

**Positronenemissionstomographie  
(PET und PET/CT) bei  
rezidivierendem kolorektalen  
Karzinom**

**Berichtsplan**

Auftrag D06-01C  
Version 1.0  
Stand: 17.12.2009

# Impressum

**Herausgeber:**

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

**Thema:**

Positronenemissionstomographie (PET und PET/CT) bei rezidivierendem kolorektalen Karzinom

**Auftraggeber:**

Gemeinsamer Bundesausschuss

**Datum des Auftrags:**

21.12.2006

**Interne Auftragsnummer:**

D06-01C

**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

**Schlagwörter:** Positronenemissionstomographie, Kolonkarzinom, Rektumkarzinom, Rezidiv, Staging, systematische Übersicht

# Inhaltsverzeichnis

	Seite
<b>Tabellenverzeichnis</b> .....	<b>v</b>
<b>Abkürzungsverzeichnis</b> .....	<b>vi</b>
<b>1 Hintergrund</b> .....	<b>1</b>
<b>1.1 Definition des Krankheitsbildes</b> .....	<b>1</b>
<b>1.2 Epidemiologie und Krankheitslast</b> .....	<b>1</b>
<b>1.3 Ursache der Erkrankung</b> .....	<b>1</b>
<b>1.4 Verlauf der Erkrankung (natürlicher Verlauf)</b> .....	<b>2</b>
<b>1.5 Klassifikationsschemata</b> .....	<b>2</b>
<b>1.6 Diagnostische Verfahren</b> .....	<b>4</b>
<b>1.7 Derzeit übliche Therapiepraxis</b> .....	<b>5</b>
<b>1.8 Positronenemissionstomographie</b> .....	<b>6</b>
<b>2 Ziele der Untersuchung</b> .....	<b>8</b>
<b>2.1 Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. PET/CT</b> .....	<b>8</b>
<b>2.2 Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT</b> .....	<b>9</b>
<b>3 Projektablauf</b> .....	<b>10</b>
<b>3.1 Verlauf des Projekts</b> .....	<b>10</b>
<b>3.2 Dokumentation der Änderungen des Berichtsplans</b> .....	<b>10</b>
3.2.1 Inhaltliche Änderungen .....	10
3.2.2 Redaktionelle Änderungen .....	11
<b>4 Methoden</b> .....	<b>12</b>
<b>4.1 Konzept der Bewertung / Auswahl der Evidenzebene</b> .....	<b>14</b>
<b>4.2 Kriterien für den Einschluss von Studien und systematischen Übersichten bzw. HTA-Berichten in die Untersuchung</b> .....	<b>16</b>
4.2.1 Nutzenbewertung der PET bzw. der PET/CT anhand von Primärstudien .....	16
4.2.1.1 Population.....	16
4.2.1.2 Prüf- und Vergleichsintervention .....	16
4.2.1.3 Zielgrößen .....	16
4.2.1.4 Studientypen.....	17

4.2.1.5	Sonstige Charakteristika.....	18
4.2.1.6	Tabellarische Zusammenfassung der Einschlusskriterien.....	18
4.2.2	Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT anhand von systematischen Übersichten bzw. HTA-Berichten.....	19
4.2.2.1	Population.....	19
4.2.2.2	Prüftechnologie, ggf. Vergleichstechnologie und Referenzstandard.....	19
4.2.2.3	Zielgrößen .....	19
4.2.2.4	Den systematischen Übersichten bzw. HTA-Berichten zugrunde liegende Studientypen.....	20
4.2.2.5	Sonstige Charakteristika.....	20
4.2.2.6	Tabellarische Zusammenfassung der Einschlusskriterien.....	20
4.2.3	Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT anhand von Primärstudien.....	20
4.2.3.1	Population.....	20
4.2.3.2	Prüftechnologie, ggf. Vergleichstechnologie (Komparator) und Referenzstandard.....	21
4.2.3.3	Zielgrößen .....	21
4.2.3.4	Studientypen.....	21
4.2.3.5	Sonstige Charakteristika.....	22
4.2.3.6	Tabellarische Zusammenfassung der Einschlusskriterien.....	22
<b>4.3</b>	<b>Informationsbeschaffung.....</b>	<b>22</b>
4.3.1	Bibliografische Literaturrecherche.....	22
4.3.2	Suche nach weiteren publizierten und nicht publizierten relevanten Informationen.....	23
4.3.3	Selektion relevanter Studien und systematischer Übersichten bzw. HTA-Berichte .....	23
<b>4.4</b>	<b>Bewertung der einzelnen Studien .....</b>	<b>24</b>
4.4.1	Bewertung von Studien zur Nutzenbewertung.....	24
4.4.2	Bewertung von systematischen Übersichten und HTA-Berichten zur diagnostischen und prognostischen Güte .....	25
4.4.3	Bewertung von Primärstudien zur diagnostischen und prognostischen Güte .....	26
<b>4.5</b>	<b>Informationssynthese .....</b>	<b>26</b>
4.5.1	Nutzenbewertung der PET bzw. der PET/CT .....	26
4.5.1.1	Kriterien zum Einschluss in die Informationssynthese .....	26
4.5.1.2	Gegenüberstellung der Ergebnisse der Einzelstudien .....	27

