

Positronenemissionstomographie (PET) und PET/CT bei Kopf- und Halstumoren

Berichtsplan (vorläufige Version)

Auftrag D06-01B
Version 0.1
Stand: 27.05.2009

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Positronenemissionstomographie (PET) und PET/CT bei Kopf- und Halstumoren

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

21.12.2006

Interne Auftragsnummer:

D06-01B

Anschrift des Herausgebers:

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

Tel.: 0221/35685-0

Fax: 0221/35685-1

Berichte@iqwig.de

www.iqwig.de

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

Schlagwörter: Positronenemissionstomographie, Computertomographie, Kopf- und Halstumore, unbekannter Primärtumor, Staging, Rezidiv, Behandlungsresiduen, systematische Übersicht

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	iv
Abkürzungsverzeichnis	v
Abkürzungsverzeichnis	v
1 Hintergrund	6
2 Ziele der Untersuchung	17
2.1 Ermittlung des patientenrelevanten Nutzens der PET bzw. PET/CT	17
2.2 Bewertung der diagnostischen und prognostischen Wertigkeit der PET bzw. PET/CT	18
3 Projektablauf	20
3.1 Verlauf des Projekts	20
4 Methoden	21
4.1 Ermittlung des patientenrelevanten Nutzens der PET bzw. PET/CT	23
4.1.1 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens der PET bzw. PET/CT anhand von systematischen Übersichten.....	24
4.1.2 Einschlusskriterien zur Bewertung des patientenrelevanten Nutzens der PET bzw. PET/CT anhand von Primärstudien.....	24
4.1.2.1 Population.....	24
4.1.2.2 Prüf- und Vergleichsintervention	24
4.1.2.3 Zielgrößen	25
4.1.2.4 Studientypen.....	26
4.1.2.5 Sonstige Studiencharakteristika	26
4.1.3 Ein- / Ausschlusskriterien	26
4.2 Bewertung der diagnostischen und prognostischen Wertigkeit der PET bzw. PET/CT	27
4.2.1 Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Wertigkeit anhand von systematischen Übersichten	27
4.2.1.1 Population.....	27
4.2.1.2 Prüftechnologie, Vergleichstechnologie („Komparator“) und Referenzstandard	27
4.2.1.3 Zielgrößen	28

4.2.1.4	Den systematischen Übersichten zugrunde liegende Studientypen	28
4.2.1.5	Sonstige Charakteristika.....	29
4.2.1.6	Ein- / Ausschlusskriterien	29
4.2.2	Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Wertigkeit anhand von Primärstudien.....	30
4.2.2.1	Population.....	30
4.2.2.2	Prüftechnologie, Vergleichstechnologie („Komparator“) und Referenzstandard	30
4.2.2.3	Zielgrößen	30
4.2.2.4	Studientypen.....	31
4.2.2.5	Sonstige Charakteristika.....	31
4.2.2.6	Ein- / Ausschlusskriterien	31
4.3	Informationsbeschaffung.....	32
4.3.1	Bibliografische Literaturrecherche.....	32
4.3.2	Suche nach weiteren publizierten und nicht publizierten Studien	32
4.3.3	Selektion relevanter Studien.....	33
4.4	Informationsbewertung	33
4.4.1	Bewertung von Studien zum patientenrelevanten Nutzen	33
4.4.2	Bewertung von systematischen Übersichten zur diagnostischen und prognostischen Wertigkeit.....	34
4.4.3	Bewertung von Primärstudien zur diagnostischen und prognostischen Wertigkeit.....	35
4.5	Informationssynthese und -analyse	35
4.5.1	Gegenüberstellung der Ergebnisse	35
4.5.1.1	Charakterisierung und Darstellung der Ergebnisse von einzelnen (Primär-) Studien zur Bewertung des patientenrelevanten Nutzens	36
4.5.1.2	Charakterisierung und Darstellung der Ergebnisse von systematischen Übersichten zur diagnostischen und prognostischen Wertigkeit	36
4.5.1.3	Charakterisierung und Darstellung der Ergebnisse von Primärstudien zur diagnostischen und prognostischen Wertigkeit.....	37
4.5.2	Meta-Analysen	37
4.5.3	Sensitivitätsanalyse	38
4.5.4	Subgruppenmerkmale und andere Effektmodifikatoren	38
5	Literaturverzeichnis.....	40

Tabellenverzeichnis

Tabelle 1: Inzidenz und Mortalität pro 100 000 Einwohner und Jahr von Mund- und Rachentumoren sowie Kehlkopftumoren in Deutschland 2004 nach Geschlecht (nach Robert Koch-Institut) [7]	6
Tabelle 2: Verteilung der häufigsten malignen Kopf- und Halstumorphistologien in den USA von 1999 bis 2001 nach Lokalisation. Mittlere Prozentzahlen für jede Lokalisation nach Davies [1] *	8
Tabelle 3: Stadieneinteilung für die nasopharyngealen Tumoren nach der Klassifikation des America Joint Committee on Cancer (AJCC) von 2002 [16,21]	11
Tabelle 4: Stadieneinteilung für labiale, orale, oropharyngeale, hypopharyngeale, laryngeale Tumoren, Tumoren der Speicheldrüsen, der Nasenhöhle und der paranasalen Sinus nach der Klassifikation des American Joint Committee on Cancer (AJCC) von 2002 [16,19]	11
Tabelle 5: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren	22
Tabelle 6: Evidenzklassifizierung des G-BA für Unterlagen zu diagnostischen Methoden [38]	23

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AJCC	America Joint Committee on Cancer
CONSORT	Consolidated Standards of Reporting Trials
CT	Computertomographie
CUP	Cancer of Unknown Primary site
DOR	Diagnostische Odds Ratio
EBV	Epstein-Barr-Virus
EGFR	Epidermal growth factor receptor
ESMO	European Society for Medical Oncology
FDG	Fluor-Deoxy-Glukose
G-BA	Gemeinsamer Bundesausschuss
HPV	Human Papillomavirus
HTA	Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IMRT	Intensitätsmodulierte Radiotherapie
is	In situ
ITT	Intention-to-Treat
KI	Konfidenzintervall
MRT	Magnet-Resonanz-Tomographie
NCCN	National Comprehensive Cancer Network
PET	Positronenemissionstomographie
PET/CT	Integration von PET (Positronenemissionstomographie) und CT (Computertomographie) in einem Untersuchungsgerät
RCT	Randomised Controlled Trial (randomisierte kontrollierte Studie)
SEER	National Cancer Institute's Surveillance Epidemiology and End Results Programme
SUV	Standardized Uptake Value
SROC	Summary-Receiver-Operating-Characteristic
VEGF	Vascular endothelial growth factor

1 Hintergrund

Definition des Krankheitsbildes

Bei Kopf- und Halstumoren handelt es sich um eine Gruppe von Tumoren, die von sehr unterschiedlichen Ursprungszellen abstammen können und deren Gemeinsamkeit zum Teil nur in der anatomischen Lokalisation liegt. Die Definition, was anatomisch zu den Kopf- und Halstumoren gehört, ist nicht eindeutig festgelegt [1]. Für diesen Bericht werden maligne Tumoren der Lippe, Mundhöhle, Pharynx, Larynx, Speicheldrüsen und des sinonasalen Bereiches zu den Kopf- und Halstumoren gezählt. Zusätzlich behandelt dieser Bericht die unbekannt Primärtumoren, insbesondere mit zervikaler Adenopathie (s. auch Abschnitt 4.1.2.1 für Details).

Bei den meisten Lokalisationen im Kopf-Hals-Bereich dominiert das Plattenepithelkarzinom [2-5] und wird auch bei diesem Bericht vor allen Dingen berücksichtigt werden können. Es gehören aber auch so unterschiedliche Tumoren wie mukoepidermoide Karzinome und Ästhesioneuroblastome dazu [1].

Epidemiologie und individuelle Krankheitslast

Die malignen Kopf- und Halstumoren sind weltweit die fünfthäufigsten Tumoren [5,6]. Zwei Drittel der Fälle treten in Entwicklungsländern auf [4]. Die Zahl der jährlichen Tumorneuerkrankungen in der Mundhöhle oder im Rachen ist in Deutschland bei Männern fast dreimal so hoch wie bei Frauen (siehe Tabelle 1), bei Larynxtumoren ist die Inzidenz bei Männern etwa achtmal so hoch wie bei Frauen [7]. Die relative 5-Jahres-Überlebensrate für Kehlkopftumoren für Männer liegt bei 6 % und für Frauen bei ca. 62 %, für die anderen Tumoren im Mund- und Rachenbereich liegt sie in Deutschland für Männer bei 47 % und für Frauen bei 55 %. Insgesamt unterscheiden sich die 5-Jahres-Überlebensraten je nach Lokalisation jedoch sehr stark. So liegen sie bei über 90 % für Lippentumoren und zwischen 30 und 40 % bei Rachentumoren [7].

Tabelle 1: Inzidenz und Mortalität pro 100 000 Einwohner und Jahr von Mund- und Rachentumoren sowie Kehlkopftumoren in Deutschland 2004 nach Geschlecht (nach Robert Koch-Institut) [7]

Erkrankung	Männer		Frauen	
	Inzidenz	Mortalität	Inzidenz	Mortalität
Mund- und Rachentumoren ^a	18,9	8,5	6,6	2,4
Kehlkopftumoren	7,4	3,3	0,9	0,4

a: umfasst bösartige Neubildungen der Lippe, der Zunge, des Mundbodens, des Gaumens, der Speicheldrüsen und des Rachens

Davies et al. untersuchten die Inzidenz je nach Histologie und Lokalisation im Kopf-Hals-Bereich für den Zeitraum von 1975 bis 2001 in den USA anhand der Daten aus dem National Cancer Institute's Surveillance Epidemiology and End Results (SEER)-Programm. Ihre Definition von Kopf- und Halstumoren umfasste alle Tumoren von den Schleimhäuten des aerodigestiven Traktes. Zusätzlich schlossen sie Tumoren der Schilddrüse, der Nebenschilddrüsen, der Speicheldrüsen, der Maxilla, der Mandibula und der Temporalknochen sowie der inneren Weichteile wie Gefäße, Nerven, Lymphknoten und Bindegewebe mit ein. Augen- und Hirntumoren sowie primäre Hauttumoren schlossen sie aus.

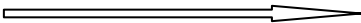
Die zweit-, dritt- und vierthäufigsten Tumorlokalisationen in der Studie von Davies et al. waren relevant für diesen Bericht: Larynx (15 %), oropharyngeale Mukosa (12 %) und Zunge (10 %). Wie aus der Tabelle 2 ersichtlich, war das Plattenepithelkarzinom in den meisten Lokalisationen der häufigste maligne Tumor [1]. Im Bereich der Kieferhöhlen und Speicheldrüsen findet sich ein relativ weites Spektrum an Tumoren, wobei die Verteilung auch nochmals je nach betroffener Speicheldrüse unterschiedlich ist [1].

