgfe.info/download/komm/definitionepilepsiezentren.pdf>.
29. Konsensus: Neuro-PET. *Nuklearmedizin* 1997; 36(7): 46-47.
30. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 3.0 [online]. 27.05.2008 [Zugriff: 17.03.2011]. URL: http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf.
31. How to read clinical journals; II: to learn about a diagnostic test. *Can Med Assoc J* 1981; 124(6): 703-710.
32. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2): 88-94.
33. Köbberling J, Trampisch HJ, Windeler J. Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem* 1990; 28(12): 873-879.
34. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. 17.12.2009 [Zugriff: 18.03.2011]. URL: http://www.g-ba.de/downloads/62-492-422/VerfO_2009-12-17.pdf.
35. Lijmer JG, Bossuyt PMM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009; 62(4): 364-373.
36. 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.
37. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004; 94(3): 361-366.
38. 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.

39. 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.
40. 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.
41. Benbadis SR, So NK, Antar MA, Barnett GH, Morris HH. The value of PET scan (and MRI and Wada test) in patients with bitemporal epileptiform abnormalities. *Arch Neurol* 1995; 52(11): 1062-1068.
42. Salanova V, Markand O, Worth R. Focal functional deficits in temporal lobe epilepsy on PET scans and the intracarotid amobarbital procedure: comparison of patients with unitemporal epilepsy with those requiring intracranial recordings. *Epilepsia* 2001; 42(2): 198-203.
43. Willmann O, Wennberg R, May T, Woermann FG, Pohlmann-Eden B. The contribution of ¹⁸F-FDG PET in preoperative epilepsy surgery evaluation for patients with temporal lobe epilepsy: a meta-analysis. *Seizure* 2007; 16(6): 509-520.
44. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics* 2001; 2(3): 249-260.
45. Jadad AR, Murray WE. *Randomized controlled trials: questions, answers and musings*. Malden: Blackwell; 2007.
46. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44(11): 1271-1278.
47. 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.
48. 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.
49. 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.
50. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. *Ann Intern Med* 2003; 138(1): 40-44.

51. 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.
52. University of Bristol. QUADAS 2 [online]. 19.07.2011 [Zugriff: 29.07.2011]. URL: <http://www.bris.ac.uk/quadas/quadas-2>.
53. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; 144(6): 427-437.
54. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ* 2001; 323(7306): 224-228.
55. Cohen J (Ed). *Statistical power analysis for the behavioral sciences*. New York: Academic Press; 1977.
56. Fayers PM, Hays R (Ed). *Assessing quality of life in clinical trials: methods and practice*. Oxford: Oxford University Press; 2005.
57. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
58. 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.
59. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156(10): 1411-1416.
60. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
61. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 243-296.
62. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
63. 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.