4.5.1.3	Meta-Analysen .....	27
4.5.1.4	Sensitivitätsanalysen .....	28
4.5.1.5	Subgruppenmerkmale und andere Effektmodifikatoren .....	28
4.5.2	Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT.....	29
4.5.2.1	Kriterien zum Einschluss in die Informationssynthese .....	29
4.5.2.2	Gegenüberstellung der Ergebnisse der einzelnen systematischen Übersichten bzw. HTA-Berichte sowie der Primärstudien .....	29
4.5.2.3	Meta-Analysen .....	30
4.5.2.4	Sensitivitätsanalysen .....	30
4.5.2.5	Subgruppenmerkmale und andere Effektmodifikatoren .....	30
<b>5</b>	<b>Literaturverzeichnis.....</b>	<b>31</b>

## Tabellenverzeichnis

Tabelle 1: Inzidenz und Mortalität pro 100 000 Einwohner in Deutschland 2004 nach Geschlecht .....	1
Tabelle 2: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren .....	13
Tabelle 3: Evidenzklassifizierung des G-BA <sup>1</sup> für Unterlagen zu diagnostischen Methoden ..	14
Tabelle 4: Einschlusskriterien für Primärstudien zur Nutzenbewertung .....	18
Tabelle 5: Einschlusskriterien für systematische Übersichten bzw. HTA-Berichte zur Bewertung der diagnostischen und prognostischen Güte.....	20
Tabelle 6: Einschlusskriterien für Primärstudien zur Bewertung der diagnostischen und prognostischen Güte .....	22

**Abkürzungsverzeichnis**

<b>Abkürzung</b>	<b>Bedeutung</b>
CDRS	Cochrane Database of Systematic Reviews
CEA	Karzinoembryonales Antigen
CENTRAL	Cochrane Central Register of Controlled Trials
CONSORT	Consolidated Standards of Reporting Trials
CT	Computertomographie
DARE	Database of Abstracts of Reviews of Effects
EMBASE	Excerpta Medica Database
EMA	European Medicines Agency
FDG	Fluor-Deoxy-Glukose
G-BA	Gemeinsamer Bundesausschuss
HTA	Health Technology Assessment
ICD	International Statistical Classification of Diseases and Related Health Problems
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention-to-Treat
KI	Konfidenzintervall
KRK	Kolorektalkarzinom
MEDLINE	Medical Literature Analysis and Retrieval System Online
MRT	Magnetresonanztomographie
PET	Positronenemissionstomographie
pTNM	Histopathologische TNM-Klassifikation
RCT	Randomised controlled trial (randomisierte kontrollierte Studie)
rKRK	Rezidivierendes Kolorektalkarzinom
ROC	Grenzwertoptimierungskurve (engl.: receiver operating characteristic)
SUV	Standardized uptake value
TNM	Tumorstadieneinteilung nach Tumorausdehnung (T = Tumor), Lymphknotenbefall (N = Nodi lymphatici) und Metastasierung (M = Metastase)

## 1 Hintergrund

### 1.1 Definition des Krankheitsbildes

Unter einem rezidivierenden kolorektalen Karzinom (rK RK) wird ein erneutes Auftreten des im Bereich des Kolons bzw. Rektums entfernten Karzinoms, (kolorektales Karzinom, KRK), verstanden. KRK sind die häufigsten bösartigen Tumoren des Kolons und Rektums.

Ein Rezidiv des kolorektalen Karzinoms kann an der Ursprungslokalisierung bzw. von dieser entfernt als sog. Metastasen vorkommen. Metastasen können der Primärlokalisation des Tumors entsprechende Lymphknoten oder weiter entfernte Organe wie z. B. Lunge oder Leber (Fernmetastasen) befallen.

### 1.2 Epidemiologie und Krankheitslast

Das kolorektales Karzinom (KRK) ist mit über 70 000 Neuerkrankungen und ca. 30 000 Todesfällen pro Jahr in Deutschland einer der häufigsten malignen Tumoren [1]. Bei der ICD-10 wird das KRK unter bösartigen Neubildungen des Darms geführt: Kolon (C18), Übergang (C19), Rektum (C20).

Darmkrebs ist bei beiden Geschlechtern die zweithäufigste Krebserkrankung. Die Zahl der jährlichen Neuerkrankungen in Deutschland wird 2004 für Männer auf über 37 000 und für Frauen auf etwa 36 000 geschätzt. Männer erkranken im Mittel mit 69, Frauen mit 75 Jahren. Darmkrebs ist darüber hinaus sowohl für Frauen als auch für Männer die zweithäufigste Krebstodesursache [2]. Die geschätzten Erkrankungsraten von Männern und Frauen bleiben nach einem seit 1980 zunehmenden Trend in den letzten 10 Jahren auf unterschiedlichem Niveau nahezu unverändert. Im Gegensatz zur Entwicklung der Inzidenz nehmen die Sterberaten für Darmkrebs für beide Geschlechter stetig ab [2].