Ursache des Gesundheitsproblems bzw. der Erkrankung

Die Tumoren in der Kopf-Hals-Region sind oft durch vermeidbare Expositionen zumindest mitbedingt. Wichtige Risikofaktoren sind Rauchen, Alkoholkonsum und oraler Genuss von Tabak (Kautabak). Holzstaub, polyzyklische aromatische Kohlenwasserstoffe und, bei einigen Tumorarten Virusinfekte (z. B. humanes Papilloma-Virus [HPV], Epstein-Barr-Virus [EBV]) sind weitere Risikofaktoren [1,4-12]. Starke Raucher haben, verglichen mit Nichtrauchern, ein 5- bis 25-fach erhöhtes Risiko, Kopf- und Halstumoren zu entwickeln. Alkoholkonsum erhöht dieses Risiko nochmals, wobei der Effekt vermutlich nicht nur additiv ist, sondern potenziert wird [4,5,9,10]. Das HPV wird vor allem mit oropharyngealen Tumoren assoziiert. Die Patienten sind tendenziell jünger. Es gibt Hinweise, dass HPV-assoziierte Kopf- und Halstumoren eine bessere Prognose – insbesondere bei Patienten, die nicht rauchen oder Alkohol trinken – haben könnten [6,8-10].

Nasopharyngeale Tumoren und Tumoren der Speicheldrüsen unterscheiden sich in ihrer Ätiologie von den klassischen Kopf- und Halstumoren [8,13]. Das EBV ist ein Risikofaktor für nasopharyngeale Tumoren [4,12]. Bei Tumoren der Speicheldrüsen ist die Wahrscheinlichkeit, dass ein Tumor maligne ist, umso größer, je kleiner die Speicheldrüse ist. So sind von den Tumoren in der Parotis 20 bis 30 % maligne. In der submandibulären, der lingualen und den anderen kleinen Speicheldrüsen sind 45 bis 60 %, 70 bis 85 % und 49 bis 80 % der Tumoren maligne [14].

Tabelle 2: Verteilung der häufigsten malignen Kopf- und Halstumorphologien in den USA von 1999 bis 2001 nach Lokalisation. Mittlere Prozentzahlen für jede Lokalisation nach Davies [1] *

Lokalisation	am häufigsten  am wenigsten häufig				
Lippe	Plattenepithelkarzinom	Andere Adenokarzinome	Mukoepidermoides Karzinom	Adenoidzystisches Karzinom	Lymphom
	93,6	2,8	1,4	1,0	0,1
Zunge	Plattenepithelkarzinom	Lymphom	Mukoepidermoides Karzinom	Adenoidzystisches Karzinom	Andere Adenokarzinome
	93,2	3,4	1,0	0,6	0,2
Tonsillen	Plattenepithelkarzinom	Lymphom	Basaloid	Plasmozytom	Undifferenziertes Karzinom
	84,8	13,8	0,3	0,3	0,3
Oropharyngeale Mukosa	Plattenepithelkarzinom	Mukoepidermoides Karzinom	Lymphom	Andere Adenokarzinome	Adenoidzystisches Karzinom
	81,6	5,5	3,8	3,2	2,1
Nase	Plattenepithelkarzinom	Lymphom	Undifferenziertes Karzinom	Ästhesioneuroblastom	Andere Adenokarzinome
	56,8	13,6	7,5	3,6	3,5
Para-nasaler Sinus	Plattenepithelkarzinom	Lymphom	Andere Adenokarzinome	Adenoidzystisches Karzinom	Ästhesioneuroblastom
	38,5	20,7	11,4	5,9	4,5
Hypo-pharynx	Plattenepithelkarzinom	Lymphom	Andere Adenokarzinome	Adenoidzystisches Karzinom	-
	96,3	0,5	0,5	0,3	-
Larynx und Trachea	Plattenepithelkarzinom	Epitheliale Neoplasien und komplex epitheliale Neoplasien	Andere Adenokarzinome	Adenoidzystisches Karzinom	Lymphom
	97,4	1,0	0,5	0,2	0,2
Speicheldrüsen	Mukoepidermoides Karzinom	Lymphom	Plattenepithelkarzinom	Andere Adenokarzinome	Azinäres Adenokarzinom
	21,7	17,2	16,1	12,8	10,3

*Es sind hier nur die für den Bericht relevanten Lokalisationen aufgeführt; es sind nur die 5 häufigsten Histologien für jede Lokalisation dargestellt.

Bei Patienten mit Kopf- und Halstumoren besteht lebenslang ein erhöhtes Risiko für Zweittumoren im Kopf-Hals-Bereich [5,6,10,13]. Slaughter et al. gingen davon aus, dass das orale Epithel durch Karzinogenexposition für die Entwicklung von Tumoren prädisponiert wird, und prägten dafür den Begriff der „field cancerization“ [9,10]. Das kumulative Risiko für Zweitneoplasien nach Kopf- und Halstumoren (außer Speicheldrüsentumoren und Tumoren des Nasopharynx) liegt einer Studie zufolge nach 20 Jahren bei 36 %. Dies schloss sowohl Zweittumoren im Kopf-Hals-Bereich als auch andere tabakassoziierte Tumoren mit ein. Tumoren, die an der gleichen Lokalisation wie der Primärtumor auftraten, wurden dabei nicht berücksichtigt, da sie in den verschiedenen Tumorregistern nicht konsistent berichtet wurden [13].

Unbekannte Primärtumoren

Die meisten Patienten, die älter als 40 Jahre sind und sich mit einer Raumforderung am Hals präsentieren, haben einen metastatischen Tumor. Der Primärtumor wird meist im Rahmen der Abklärungen im Kopf-Hals-Bereich gefunden. Bestätigt die Histologie das Vorliegen einer Metastase, ohne dass der Primärtumor im Rahmen der Routinediagnostik gefunden wird, ist die Definition eines unbekanntes Primärtumors oder CUPs (Cancer of unknown primary site) erfüllt [15]. Bei Patienten mit CUP findet sich der Primärtumor auch bei intensiver Abklärung meist nicht [16]. Unbekannte Primärtumoren machen zwischen 2 bis 5 % aller Tumoren im Kopf-Hals-Bereich aus [4,15].

Verlauf der Erkrankung (natürlicher Verlauf)

Das mediane Erkrankungsalter liegt für Plattenepithelkarzinome bei ca. 60 Jahren, aber die Inzidenz bei jüngeren Patienten (< 40 Jahre) scheint zuzunehmen [5,6,8]. Die Symptome variieren je nach Lokalisation und Stadium des Tumors [4,9]. In den frühen Stadien haben die Patienten z. T. vage Symptome oder können asymptomatisch sein und der Befund wird nur als Zufallsbefund erhoben [4,6,8,9].

Patienten mit nasopharyngealen Karzinomen können sich mit Epistaxis, Nasenobstruktion, Ausfluss, Tinnitus, Schwerhörigkeit, Raumforderungen am Hals und zum Teil mit Kopfschmerzen, Diplopie, Schmerzen oder verringerter Sensibilität im Gesicht präsentieren [4,9,17].

Orale oder oropharyngeale Tumoren können insbesondere unspezifische Schmerzen, nicht heilende Ulzera, lockere Zähne, Blutungen, Dysarthrie, Dysphagie, Odynophagie, Otalgie sowie Raumforderungen an der Stelle des Primärtumors oder zervikale Lymphknotenmetastasen verursachen. Es werden auch Nervenläsionen durch die Tumoren beschrieben [4,6]. Patienten mit Tumoren im Hypopharynx präsentieren sich häufig erst spät im Krankheitsverlauf mit Dysphagie, Otalgie, Heiserkeit und oft mit zervikalen Lymphknotenmetastasen [4].

Die Symptomatik von Larynxtumoren hängt von der genauen Lokalisation des Befalls ab. Tumoren der Glottis verursachen Heiserkeit und in fortgeschrittenen Stadien gelegentlich Kurzatmigkeit und werden meist frühzeitig entdeckt [4,9]. Supraglottische Tumoren werden erst in einem späteren Stadium entdeckt und es finden sich meist schon Raumforderungen am Hals. Sie können auch mit Schmerzen, Otalgie, Dysphagie oder Dysphonie verbunden sein [4,9].

Kopf- und Halstumoren metastasieren am häufigsten in die Lunge, aber auch in Lymphknoten, Leber und Knochen [6,8,9,18]. Patienten mit z. B. Lungenmetastasen eines oralen Plattenepithelkarzinoms haben eine mittlere Überlebenszeit von 8,9 Monaten [6]. Je nach Histologie sind die Prognose und die Tendenz zu metastasieren bei den Speicheldrüsen unterschiedlich [16]. Nasopharyngeale Tumoren metastasieren von den Kopf- und Halstumoren am häufigsten [16].

Unbekannte Primärtumoren mit Erstmanifestation im Kopf-Hals-Bereich präsentieren sich meist als Raumforderung im oberen oder mittleren Hals [4].

Klassifikationsschemata

Die Tumoren werden nach der TNM-Klassifikation in Stadien eingeteilt. Dabei werden 2 Klassifikationen unterschieden: cTNM, was der klinischen Klassifikation vor einer Behandlung entspricht, und pTNM, was der histopathologischen Klassifikation nach der Operation entspricht. In Tabelle 3 ist diese für die nasopharyngealen Tumoren und in Tabelle 4 für alle anderen Lokalisationen im Kopf-Hals-Bereich aufgeführt. Hierbei ist zu beachten, dass insbesondere die lokale Ausdehnung des Primärtumors je nach Lokalisation unterschiedlich definiert ist [16,19].

Eine differenziertere Klassifikation ist anhand einer zusätzlichen Einteilung nach histopathologischem Grading, Infiltration der Lymphbahnen, Infiltration der Venen sowie An- oder Abwesenheit von Resttumor nach Behandlung möglich [20].

Das Tumorstadium zum Zeitpunkt der Diagnosestellung bestimmt die Prognose für das Überleben am besten [6,16,17]. Im Allgemeinen beträgt die Überlebensrate für Patienten mit Kopf- und Halstumoren im Stadium III oder IV weniger als 50 % der Überlebensrate von Patienten mit frühen Tumorstadien [16].

Tabelle 3: Stadieneinteilung für die nasopharyngealen Tumoren nach der Klassifikation des America Joint Committee on Cancer (AJCC) von 2002 [16,21]

Stadium	Primärtumor	Regionaler Lymphknotenbefall	Fernmetastasen
0	T _{is}	N ₀	M ₀
I	T ₁	N ₀	M ₀
II A	T _{2a}	N ₀	M ₀
II B	T _{2b}	N ₀₋₁	M ₀
	T ₁₋₂ oder T _{2a}	N ₁	M ₀
III	T _{1-2a/b}	N ₂	M ₀
	T ₃	N ₀₋₂	M ₀
IV A	T ₄	N ₀₋₂	M ₀
IV B	Jedes Primärtumorstadium	N ₃	M ₀
IV C	Jedes Primärtumorstadium	Jedes Lymphknotenstadium	M ₁
is: in situ			

Tabelle 4: Stadieneinteilung für labiale, orale, oropharyngeale, hypopharyngeale, laryngeale Tumoren, Tumoren der Speicheldrüsen, der Nasenhöhle und der paranasalen Sinus nach der Klassifikation des American Joint Committee on Cancer (AJCC) von 2002 [16,19]

Stadium	Primärtumor^a	Regionaler Lymphknotenbefall	Fernmetastasen
0 ^b	T _{is}	N ₀	M ₀
I	T ₁	N ₀	M ₀
II	T ₂	N ₀	M ₀
III	T ₃	N ₀	M ₀
	T ₁₋₃	N ₁	M ₀
IVA	T _{4a}	N ₀₋₁	M ₀
	T ₁₋₃ oder T _{4a}	N ₂	M ₀
IVB	T _{4b}	Jeder Lymphknotenstatus	M ₀
	Jeder Primärtumorstatus	N ₃	M ₀
IV C	Jeder Primärtumorstatus	Jeder Lymphknotenstatus	M ₁
a: Je nach Lokalisation des Primärtumors unterscheidet sich diese Definition.			
b: Bei den Speicheldrüsen gibt es kein Stadium 0.			

Diagnostische Verfahren

Es gibt keine aktuellen deutschen Leitlinien zur Diagnostik und Therapie von Kopf- und Halstumoren. Daher werden hier die Empfehlungen des National Comprehensive Cancer Network (NCCN, USA) und der European Society for Medical Oncology (ESMO) präsentiert.

Neben der Anamnese und Palpation sollten alle Tumoren bioptisch untersucht werden [16]. Abhängig von der Histologie und Lokalisation des Primärtumors werden weitere Abklärungen wie z. B. Laboruntersuchungen auf HPV oder EBV, Endoskopien und die Untersuchung der Sprache und des Schluckens empfohlen [16].

Auch die empfohlene Bildgebung ist unterschiedlich, je nach befallenem Areal im Kopf-Hals-Bereich. Eine Bildgebung des Thorax wird sehr häufig und Panorama-Aufnahmen des Kiefers bei Tumoren der Mundhöhle, des Oropharynx und der Lippe bei entsprechender Indikation empfohlen [16].

Konventionelle Bildgebung: CT und MRT

In den Leitlinien des NCCN von 2008 wird für Tumoren des ethmoidalen und der maxillären Sinus, der Speicheldrüsen, des Oropharynx, des Hypopharynx, des Nasopharynx und des Larynx für das Staging eine Bildgebung mit CT (Computertomographie) und / oder MRT (Magnet-Resonanz-Tomographie) empfohlen [16]. Für Tumoren der Mundhöhle soll eine Abklärung mit CT/MRT nur bei entsprechender Indikation erfolgen, die aber nicht weiter spezifiziert wird. Eine CT oder MRT sollte bei Primärtumoren der Lippe durchgeführt werden, wenn ein Knochenbefall vermutet wird [16].

Die ESMO empfiehlt für Tumoren des Nasopharynx und für das Plattenepithelkarzinom des Kopfes und des Halses routinemäßig eine Bildgebung mit CT oder MRT beim Staging [2,22]. Beim Plattenepithelkarzinom im Kopf-Hals-Bereich könne eine Thorax-CT durchgeführt werden, um Lungenmetastasen auszuschließen [2].

Für die Nachbeobachtung empfehlen die NCCN-Leitlinien eine Bildgebung als Baseline bei Tumoren des ethmoidalen und der maxillären Sinus mit CT/MRT sowie bei den fortgeschritteneren Stadien der Tumoren des Oropharynx, des Hypopharynx, des Nasopharynx und des glottischen und supraglottischen Larynx. Je nach Klinik empfehlen sie bei diesen fortgeschritteneren Stadien weitere Bildgebung im Verlauf. Eine Bildgebung des Thorax empfehlen sie bei allen Tumoren außer denen der Lippe, abhängig von der klinischen Indikation [16].

Laut ESMO soll das Therapieansprechen beim Plattenepithelkarzinom im Kopf-Hals-Bereich mit der gleichen Technologie (CT oder MRT) wie beim Staging untersucht werden. Bei der Nachbeobachtung soll eine radiologische Bildgebung eingeschlossen werden [2].

PET und PET/CT

Das NCCN erwähnt die PET oder PET/CT zum Staging bei folgenden Indikationen: Beim CUP sollte PET/CT nur durchgeführt werden, wenn mit den anderen Tests kein Primärtumor identifiziert werden konnte. PET kann genutzt werden, um einen klinischen Eindruck zu bestätigen, einen unbekanntem Primärtumor zu finden und zur Überwachung. Die WHO-Klassifikation wird laut NCCN am häufigsten für die Einteilung der nasopharyngealen Tumoren benutzt. Typ 1 nach WHO umfasst gut bis mäßig gut differenzierte Plattenepithelkarzinome, Typ 2 nicht verhornende Tumoren, inklusive Schneider'sches Papillom und Lymphoepitheliom, und Typ 3 undifferenzierte Karzinome, inklusive Lymphoepitheliom, anaplastischen, klar- und spindelzelligen Varianten. Bei der Abklärung von Fernmetastasen bei Tumoren des Nasopharynx WHO-Klasse 2 bis 3 mit N2-3 (AJCC) kann die Bildgebung für Fernmetastasen PET und / oder CT einschließen [16]. Die ESMO-Leitlinien beschreiben die PET für die bildgebende Abklärung von Fernmetastasen bei Hochrisikopatienten mit nasopharyngealem Tumor als vielversprechend. Andere Möglichkeiten seien Knochenszintigraphie und CT [22]. Wenn die PET für das Follow-up von laryngealen oder oropharyngealen Tumoren genutzt wird, dann empfehlen die NCCN-Leitlinien, die PET-Untersuchung frühestens 12 Wochen nach Therapieende durchzuführen, um die Rate der falsch positiven Befunde zu reduzieren [16].

Gründe für eine falsch negative PET bei Kopf- und Halstumoren sind fehlende Fluor-Deoxy-Glukose (FDG)-Avidität des Tumors, zu kleines Tumolvolumen, zu frühe Bildgebung nach Strahlentherapie oder Chemotherapie und Tumoren in Strukturen mit physiologisch erhöhtem Stoffwechsel [3]. Als Ursachen für falsch positive PET-Befunde im Kopf-Hals-Bereich werden Infektionen, Strukturen mit physiologischerweise erhöhtem Metabolismus wie Tonsillen und Speicheldrüsen, reaktive Lymphknoten, braunes Fett und nicht infektiöse Entzündungsreaktionen innerhalb von 4 bis 6 Monaten nach Operation (Flare-Phänomen) beschrieben. Auch in den Muskeln der Mundhöhle, des Halses und der Larynx sowie der Kaumuskulatur kann eine erhöhte FDG-Aufnahme auftreten [3,23,24]. Zahnfüllungen und -implantate mit Metall können die Bildqualität bei der PET/CT verschlechtern, die gemessenen FDG-Konzentrationen im Gewebe neben dem Metall beeinflussen und so zu falsch positiven Befunden führen [24].

Derzeit übliche Therapiepraxis

Therapieziele bei der Entwicklung von Therapieschemata für Kopf- und Halstumoren sind neben der Heilung der Erhalt der Funktion, eine Reduktion der Therapienebenwirkungen und ein kosmetisch zufriedenstellendes Ergebnis [4,6,8-10,17].

Die Behandlung von Kopf-Hals-Tumoren ist komplex und sollte bevorzugt von einem multidisziplinären Team durchgeführt werden [2,4,10,16]. Eine differenzierte Abhandlung der Behandlungsschemata sprengt den Rahmen dieser Übersicht, da sich die Behandlung je nach Tumorart, Grading, Stadium, Lokalisation und Resezierbarkeit unterscheiden kann. Es gibt

keine allgemein anerkannte Definition für die Resezierbarkeit. Diese hängt zum Teil von der Erfahrung des operativen Teams sowie dem Bestehen eines multidisziplinären Teams an einem Krankenhaus ab [8,9,16]. Es sollen daher hier nur einige Möglichkeiten umrissen werden.

Allgemeine Prinzipien der Behandlung

Lange Zeit wurden Kopf- und Halstumoren vornehmlich chirurgisch und / oder mit Strahlentherapie behandelt, aber zunehmend spielt auch die Chemotherapie eine Rolle bei der Therapie [6,8,10]. Monotherapie mit Chirurgie oder Strahlentherapie wird im Allgemeinen für die Patienten empfohlen, die sich in einem frühen Stadium (Stadium I oder II) präsentieren [16]. Die Empfehlungen der ESMO für das Plattenepithelkarzinom des Kopf-Hals-Bereiches sind gleich, wobei die Autoren einschränken, dass die Evidenz nur auf retrospektiven Studien basiert und randomisierte Studien fehlen [2].

Die Therapie beim CUP unterscheidet sich je nach Histologie und TNM-Stadium. Der CUP ist sehr oft kurierbar, obwohl der Primärtumor nur in wenigen Fällen im Verlauf identifiziert werden kann [16].

Neck Dissection

Eine häufige chirurgische Intervention bei Kopf- und Halstumoren – neben der Operation des Primärtumors – ist die Neck Dissection (Halsausträumung). Die Terminologie für die verschiedenen Formen der Neck Dissection ist vielfältig, auch wenn es Bemühungen gibt, diese zu vereinheitlichen [25].

Bei der radikalen Neck Dissection entfernt man alle lymphatischen und nicht lymphatischen Strukturen von der Mandibula zur Klavikula zwischen dem Platysma und der prävertebralen Faszie außer Arteria carotis, Nervus hypoglossus, Nervus lingualis, Nervus vagus, Nervus phrenicus und Plexus brachialis [26].

Bei der selektiven Neck Dissection werden Lymphknotengruppen entfernt, die sich entlang der Ausbreitungswege der jeweiligen Primärtumoren befinden. Sie sollte laut NCCN im Allgemeinen nur bei Patienten durchgeführt werden, die keinen Lymphknotenbefall haben. Aufgrund der Ergebnisse der selektiven Neck Dissection wird dann entschieden, welche Patienten adjuvante Strahlentherapie erhalten [16].

Unter einer umfassenden Neck Dissection wird in den NCCN-Leitlinien die Entfernung aller Lymphknotengruppen, die auch bei einer radikalen Neck Dissection entfernt worden wären, verstanden – unabhängig davon, ob andere Strukturen wie der Musculus sternocleidomastoideus, die Vena jugularis oder der Nervus accessorius Radix spinalis erhalten werden. Die Indikation hierfür unterscheidet sich bei den einzelnen Tumoren [16].

Für Patienten im Stadium N2 oder N3 (AJCC), die eine präoperative Bestrahlung erhalten haben und ein komplettes Ansprechen zeigen, besteht kein Konsens bezüglich der Durchführung einer Neck Dissection oder Beobachtung [16]. In den meisten Institutionen, die Mitglieder des NCCN sind, würden Patienten mit extrakapsulärer Tumorausdehnung und / oder keiner sicheren Resektion im Gesunden eine adjuvante Chemoradiotherapie nach der Operation erhalten [16].

Strahlentherapie

Die Strahlentherapie für Kopf- und Halstumoren ist sehr komplex und es gibt große Unterschiede in der Anwendung zwischen verschiedenen Ländern und auch zwischen unterschiedlichen Institutionen im gleichen Land [16]. Ein limitierender Faktor der Strahlentherapie im Kopf-Hals-Bereich ist, dass sie radiosensitive Strukturen wie z. B. Hirnstamm, Augen und Speicheldrüsen schädigen kann [17]. Diverse Nebenwirkungen wie z. B. Mukositis, subkutane Fibrose, Heiserkeit durch Ödem des Larynx, Dysphagie und Nekrose des Schläfenlappens sind bekannt [6,8,9,17,27]. Um diese Schäden zu limitieren, wurden Methoden wie die Brachytherapie, die Hyperfraktionierung oder die stereotaktische Radiochirurgie entwickelt [8-10,17,23]. Eine der neueren Entwicklungen, die vielversprechend zu sein scheinen, ist z. B. die IMRT (intensitätsmodulierte Radiotherapie) [8-10,16,17].

Neuere Therapien

Andere neuere und gezieltere Therapien, die derzeit evaluiert werden, sind zum Beispiel Antikörper gegen EGFR (epidermal growth factor receptor) oder gegen VEGF (vascular endothelial growth factor) [8-10].

Stand der Technologie

Die PET ist ein nicht invasives diagnostisches Verfahren, das geeignet ist, mittels kurzlebiger Radioisotope metabolische Prozesse abzubilden. Im Gegensatz zu den klassischen bildgebenden Verfahren, bei denen Informationen über die anatomische Struktur und die Lokalisation von Läsionen erhoben werden, kann die PET komplementär dazu Funktion und Metabolismus der Gewebe charakterisieren. Wegen ihres verstärkten Glukosemetabolismus wird bei Kopf- und Halstumoren in der Regel, aber nicht ausschließlich, das Glukoseanalogon FDG als Tracer eingesetzt [23]. Die durch den Zerfall des radioaktiven Tracers freigesetzten Positronen werden von einem PET-Scanner detektiert und in ein Bild umgewandelt. Die Befundung ergibt sich aus dem Vergleich zwischen der Aktivität eines bestimmten Areals und der Aktivität aus dem umliegenden „normalen“ Gewebe. Aufgrund der Unterschiede in der physiologischen Glukoseaufnahme in den verschiedenen Körperregionen kann die Zuordnung einer bestimmten Region als „erhöhter Uptake“ schwierig sein und erfordert erhebliche Expertise. Um bei der Interpretation eine gewisse Standardisierung zu erreichen, wird der Glukosemetabolismus semiquantitativ als Standardized Uptake Value (SUV) wiedergegeben.

Damit kann die morphologisch orientierte CT- und MRT-Bildgebung bei spezifischen Fragestellungen um weitere Informationen ergänzt werden. Allerdings ist ein positiver Befund nicht malignomspezifisch: vielmehr können sich auch andere Prozesse mit erhöhtem Glukoseumsatz, wie z. B. Infektionen, dahinter verbergen. PET-Vollringscanner werden im Allgemeinen als Technologie der Wahl angesehen [28,29].

Eine technische Weiterentwicklung stellt die PET/CT dar, eine Integration von PET und CT in einem Untersuchungsgerät. Hier wird der Patient in einem Untersuchungsgang durch die beiden Detektorringe von der CT und PET gefahren. Die entstehenden Bilder werden im Computer fusioniert, wobei die CT-Information meist in Graustufen abgebildet und von der PET-Information farbig überlagert wird. Mit dieser Methode wird die hohe Auflösung der CT (≤ 1 mm) mit den Stoffwechsellinformationen der PET verknüpft.

2 Ziele der Untersuchung

Die vorliegende Untersuchung verfolgt 2 Ziele, nämlich die Untersuchung des patientenrelevanten Nutzens sowie der diagnostischen und prognostischen Wertigkeit der PET bzw. PET/CT. Folgende Fragestellungen werden untersucht:

1. Bestimmung des Tumorstadiums von Kopf- und Halstumoren, einschließlich der Tumorausdehnung und der Unterstützung bei einer gezielten Biopsie.
 - a. Detektion von zervikalen Lymphknotenmetastasen
 - b. Detektion von Fernmetastasen
 - c. Detektion von synchronen Zweittumoren
2. Beurteilung des Behandlungsansprechens bei Kopf- und Halstumoren
 - a. Nach Therapie
 - b. Zuverlässigkeit der PET nach Strahlentherapie
3. Nachweis von Rezidiven (bei begründetem Verdacht) bei Kopf- und Halstumoren
4. Unbekannter Primärtumor, insbesondere Diagnostik bei metastatischer zervikaler Adenopathie unbekanntem Ursprungs

Da die Punkte 2 und 3 sehr ähnliche Konzepte behandeln, die sich im Wesentlichen durch den Zeitpunkt der Untersuchung unterscheiden, werden diese zusammen behandelt.

2.1 Ermittlung des patientenrelevanten Nutzens der PET bzw. PET/CT

In erster Linie geht es um die Beschreibung des patientenrelevanten Nutzens, den Ärzte und Patienten von der Bildgebung mit PET bzw. PET/CT bei Kopf- und Halstumoren bzw. bei unbekanntem Primärtumor, d. h. insbesondere von der Diagnostik bei metastatischer zervikaler Adenopathie unbekanntem Ursprungs, erwarten können.

Als Indikationen für den Einsatz bei Kopf-Hals-Tumoren werden a) die Bestimmung des Tumorstadiums („staging“) und b) das Ansprechen des Tumors auf die Behandlung („residual disease evaluation“ / „re-staging“) bzw. der Nachweis eines Rezidivs bei begründetem Verdacht betrachtet.

Darüber hinaus soll beim unbekanntem Primärtumor untersucht werden, ob sich aufgrund der Diagnostik mittels PET oder PET/CT im Vergleich zur konventionellen Diagnostik im Rahmen der Tumorsuche und des TNM-Stagings ein patientenrelevanter Nutzen ergibt.

Unter Nutzen werden hierbei Veränderungen verstanden, die für den Patienten fassbare Konsequenzen haben, wie z. B. die Auswirkung auf die Mortalität und Morbidität, die optimierte Auswahl von vorhandenen Behandlungsoptionen mit mehr oder weniger toxischen Nebenwirkungen und Veränderungen in der Lebensqualität.

1. Es werden diagnostisch-therapeutische Strategien, die eine PET bzw. PET/CT enthalten, mit diagnostisch-therapeutischen Strategien ohne eine PET oder PET/CT bzw. ohne Berücksichtigung der PET- oder PET/CT-Resultate verglichen.
2. Zusätzlich werden diagnostisch-therapeutische Strategien mit PET Strategien mit PET/CT gegenübergestellt.

Untersuchungen der konventionellen Diagnostik, die bei den Testvergleichen in diesem Bericht berücksichtigt werden, sind primär die Computertomographie, die MRT und die Endoskopie (z. B. Laryngoskopie, Panendoskopie usw.).

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

Läge ausreichend Evidenz zu Ziel 1 vor, wäre eine Nutzenbewertung im Prinzip alleine anhand der Beantwortung dieser Fragestellung möglich. Vorrecherchen haben jedoch gezeigt, dass zu diesem Ziel kaum vergleichende Primärstudien vorliegen. Sollte sich dieser Mangel an aussagekräftigen Primärstudien nach der systematischen Literaturrecherche bestätigen, werden als zweites Ziel auch die Ergebnisse von Diagnose- und Prognosestudien zur PET bzw. PET/CT zusammengefasst.

Bei der Bewertung der diagnostischen und prognostischen Wertigkeit der PET bzw. PET/CT geht es primär um die Frage, inwieweit sich durch den Einsatz der PET bzw. PET/CT

- a) die Zuordnung in die jeweilige Staginggruppe mit den unterschiedlichen prognostischen Konsequenzen,
- b) das Erkennen von Patienten mit Residuen nach Behandlungsabschluss bzw. die Diagnose von Rezidiven

verbessert. Es sollen nach Möglichkeit auch Aussagen über den besten Zeitpunkt für die PET bzw. PET/CT nach Therapie getroffen werden.

Ähnlich gelagert ist die Frage, ob sich mittels PET und PET/CT zuverlässigere prognostische Aussagen über die Wahrscheinlichkeit für ein Rezidiv machen lassen, als es mit den bisherigen konventionellen diagnostischen Verfahren möglich ist.

Dieses Ziel setzt den Vergleich mit einem sogenannten Referenzstandard voraus. Der Referenzstandard muss für eine adäquate Interpretation der Ergebnisse idealerweise

unabhängig von dem zu prüfenden Diagnoseverfahren erhoben werden, da es anderenfalls zu Verzerrungen der Güte- und / oder Variabilitätsschätzer kommen kann (sogenannter Verifikationsbias oder Inkorporationsbias) [30-34].

3 Projektablauf

3.1 Verlauf des Projekts

Der Gemeinsame Bundesausschuss (G-BA) hat mit Schreiben vom 21.12.2006 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Recherche, Darstellung und Bewertung des aktuellen medizinischen Wissensstandes zur Positronen-emissionstomographie (PET), auch als Integrationsgerät von PET und Computertomographie (PET/CT), bei 14 verschiedenen Krankheitsbildern bzw. insgesamt 21 Indikationen beauftragt. Der vorliegende Berichtsplan (vorläufige Version) bezieht sich auf den Teilauftrag zu den Indikationen bei Kopf-Hals-Tumoren und unbekanntem Primärtumor.

Das abschließende Gespräch zur Auftragskonkretisierung hierzu erfolgte mit der zuständigen Themengruppe im G-BA am 21.03.2007. Hierbei wurde u. a. vorgeschlagen, dass zu den Kopf- und Halstumoren im Allgemeinen Malignome im Bereich der Lippe, Mundhöhle, Pharynx, Larynx, Speicheldrüsen und des sinonasalen Bereiches zählen.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Zur Erstellung des Berichtsplans wurden am 29.01.2008 Patientenvertreter der folgenden Organisationen konsultiert:

- Ohne Schilddrüse leben e. V.
- Selbsthilfeverein der Kehlkopfoperierten

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

4 Methoden

In der Vergangenheit verband man mit dem Denken über die Bewertung von diagnostischen Tests primär die Beschreibung der technischen Merkmale des Tests und die Evaluation seiner diagnostischen Güte, d. h. seine Fähigkeit, kranke von gesunden Personen zu unterscheiden [35], in der impliziten Erwartung, dass sich eine bessere diagnostische Güte automatisch in eine bessere Patientenversorgung umsetzen würde. Seitdem hat sich das Denken über den Stellenwert diagnostischer Tests im therapeutischen Management weiterentwickelt, wobei neben der verbesserten diagnostischen Güte die Konsequenzen, die aus dem diagnostischen Test für den Arzt, den Patienten und die Gesellschaft resultieren, in das Zentrum der Betrachtung rücken [36,37]. Fryback und Thornbury haben dazu in ihrer Publikation „The efficacy of diagnostic imaging“ von 1991 [37] 6 Stufen identifiziert, die von den technischen Auswirkungen („technical efficacy“) als unterste Stufe über die diagnostische Genauigkeit („diagnostic accuracy efficacy“) als vorletzte Stufe bis hin zu den therapeutischen Auswirkungen („therapeutic efficacy“) als vierte Stufe und den patientenrelevanten Auswirkungen („patient outcome efficacy“) als fünfte Stufe reichen. Als oberste Stufe betrachten sie schließlich die gesellschaftlichen Konsequenzen („societal efficacy“) der diagnostischen Technologie (s. auch Tabelle 5). 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 diagnostischer Tests für Arzt, Patient und Gesellschaft in Betracht, wenn sie über den Wert eines diagnostischen Verfahrens nachdenkt. Auch der Gemeinsame Bundesausschuss hat sich in seiner Verfahrensordnung von 2006 bei der Bewertung diagnostischer Tests (§ 18 Abs. 7) dieser Sichtweise angeschlossen [38]. In Tabelle 5 sind verschiedene Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren einander gegenübergestellt. Neben der Einteilung von Fryback und Thornbury und der Evidenzklassifizierung des G-BA (s. Tabelle 5) wurde auch das Phasenschema nach Köbberling et al. [36] mit aufgenommen.

Tabelle 5: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren

	Köbberling et al. [36]	Fryback & Thornbury [37]	Evidenzklassifi- zierung G-BA [38]
Patientenrelevante Zielgrößen	Phase 4: Wirksamkeit	6. Stufe: Auswirkung auf Systemebene	Evidenzstufe I
		5. Stufe: Auswirkung im Hinblick auf patientenrelevante Endpunkte	
Surrogatzielgröße (Managementänderung)		4. Stufe: Auswirkung im Hinblick auf das therapeutische Denken des Behandelnden	Evidenzstufe II
		3. Stufe: Auswirkung im Hinblick auf das (differenzial)diagnostische Denken („ <i>Nachtest- wahrscheinlichkeit</i> “)	
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	

Tabelle 6: Evidenzklassifizierung des G-BA für Unterlagen zu diagnostischen Methoden [38]

Evidenzklassifizierung G-BA [38]	Kriterien
I a	Systematische Übersichtsarbeiten von Studien der Evidenzstufe I b
I b	Randomisierte kontrollierte Studien
I c	Andere Interventionsstudien
II a	Systematische Übersichtsarbeiten von Studien zur diagnostischen Testgenauigkeit der Evidenzstufe II b
II b	Querschnitt- und Kohortenstudien, aus denen sich alle diagnostischen Kenngrößen zur Testgenauigkeit (Sensitivität und Spezifität, Wahrscheinlichkeitsverhältnisse, positiver und negativer prädiktiver Wert) berechnen lassen
III	Andere Studien, aus denen sich die diagnostischen Kenngrößen zur Testgenauigkeit (Sensitivität und Spezifität, Wahrscheinlichkeitsverhältnisse) berechnen lassen
IV	Assoziationsbeobachtungen, pathophysiologische Überlegungen, deskriptive Darstellungen, Einzelfallberichte u. Ä.; nicht mit Studien belegte Meinungen anerkannter Experten, Berichte von Expertenkomitees und Konsensuskonferenzen

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

Um den patientenrelevanten Nutzen der bildgebenden diagnostischen Technologien PET bzw. PET/CT für die Behandlung von Patienten mit Kopf- und Halstumoren zu bewerten, genügt es nicht festzustellen, inwieweit die PET bzw. die PET/CT den gängigen diagnostischen Methoden hinsichtlich einer korrekten Identifizierung des Tumors (bzw. seiner Ausbreitung im Körper) überlegen ist. Vielmehr geht es bei einer erweiterten Bewertung darum festzustellen, inwieweit sich eine präzisere Diagnosestellung auf die Behandlung und das weitere Management des Patienten auswirkt und inwieweit, in letzter Konsequenz, das veränderte Management patientenrelevante Endpunkte wie z. B. Mortalität, Morbidität, Rezidivraten, Krankheitsprogression oder die Lebensqualität beeinflusst.

Diese Fragen lassen sich durch Managementstudien untersuchen, in denen die Patienten z. B. je nach PET-Befund randomisiert entweder der bisherigen, von der PET unabhängigen Standardtherapie oder einer an den PET-Befund adaptierten Therapie zugeteilt werden. Aber auch andere Studiendesigns kommen infrage [39]. Wie sich die Behandlungsoptionen (aggressiveres oder weniger aggressives Vorgehen, Chemo- oder Strahlentherapie oder beides) verändern und welche Auswirkungen die unterschiedlichen Strategien auf patientenrelevante Endpunkte wie Lebensqualität, Morbidität und Mortalität haben, lässt sich am besten durch Studien mit ausreichender Nachbeobachtungszeit beantworten (s. auch Abschnitt 4.2.1.3).

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

Vorrecherchen haben ergeben, dass vermutlich kaum qualitativ ausreichende systematische Übersichten zur Bewertung des patientenrelevanten Nutzens vorliegen. Daher soll der patientenrelevante Nutzen in dem zu erstellenden Bericht anhand von Primärstudien untersucht werden.

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

4.1.2.1 Population

Folgende Patientenpopulationen werden eingeschlossen:

- Patienten mit nachgewiesenen Kopf- und Halstumoren, die ein Staging für die Primärtherapie erhalten
- Patienten nach Therapie zur Erfassung des Therapieansprechens sowie Patienten mit begründetem Verdacht auf ein Rezidiv
- Patienten mit unbekanntem Primärtumor und insbesondere mit metastatischer zervikaler Adenopathie unbekanntem Ursprungs

Zu den Kopf- und Halstumoren zählen für diesen Bericht Malignome im Bereich der Lippe, Mundhöhle, Pharynx, Larynx, Speicheldrüsen und des sinonasalen Bereiches. Für diesen Bericht werden die Anteile der Lippe, die mit zur Mundhöhle gehören, berücksichtigt werden. Durch die anatomische Definition der Tumoren des Kopf-Hals-Bereiches gibt es Überlappungen mit anderen Tumorgruppen, die z. T. auch in anderen Berichten abgehandelt werden. Tumoren der Haut, Knochen, Lymphome sowie Tumoren der Weichteile im Kopf-Hals-Bereich sowie des Ösophagus und der Schilddrüse werden daher in diesem Bericht nur berücksichtigt, wenn sie als Primärtumoren im Rahmen einer CUP-Abklärung identifiziert werden.

Unbekannte Primärtumoren sind Tumoren, bei denen mittels konventioneller Diagnostik der Primärtumor nicht auffindig gemacht wurde. Es muss aus der Studie ersichtlich sein, welche Tests im Rahmen der konventionellen Diagnostik durchgeführt wurden.

4.1.2.2 Prüf- und Vergleichsintervention

Als Prüfindervention gilt eine diagnostisch-therapeutische Strategie mit Anwendung einer Vollring-PET bzw. -PET/CT. Als Vergleichsintervention wird eine diagnostisch-therapeutische Strategie ohne Anwendung der PET bzw. PET/CT betrachtet.

Dabei ist eine Vielzahl von Vergleichen möglich. Diese können die Randomisierung zu einer der beiden Strategien umfassen oder eine Vorschaltung der Diagnostik mit und ohne PET und mit konsekutiver Randomisierung zu verschiedenen Therapieoptionen beinhalten, wie z. B.:

- Prüfintervention: therapeutische Strategie basierend auf dem PET- oder PET/CT-Befund,
- Vergleichsintervention: therapeutische Strategie unabhängig vom PET- oder PET/CT-Befund

oder

- Prüfintervention: diagnostische Strategie mit Anwendung der PET bzw. PET/CT und konsekutiver Therapie
- Vergleichsintervention: diagnostische Strategie ohne Anwendung der PET bzw. PET/CT und konsekutiver Therapie

Weitere Vergleiche sind möglich [39]. Unabdingbar ist dabei, dass für ein bestimmtes Testergebnis die therapeutische Strategie vorab festgelegt ist. Im Rahmen dieser Nutzenstudien kann die PET/CT auch mit der PET verglichen werden.

Eine Restriktion auf eine bestimmte (erlaubte) Begleitdiagnostik bzw. auf eine erlaubte Verteilung der Begleitdiagnostik zwischen der Prüf- und Vergleichsintervention ist im Berichtsplan nicht geplant. Mögliche Unterschiede in den verwendeten diagnostischen Maßnahmen – außer bei der Anwendung der PET bzw. PET/CT – zwischen den zu vergleichenden Strategien werden bei der Interpretation der Ergebnisse aufgegriffen.

4.1.2.3 Zielgrößen

Für die Untersuchung werden folgende primäre Zielgrößen verwendet, die eine Beurteilung patientenrelevanter Therapieziele ermöglichen:

- Mortalität
- krankheitsfreies Überleben
- Morbidität, z. B. Auftreten von Rezidiven, Nebenwirkungen durch aggressivere / weniger aggressive Behandlung
- Funktionalität (z. B. Sprechen, Kauen), gesundheitsbezogene Lebensqualität

Als ergänzende Information sollen zudem Ergebnisse berichtet werden, die sich auf Änderungen im klinischen Management der Patienten (z. B. Wahl der Therapie, Größe des Bestrahlungsfeldes) beziehen, die a priori als primäre oder sekundäre Endpunkte definiert und systematisch dokumentiert wurden.

4.1.2.4 Studientypen

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

Finden sich keine bzw. nur eine geringe Zahl von relevanten RCTs, anhand derer der patientenrelevante Nutzen der PET bzw. PET/CT nicht mit ausreichender Ergebnissicherheit geschätzt werden kann (z. B. kleine Fallzahl, wenige Ereignisse [geringe Präzision], hohes Verzerrungspotenzial), werden gegebenenfalls auch verfügbare eindeutig prospektiv geplante, nicht randomisierte kontrollierte (Interventions-)Studien herangezogen, sofern das Problem einer möglichen Strukturungleichheit (unfairer Vergleich) adäquat in der Planung und Auswertung der entsprechenden Studien berücksichtigt wurde. Solche Studien können zwar die Aussage von Ergebnissen aus qualitativ belastbaren RCTs in der Regel nicht qualitativ ändern, diese aber ggf. substantiieren. 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 entfaltet. Weiterhin erlauben möglicherweise nicht randomisierte kontrollierte (Interventions-)Studien – wie oben skizziert – wertvolle Hinweise auf die Ausgestaltung aussagekräftiger(er) Studien.

4.1.2.5 Sonstige Studiencharakteristika

Es erfolgt keine weitere Einschränkung.

4.1.3 Ein- / Ausschlusskriterien

In die Nutzenbewertung werden alle Studien einbezogen, die alle nachfolgenden Einschlusskriterien und keines der nachfolgenden Ausschlusskriterien erfüllen.

Einschlusskriterien für Primärstudien zur Evaluierung des patientenrelevanten Nutzens	
E1a	Patienten mit Kopf- und Halstumoren oder unbekanntem Primärtumor, wie in Abschnitt 4.1.2.1 beschrieben
E2a	Vergleich von 2 Strategien wie unter 4.1.2.2 beschrieben
E3a	Patientenrelevante Endpunkte wie unter 4.1.2.3 beschrieben
E4a	Randomisierte kontrollierte Studien (bei fehlenden randomisiert-kontrollierten Studien werden ggf. prospektive nicht randomisierte kontrollierte Studien wie in Abschnitt 4.1.2.4 beschrieben eingeschlossen)
E5a	Sprachen: Englisch, Deutsch, Französisch, Spanisch

Ausschlusskriterien für Primärstudien zur Evaluierung des patientenrelevanten Nutzens	
A1a	Mehrfachpublikationen ohne relevante Zusatzinformation
A2a	Keine Vollpublikation verfügbar ^a
a: Als Vollpublikation gilt in diesem Zusammenhang auch die nicht vertrauliche Weitergabe eines Studienberichts an das Institut oder die nicht vertrauliche Bereitstellung eines Berichts über die Studien, der den Kriterien des CONSORT- [40] oder TREND-Statements [41] genügt und eine Bewertung der Studie ermöglicht.	

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

In den orientierenden Vorrecherchen fanden sich 3 systematische Übersichten, welche Teilaspekte der Fragestellungen zur diagnostischen Güte der PET oder PET/CT bei Kopf- und Halstumoren untersuchten. Um Redundanzen und Doppelarbeit zu vermeiden, soll die Bewertung der diagnostischen Güte der PET und PET/CT primär auf Grundlage von systematischen Übersichten erfolgen, sofern sichergestellt ist, dass die systematischen Übersichten bestimmte Mindestanforderungen erfüllen. Findet sich keine geeignete systematische Übersicht zu einer Fragestellung, wird die diagnostische und prognostische Wertigkeit anhand von Primärstudien beurteilt. Ist die systematische Übersicht nicht mehr aktuell (Suchzeitraum 6 Monate vor Berichterstellung und älter), wird für den Zeitraum, den die systematische Übersicht nicht abdeckt, relevante Primärliteratur herangezogen (Ergänzungsrecherche).

Nachfolgend werden zunächst die Kriterien für einen Einschluss von systematischen Übersichten in die geplante Bewertung aufgeführt. Anschließend erfolgt die Darstellung für (s. o.) einzuschließende Primärstudien.

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

4.2.1.1 Population

Patienten mit Kopf- und Halstumoren oder unbekanntem Primärtumor wie unter Abschnitt 4.1.2.1 beschrieben.

4.2.1.2 Prüftechnologie, Vergleichstechnologie („Komparator“) und Referenzstandard

Prüftechnologie ist die PET bzw. -PET/CT, als Vergleichstechnologie („Komparator“) dient die konventionelle Diagnostik. Außerdem wird der Vergleich PET vs. PET/CT zugelassen (s. Abschnitt 2.1). Prüf- und Vergleichstechnologie müssen zu einem sogenannten Referenzstandard in Beziehung gebracht worden sein (s. Abschnitt 2.2). In den einzuschließenden systematischen Übersichten und Meta-Analysen muss ein solches

Ordnungsprinzip erkennbar und die Ergebnisse aus den Einzelstudien entsprechend dargestellt sein. Als Referenzstandard für die diagnostische Güte gilt die Histologie, ggf. in Kombination mit einer Nachbeobachtung, für die prognostische Güte die Nachbeobachtung.

Konkordanzstudien müssen so konzipiert sein, dass alle diskordanten Befunde analog zu den Studien zur diagnostischen und prognostischen Güte mit dem gleichen, validen Referenzstandard untersucht wurden (s. Abschnitt 4.2.1.4).

Zur Bestimmung des besten Zeitpunktes für eine PET nach Strahlentherapie werden Studien herangezogen, welche die diagnostische Wertigkeit der PET zu unterschiedlichen Zeitpunkten nach der Therapie bestimmen und direkt vergleichen.

4.2.1.3 Zielgrößen

- Beschreibung der diagnostischen Güte mittels Sensitivität / Spezifität (ggf. ROC-Kurven) und daraus abgeleitet ggf. prädiktive Aussagen z. B. mittels positivem / negativem prädiktivem Wert. Dabei muss die Information zur diagnostischen Güte aus jeder Primärstudie zur Verfügung stehen. Die diagnostische Güte der Prüf- und Vergleichstechnologie zur Beurteilung des Lymphknotenstadiums muss pro Patient erhoben worden sein. Zusätzliche Angaben, bezogen auf die befallene Halsseite und die Lymphknotenlevels (entsprechend den Levels I bis VI der Neck Dissection [25]), werden ebenfalls extrahiert und ausgewertet. Die Bestimmung der patientenbasierten diagnostischen Güte birgt dabei das geringste Verzerrungspotenzial in sich.
- Prognostische Aussagen bzgl. des Überlebens (z. B. gesamt, progressionsfrei) anhand geeigneter statistischer Methoden (z. B. Kaplan-Meier-Schätzungen) und multifaktorieller Analysen (z. B. Cox-Regression) zur Adjustierung von Confounding. Für eine adäquate Beurteilung der prognostischen Güte der PET bzw. PET/CT sollten in der eingeschlossenen Literatur mögliche Confounder (wie z. B. unterschiedliche Histologie, unterschiedliche Therapieschemata usw.) untersucht werden, die zu verzerrten Einschätzungen der prognostischen Güte führen können. Studien ohne multifaktorielle Analysen werden jedoch nicht ausgeschlossen. Für die Interpretation des Überlebens ist ein durchschnittliches Follow-up von mindestens 12 Monaten entscheidend.
- Aussagen zur Konkordanz / Diskordanz unterschiedlicher diagnostischer Verfahren (s. Abschnitt 4.2.1.4)

4.2.1.4 Den systematischen Übersichten zugrunde liegende Studientypen

- Querschnittstudien, ggf. mit klinischer Nachbeobachtung, durchgeführt in der Anwendungssituation (gemäß Phase 3 nach Köbberling et al. [36])
- Prospektive und retrospektive Kohortenstudien

- Ist die Datenlage aus Studien mit einem direkten Vergleich der alternativen diagnostischen Vorgehensweisen mit dem Referenzstandard unzureichend, können systematische Übersichten von Querschnittstudien bzw. prospektiven oder retrospektiven Kohortenstudien herangezogen werden, die lediglich die diskordanten Befunde zwischen 2 Technologien untersuchen.

4.2.1.5 Sonstige Charakteristika

- Es soll sich um Publikationen handeln, die auf systematische, reproduzierbare und transparente Weise Ergebnisse von Primärstudien zusammenfassend dargestellt und bewertet haben. Dazu müssen die Publikationen auf dem Qualitätsindex von Oxman und Guyatt mindestens 5 von 7 möglichen Punkten erreichen [42,43].
- Die Forschungsfrage wurde im Format einer sog. PICO-Frage formuliert, die Population, Indextest, Referenzstandard und Endpunkte explizit benennt und nur wenige, spezifisch benannte Aspekte untersucht.
- Adressierung eines möglicherweise in den Primärstudien zu konstatierenden Verifikations- bzw. Inkorporationsbias
- Die Ergebnisse der Qualitätsbewertung, die Kerninformationen über die Population und die Outcomes einer jeden Studie sind in tabellarischer Form, transparent und vergleichbar abgebildet und die Daten können eindeutig mit den zugehörigen Studien verknüpft werden.

4.2.1.6 Ein- / Ausschlusskriterien

Es werden alle Studien einbezogen, die alle nachfolgenden Einschlusskriterien und keines der nachfolgenden Ausschlusskriterien erfüllen.

Einschlusskriterien zur Evaluierung des diagnostischen und prognostischen Wertes anhand von systematischen Übersichten	
E1b	Wie E1a (s. Abschnitt 4.1.2.1)
E2b	Prüftechnologie: PET bzw. PET/CT, Evaluierung wie in Abschnitt 4.2.1.2 beschrieben
E3b	Zielgrößen wie in Abschnitt 4.2.1.3 beschrieben
E4b	Systematische Übersichten aus Querschnitt- oder Kohortenstudien wie in den Abschnitten 4.2.1.4 und 4.2.1.5 beschrieben
E5b	Sprachen: Englisch, Deutsch, Französisch, Spanisch

Ausschlusskriterien zur Evaluierung des diagnostischen und prognostischen Wertes anhand von systematischen Übersichten	
A1b	Mehrfachpublikationen ohne relevante Zusatzinformation
A2b	Keine Volltextpublikation 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 QUOROM-Statements [44] genügt und eine Bewertung der Studie ermöglicht.	

4.2.2 Einschlusskriterien zur Ermittlung der diagnostischen und prognostischen Wertigkeit anhand von Primärstudien

4.2.2.1 Population

Patienten mit Kopf- und Halstumoren oder unbekanntem Primärtumor, wie in Abschnitt 4.1.2.1 beschrieben.

4.2.2.2 Prüftechnologie, Vergleichstechnologie („Komparator“) und Referenzstandard

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

4.2.2.3 Zielgrößen

- Beschreibung der diagnostischen Genauigkeit mittels Sensitivität / Spezifität (ggf. ROC-Kurven) und daraus abgeleitet ggf. prädiktive Aussagen z. B. mittels positivem / negativem prädiktivem Wert. Dabei muss die vollständige Darstellung der Information aus Vierfeldertafeln gewährleistet sein. Es gilt als nicht ausreichend, wenn z. B. allein Werte für die Sensitivität angegeben werden. Die diagnostische Güte der Prüf- und Vergleichstechnologie zur Beurteilung des Lymphknotenstadiums muss pro Patient, erhoben worden sein. Zusätzliche Angaben, bezogen auf die befallene Halsseite und die Lymphknotenlevels (entsprechend den Levels I bis VI der Neck Dissection [25]), werden ebenfalls extrahiert und ausgewertet. Die Bestimmung der patientenbasierten diagnostischen Güte birgt dabei das geringste Verzerrungspotenzial in sich.
- Prognostische Aussagen bzgl. des Überlebens (z. B. gesamt, progressionsfrei) anhand geeigneter statistischer Maßzahlen (z. B. Kaplan-Meier-Schätzungen) und multifaktorieller Analysen (z. B. Cox-Regression) zur Adjustierung von Confounding. Für eine adäquate Beurteilung der prognostischen Güte der PET bzw. PET/CT sollten in der eingeschlossenen Literatur mögliche Confounder (wie z. B. unterschiedliche Histologie, unterschiedliche Therapieschemata usw.) untersucht werden, die zu falschen Einschätzungen der prognostischen Güte führen können. Studien ohne multifaktorielle

Analysen werden jedoch nicht ausgeschlossen. Für die Interpretation des Überlebens ist ein durchschnittliches Follow-up von mindestens 12 Monaten entscheidend.

- Aussagen zur Konkordanz / Diskordanz unterschiedlicher diagnostischer Verfahren (s. 4.2.2.4)
- Zur Bestimmung des besten Zeitpunktes für die PET nach Strahlentherapie werden Studien herangezogen, welche die diagnostische Wertigkeit der PET zu unterschiedlichen Zeitpunkten nach der Therapie bestimmen und direkt vergleichen.

4.2.2.4 Studientypen

- Querschnittstudien, ggf. mit klinischer Nachbeobachtung, durchgeführt in der Anwendungssituation (gemäß Phase 3 nach Köbberling et al. [36])
- Prospektive Kohortenstudien
- Ist die Datenlage aus Studien mit einem direkten Vergleich der alternativen diagnostischen Vorgehensweisen mit dem Referenzstandard unzureichend, können Querschnittstudien und prospektive Kohortenstudien herangezogen werden, die lediglich die diskordanten Befunde zwischen 2 Technologien untersuchen.

4.2.2.5 Sonstige Charakteristika

Es erfolgt keine weitere Einschränkung.

4.2.2.6 Ein- / Ausschlusskriterien

Es werden alle Studien einbezogen, die alle nachfolgenden Einschlusskriterien und keines der nachfolgenden Ausschlusskriterien erfüllen.

Einschlusskriterien für Primärstudien zur Evaluierung des diagnostischen und prognostischen Wertes	
E1c	Patienten mit Kopf- und Halstumoren oder unbekanntem Primärtumor wie in Abschnitt 4.1.2.1 beschrieben
E2c	Prüftechnologie: PET bzw. PET/CT, Evaluierung wie in Abschnitt 4.1.2.2 beschrieben
E3c	Zielgrößen wie in Abschnitt 4.2.2.3 beschrieben
E4c	Querschnitt- oder prospektive Kohortenstudien wie in Abschnitt 4.2.2.4 beschrieben
E5c	Sprachen: Englisch, Deutsch, Französisch, Spanisch

Ausschlusskriterien für Primärstudien zur Evaluierung des diagnostischen und prognostischen Wertes
A1c Mehrfachpublikationen ohne relevante Zusatzinformation
A2c Keine Volltextpublikation verfügbar ^a
a: Als Vollpublikation gilt in diesem Zusammenhang auch die nicht vertrauliche Weitergabe eines Studienberichts an das Institut oder die nicht vertrauliche Bereitstellung eines Berichts über die Studien, der den Kriterien des STARD-Statements [45] oder TREND-Statements [41] genügt und eine Bewertung der Studie ermöglicht.

4.3 Informationsbeschaffung

4.3.1 Bibliografische Literaturrecherche

Die systematische Suche nach relevanten Studien soll in folgenden Quellen durchgeführt werden:

- Suche nach Primärstudien in den bibliografischen Datenbanken MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials (Clinical Trials)
- Suche nach relevanten Sekundärpublikationen: Identifizierung mittels Suche in den Datenbanken MEDLINE und EMBASE parallel zur Suche nach relevanter Primärliteratur sowie mittels Suche in den Datenbanken Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE) und Health Technology Assessment Database (HTA)

4.3.2 Suche nach weiteren publizierten und nicht publizierten Studien

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

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

Die Literaturrecherche wurde aufgrund der Erfahrungen aus dem Projekt zur PET bei Lymphomen und der Studie über Suchstrategien für HTA-Berichte adaptiert [46]. Finden sich

relevante HTA-Berichte, welche die Einschlusskriterien für systematische Übersichten erfüllen, werden die jeweiligen HTA-Organisationen angeschrieben, um nach eventuellen Updates, welche noch nicht in den Datenbanken sind, zu fragen.

4.3.3 Selektion relevanter Studien

Die Selektion relevanter Publikationen bzw. Studien aus den in der bibliografischen Literaturrecherche identifizierten Literaturzitationen wird auf Basis des Titels und, sofern vorhanden, Abstracts von 2 Gutachtern unabhängig voneinander vorgenommen. Es werden in einem 1. Selektionsschritt alle Referenzen von sicher nicht relevanten Dokumenten aussortiert. Alle verbleibenden Dokumente werden in einem 2. Selektionsschritt im Volltext gesichtet und abschließend als relevante bzw. nicht relevante Dokumente eingestuft. Diskrepanzen im Selektionsprozess zwischen den Gutachtern werden durch Diskussion aufgelöst.

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

Die Bewertung der Informationen der eingeschlossenen Studien hängt stark von den verfügbaren Angaben und der Qualität der jeweiligen Publikationen und weiteren Informationsquellen ab. Alle für die Nutzenbewertung relevanten, mit validierten Messinstrumenten erhobene Ergebnisse werden hinsichtlich ihrer Ergebnissicherheit, bestehend aus dem Verzerrungspotenzial und der Präzision 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
- 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 davon ausgegangen werden kann, dass die Ergebnisse nicht relevant verzerrt sind. Bei einem hohen Verzerrungspotenzial ist eine relevante Verzerrung nicht ausgeschlossen. Unter einer relevanten Verzerrung ist zu verstehen, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

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

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

4.4.2 Bewertung von systematischen Übersichten zur diagnostischen und prognostischen Wertigkeit

Die Bewertung der allgemeinen Qualität einer systematischen Übersicht erfolgt anhand des validierten Qualitätsindex für systematische Übersichten von Oxman und Guyatt (s. Abschnitt 4.2.1.6). Dieser kann bei Bedarf an die Fragestellungen angepasst werden, z. B. zum Umgang mit Verifikations- bzw. Inkorporationsbias, zur Berücksichtigung der Qualität von Primärstudien in den Schlussfolgerungen, oder auch zur Frage nach Qualität der Suchstrategie und / oder Aktualität.

4.4.3 Bewertung von Primärstudien zur diagnostischen und prognostischen Wertigkeit

Die Qualitätsbewertung der Primärstudien zur diagnostischen Güte wird mithilfe des QUADAS-Instruments erfolgen [32]. Dieses kann bei Bedarf an die Fragestellungen angepasst werden (z. B. Konkordanzstudien).

Die 14 QUADAS-Kriterien beinhalten Fragen zum Verzerrungspotenzial, zu der Übertragbarkeit und der Publikationsqualität der Studie. Die meisten QUADAS-Kriterien beziehen sich auf das Verzerrungspotenzial (Frage 3, 4, 5, 6, 7, 10, 11, 12 und 14), 2 Fragen betreffen die Übertragbarkeit (Frage 1 und 2) und 3 Fragen beurteilen die Qualität der Publikation (Frage 8, 9, 13).

- Für die Qualitätsbewertung von prognostischen Studien wird zusätzlich zu QUADAS die adäquate Berücksichtigung von möglichen Confoundern ein weiteres Qualitätskriterium sein [47].

Die Gesamtbewertung der Güte der Evidenz umfasst folgende 4 Kategorien:

- Hohe bis sehr hohe Güte der Evidenz
- Mäßig hohe Güte der Evidenz
- Niedrige Güte der Evidenz
- Sehr niedrige Güte der Evidenz

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 eingesetzt. Eine abschließende zusammenfassende Bewertung der Informationen erfolgt darüber hinaus in jedem Fall.

4.5.1 Gegenüberstellung der Ergebnisse

Die Ergebnisse zu den in den Studien berichteten patientenrelevanten Endpunkten werden im Bericht vergleichend beschrieben.

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 Wertigkeit einbezogen. Dies trifft insbesondere zu, wenn viele Patienten nicht in der Auswertung enthalten sind. Ergebnisse fließen i. d. R. nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patienten basieren, d. h. wenn der Anteil der fehlenden Werte größer als 30 % ist. In der Literatur werden zum Teil bereits Ausfallraten größer als 20 Prozentpunkte als nicht mehr

aussagekräftig betrachtet [48]. Die Festlegung der Grenze auf 30 % kann daher als vergleichsweise liberal angesehen werden.

Ausnahmen von dieser Regel können z. B. dann gemacht werden, wenn aus logistischen Gründen für ganze Zentren (ganze Randomisierungsblöcke) keine Daten erhoben wurden und dies bereits bei der Studienplanung vorgesehen war [49]. Das beschriebene Vorgehen wird auch angewendet, wenn der Unterschied der fehlenden Werte zwischen den Gruppen größer als 15 Prozentpunkte ist.

4.5.1.1 Charakterisierung und Darstellung der Ergebnisse von einzelnen (Primär-) Studien zur Bewertung des patientenrelevanten Nutzens

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, Vorselektion) dargestellt. Die Studienpopulationen werden weiterhin durch demografische Daten (Alter, Geschlecht), durch Charakteristika der betrachteten Krankheitsentitäten, durch prognostische Angaben und durch die Zahl der Studienabbrecher beschrieben. Die Studienpopulationen in den einzelnen Studien sowie die Interventionsgruppen werden im Bericht 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 Zielgrößen (s. Abschnitt 4.1.2.3) zugeordnet. Endpunkte aus Studien, die keiner vordefinierten Zielgröße gemäß Abschnitt 4.1.2.3 zugeordnet werden können, sind als nicht patientenrelevant anzusehen und werden daher in der Nutzenbewertung nicht berücksichtigt.

4.5.1.2 Charakterisierung und Darstellung der Ergebnisse von systematischen Übersichten zur diagnostischen und prognostischen Wertigkeit

Bei der Bewertung der diagnostischen und prognostischen Wertigkeit werden die Ergebnisse der systematischen Übersichten unter den jeweiligen Fragestellungen des IQWiG-Berichtes abgebildet. Die in den systematischen Übersichten dargestellten Ergebnisse aus den zugrunde liegenden Primärstudien werden dabei – soweit notwendig und möglich – zunächst nach den Fragestellungen und Ordnungskriterien „Prüftechnologie“, „Komparator“ und „Referenzstandard“ klassifiziert und qualitativ zusammengefasst. Weiterhin 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 [50].

Sollten nach Abschluss des Recherchezeitraums der eingeschlossenen systematischen Übersicht noch weitere relevante Primärstudien publiziert worden sein, so werden diese separat und analog zu den (Primär-)Studien zur Bewertung des patientenrelevanten Nutzens präsentiert (s. Abschnitt 4.5.1.3).

4.5.1.3 Charakterisierung und Darstellung der Ergebnisse von Primärstudien zur diagnostischen und prognostischen Wertigkeit

Im Bericht werden die Studien anhand des Designs (Studiendesign, Studiendauer, Ort und Zeitraum der Studiendurchführung, relevante Endpunkte) beschrieben. Die Studienpopulationen werden außerdem durch demografische Daten (Alter, Geschlecht), durch Charakteristika der betrachteten Krankheitsentitäten, durch prognostische Angaben und durch die Zahl der Studienabbrecher beschrieben.

Bei der Bewertung der diagnostischen und prognostischen Wertigkeit werden die Ergebnisse der Primärstudien unter den jeweiligen Fragestellungen abgebildet. Die Ergebnisse werden dabei zunächst nach den Fragestellungen und Ordnungskriterien „Prüftechnologie“, „Komparator“ und „Referenzstandard“ klassifiziert und qualitativ zusammengefasst. Bei prognostischen Primärstudien werden – soweit untersucht – multifaktorielle Analysen und der Einfluss von möglichen Confoundern auf die prognostische Wertigkeit präsentiert.

4.5.2 Meta-Analysen

Sofern die Primärstudien hinsichtlich der Fragestellung und relevanter Charakteristika vergleichbar sind, werden die Einzelergebnisse mithilfe von Meta-Analysen quantitativ zusammengefasst. Für die statistische Auswertung werden primär die Ergebnisse aus Intention-to-Treat-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet. Die Meta-Analysen erfolgen i. d. R. auf Basis von Modellen mit zufälligen Effekten [51]. In begründeten Ausnahmefällen werden Modelle mit festen Effekten eingesetzt. Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, werden diese nach Möglichkeit anhand der vorhandenen Informationen eigenständig berechnet bzw. näherungsweise bestimmt.

Bei Studien zum patientenrelevanten Nutzen wird für kontinuierliche Variablen die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges g , als Effektmaß eingesetzt. Bei binären Variablen werden Meta-Analysen primär anhand des Odds Ratios durchgeführt. Bei Diagnostikstudien wird für Testgütekriterien die Meta-Analyse anhand der Sensitivität und Spezifität in einer bivariaten Analyse durchgeführt [52].

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 [53]. Ist die Heterogenität der Studienergebnisse nicht

bedeutsam ($p > 0,2$ für Heterogenitätstest), wird der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt. Bei bedeutsamer Heterogenität werden die Ergebnisse nur in begründeten Ausnahmefällen gepoolt. Außerdem wird untersucht, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt 4.5.3) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.5.4).

4.5.3 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, z. B.:

Für Studien zur Nutzenbewertung der PET bzw. PET/CT:

- Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ oder „niedrig“

Für Studien zur diagnostischen und prognostischen Wertigkeit:

- Verblindete Beurteilung der Prüf- und Vergleichstechnologie vs. unverblindete Beurteilung
- Konsekutive vs. nicht konsekutive Patientenrekrutierung
- Diagnostische Wertigkeit, basierend auf Patienten vs. auf Halsseite vs. auf Lymphknotenlevel (s. Abschnitt 4.2.2.3)

Das Ergebnis der Sensitivitätsanalysen kann die Sicherheit der aus den beobachteten Effekten abgeleiteten Aussagen beeinflussen.

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 auf Effektmodifikation zu untersuchen:

- Alter (insbesondere Kinder)
- Geschlecht

Nur bei Kopf- und Halstumoren:

- PET bis zu 3 Monate nach Strahlentherapie vs. PET mehr als 3 Monate nach Strahlentherapie
- Stadium cN0 vs. cN+ vs. gemischt (cN0 und cN+)

Darüber hinaus sollen die diagnostische Wertigkeit und die Zuverlässigkeit der PET oder PET/CT nach Strahlentherapie (stärkere Entzündungsreaktion) mit der diagnostischen Wertigkeit nach Chemotherapie verglichen werden.

Nur beim unbekanntem Primärtumor:

- Metastatische zervikale Adenopathie unbekanntem Ursprungs vs. unbekannter Primärtumor mit Erstmanifestation an anderer Lokalisation
- Primärtumor mit konventioneller Diagnostik inkl. CT und / oder MRT nicht gefunden vs. Primärtumor mit konventioneller Diagnostik ohne CT oder MRT nicht gefunden
- Ganzkörper-PET bzw. -PET/CT vs. PET bzw. PET/CT im Kopf-Hals-Bereich

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

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

5 Literaturverzeichnis

1. Davies L, Welch HG. Epidemiology of head and neck cancer in the United States. *Otolaryngol Head Neck Surg* 2006; 135(3): 451-457.
2. Pivot X, Felip E. Squamous cell carcinoma of the head and neck: ESMO clinical Recommendations for diagnosis, treatment and follow-up. *Ann Oncol* 2008; 19(Suppl 2): ii79-ii80.
3. Kapoor V, Fukui MB, McCook BM. Role of 18FFDG PET/CT in the treatment of head and neck cancers: posttherapy evaluation and pitfalls. *AJR Am J Roentgenol* 2005; 184(2): 589-597.
4. Marur S, Forastiere AA. Head and neck cancer: changing epidemiology, diagnosis, and treatment. *Mayo Clin Proc* 2008; 83(4): 489-501.
5. Sturgis EM, Wei Q, Spitz MR. Descriptive epidemiology and risk factors for head and neck cancer. *Semin Oncol* 2004; 31(6): 726-733.
6. Kadmani D. Oral cancer. *Mayo Clin Proc* 2007; 82(7): 878-887.
7. Robert Koch-Institut. Krebs in Deutschland 2003-2004: Häufigkeiten und Trends. Berlin: RKI; 2008. URL: <http://www.ekr.med.uni-erlangen.de/GEKID/Doc/kid2008.pdf>.
8. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet* 2008; 371(9625): 1695-1709.
9. Choong N, Vokes E. Expanding role of the medical oncologist in the management of head and neck cancer. *CA Cancer J Clin* 2008; 58(1): 32-53.
10. Cagnetti DM, Weber RS, Lai SY. Head and neck cancer: an evolving treatment paradigm. *Cancer* 2008; 113(7 Suppl): 1911-1932.
11. Goldenberg D, Benoit NE, Begum S, Westra WH, Cohen Y, Koch WM et al. Epstein-Barr virus in head and neck cancer assessed by quantitative polymerase chain reaction. *Laryngoscope* 2004; 114(6): 1027-1031.
12. Kalogera-Fountzila A, Karanikolas D, Katodritis N, Samantas E, Sarafopoulos A, Ikonomidou I et al. Prognostic factors and significance of the revised 6th edition of the AJCC classification in patients with locally advanced nasopharyngeal carcinoma. *Strahlenther Onkol* 2006; 182(8): 458-466.
13. Chuang SC, Scelo G, Tonita JM, Tamaro S, Jonasson JG, Kliewer EV et al. Risk of second primary cancer among patients with head and neck cancers: a pooled analysis of 13 cancer registries. *Int J Cancer* 2008; 123(10): 2390-2396.
14. Lee YY, Wong KT, King AD, Ahuja AT. Imaging of salivary gland tumours. *Eur J Radiol* 2008; 66(3): 419-436.

15. Pentheroudakis G, Briasoulis E, Pavlidis N. Cancer of unknown primary site: missing primary or missing biology? *Oncologist* 2007; 12(4): 418-425.
16. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: head and neck cancers [Online]. 22.05.2008 [Zugriff am 28.01.2009]. URL: <http://www.nccn.org>.
17. Wei WI, Sham JS. Nasopharyngeal carcinoma. *Lancet* 2005; 365(9476): 2041-2054.
18. Leon X, Quer M, Orus C, Del Prado V, Lopez M. Distant metastases in head and neck cancer patients who achieved loco-regional control. *Head Neck* 2000; 22(7): 680-686.
19. American Cancer Society. Detailed guide: laryngeal and hypopharyngeal cancer; how are laryngeal and hypopharyngeal cancers staged? [Online]. 14.05.2007 [Zugriff am 18.12.2007]. URL: http://www.cancer.org/docroot/cric/content/cric_2_4_3x_how_is_laryngeal_and_hypopharyngeal_cancer_staged_23.asp?sitearea=cric.
20. Watkinson JC, Gaze MN, Wilson JA. *Stell and Maran's head and neck surgery*. Oxford: Butterworth-Heinemann; 2000.
21. American Cancer Society. Detailed guide: nasopharyngeal cancer; how is nasopharyngeal cancer staged? [Online]. 03.08.2006 [Zugriff am 18.12.2007]. URL: http://www.cancer.org/docroot/cric/content/cric_2_4_3x_how_is_nasopharyngeal_cancer_staged_17.asp?sitearea=cric.
22. Chan AT, Felip E. Nasopharyngeal cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol* 2008; 19(Suppl 2): ii81-ii82.
23. Newbold K, Partridge M, Cook G, Sohaib SA, Charles-Edwards E, Rhys-Evans P et al. Advanced imaging applied to radiotherapy planning in head and neck cancer: a clinical review. *Br J Radiol* 2006; 79(943): 554-561.
24. Rosenbaum SJ, Lind T, Antoch G, Bockisch A. False-positive FDG PET uptake: the role of PET/CT. *Eur Radiol* 2006; 16(5): 1054-1065.
25. Robbins KT, Shaha AR, Medina JE, Califano JA, Wolf GT, Ferlito A et al. Consensus statement on the classification and terminology of neck dissection. *Arch Otolaryngol Head Neck Surg* 2008; 134(5): 536-538.
26. Ferlito A, Rinaldo A. Paraneoplastic syndromes in patients with cancer of the larynx and hypopharynx. *Ann Otol Rhinol Laryngol* 2007; 116(7): 502-513.
27. Abendstein H, Nordgren M, Boysen M, Jannert M, Silander E, Ahlner-Elmqvist M et al. Quality of life and head and neck cancer: a 5 year prospective study. *Laryngoscope* 2005; 115(12): 2183-2192.
28. Gemeinsamer Bundesausschuss. Abschlussbericht des G-BA nach § 91 Abs. 7 SGB V Krankenhausbehandlung: Positronenemissionstomographie beim nichtkleinzelligen

Lungenkarzinom [Online]. 02.03.2006 [Zugriff am 28.04.2008]. URL: <http://www.g-ba.de/downloads/40-268-86/2005-12-20-khb-PET-NSCLC-Abschluss.pdf>.

29. Gemeinsamer Bundesausschuss. Positronen-Emissionstomographie: zusammenfassender Bericht des Arbeitsausschusses "Ärztliche Behandlung" des Bundesausschusses der Ärzte und Krankenkassen über die Beratungen gemäss §135 Abs.1 SGB V [Online]. 23.05.2002 [Zugriff am 26.09.2007]. URL: <http://www.g-ba.de/downloads/40-268-238/2003-05-23-Abschluss-PET.pdf>.

30. Cheran SK, Herndon JE, Patz EF. Comparison of whole-body FDG-PET to bone scan for detection of bone metastases in patients with a new diagnosis of lung cancer. *Lung Cancer* 2004; 44(3): 317-325.

31. Kirby AM, George MN. The role of FDG PET in the management of lymphoma: practical guidelines. *Nucl Med Commun* 2007; 28(5): 355-357.

32. 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.

33. 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.

34. 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.

35. How to read clinical journals; II: to learn about a diagnostic test. *Can Med Assoc J* 1981; 124(6): 703-710.

36. Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem* 1990; 28(12): 873-879.

37. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2): 88-94.

38. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [Online]. 18.04.2006 [Zugriff am 22.08.2007]. URL: http://www.g-ba.de/downloads/62-492-83/VerfO_2006-04-18.pdf.

39. Lijmer JG, Bossuyt PMM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009; 62(4): 364-373.

40. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001; 134(8): 657-662.

41. 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.
42. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44(11): 1271-1278.
43. 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.
44. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; 354(9193): 1896-1900.
45. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138(1): 1-12.
46. Raatz H, Suter K, Simon M, Scheibler F, Hausner E, Lange S. Search strategies for retrieving HTA-reports [Online]. In: *EbM 2008: evidenzbasierte Primärversorgung und Pflege*; 9. Jahrestagung Deutsches Netzwerk Evidenzbasierte Medizin und Kongress der Deutschen Gesellschaft für Pflegewissenschaft; 22.-23.02.2008; Witten, Deutschland. 12.02.2008 [Zugriff am 22.04.2009]. URL: <http://www.egms.de/en/meetings/ebm2008/08ebm52.shtml>.
47. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; 144(6): 427-437.
48. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
49. 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.
50. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156(10): 1411-1416.
51. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
52. 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.
53. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.