Die Inzidenz und die Mortalität in Deutschland sind in Tabelle 1 aufgeführt.

Tabelle 1: Inzidenz und Mortalität pro 100 000 Einwohner in Deutschland 2004 nach Geschlecht

Erkrankung	Männer		Frauen	
	Inzidenz	Mortalität	Inzidenz	Mortalität
Darmkrebs (ICD-10: C18-21)	92,3	34,1	85,4	33,3

nach Robert Koch-Institut [2]

### 1.3 Ursache der Erkrankung

Die genauen Ursachen der KRK sind nicht bekannt; es wird von Risikofaktoren gesprochen. Ernährungsbedingte Faktoren, insbesondere eine ballaststoffarme, fettreiche Nahrung, ein

hoher Anteil an rotem (eisenhaltigem) Fleisch, ein geringer Anteil an Gemüse sowie regelmäßiger Alkoholkonsum, erhöhen das Risiko, an Darmkrebs zu erkranken, ebenso wie Übergewicht und Bewegungsmangel. Verwandte ersten Grades von Patienten mit Darmkrebs sind selbst überdurchschnittlich häufig betroffen – ob aufgrund von Genveränderungen oder eines ähnlichen Lebensstils, ist bisher nicht abschließend geklärt. Bei der seltenen erblichen familiären adenomatösen Polyposis sowie dem erblichen nicht polypösen Kolorektalkarzinom treten mit hoher Wahrscheinlichkeit bereits in jüngeren Jahren Darmkrebsfälle in den betroffenen Familien auf. In geringerem Maße erhöhen auch chronisch entzündliche Darmerkrankungen wie beispielsweise eine Colitis ulcerosa das Krebsrisiko [2].

#### **1.4 Verlauf der Erkrankung (natürlicher Verlauf)**

Das KRK verläuft häufig lange symptomlos. Häufig kommt es erst zu Symptomen, wenn der Tumor in größerem Maße blutet oder wenn er das Darmlumen nennenswert einengt. Typische Symptome sind Blut oder Schleim im Stuhl, Darmkrämpfe sowie Durchfälle, Verstopfungen und Flatulenz. Bedingt durch einen Blutverlust kann es zu einer Anämie (Blutarmut) mit Symptomen wie Leistungsabfall, Müdigkeit und allgemeine Schwäche kommen. Darüber hinaus kommt es infolge der Tumorerkrankung zu einem Gewichtsverlust eventuell bis hin zur Auszehrung.

Die wichtigsten Komplikationen beim KRK sind Darmverschluss durch den Tumor und Durchbruch durch die Darmwand (Perforation) mit Entwicklung einer Bauchfellentzündung (Peritonitis). Diese Komplikationen sind Zeichen eines fortgeschrittenen Tumorleidens, können dennoch die ersten Symptome eines Darmkrebses sein. Fisteln, Einbruch in Nachbarorgane, Zeichen des Leberversagens aufgrund von Lebermetastasen sowie Kompression von anderen Organen gelten als weitere relevante Komplikationen.

#### **1.5 Klassifikationsschemata**

Das rezidivierende Kolorektalkarzinom kann analog dem Kolorektalkarzinom nach einem TNM-System klassifiziert werden. Das TNM-System ist eine Tumorstadieneinteilung (Staging) nach der klinisch und (soweit möglich) histopathologisch bestimmten anatomischen Ausdehnung eines Tumors aufgrund folgender Kriterien: Größe und Ausdehnung des Primärtumors (T); Fehlen bzw. Vorhandensein von regionalen Lymphknotenmetastasen (N von „Nodi lymphatici“); Fehlen bzw. Vorhandensein von Fernmetastasen (M) [3]:

T – Primärtumor (bezieht sich hier auf Lokalisation des rKRK):

- TX – Primärtumor kann nicht beurteilt werden
- T0 – kein Anhalt für Primärtumor
- Tis – Carcinoma / Tumor in situ
- T1 – Infiltration der Tela submucosa
- T2 – Infiltration der Tunica muscularis
- T3 – Infiltration der Subserosa

- T4 – Infiltration von Nachbarorganen oder des Bauchfells (Peritoneum viscerale).

N – Regionale Lymphknoten:

- NX – Regionale Lymphknoten können nicht beurteilt werden
- N0 – keine regionale Lymphknotenmetastasen
- N1 – Metastasen in 1 bis 3 perikolischen (perirektalen) Lymphknoten
- N2 – Metastasen in mehr als 3 perikolischen (perirektalen) Lymphknoten
- N3 – Metastasen entlang eines benannten Gefäßstamms und / oder apikale Lymphknotenmetastasen

M – Fernmetastasen:

- MX – Das Vorliegen von Fernmetastasen kann nicht beurteilt werden
- M0 – keine Fernmetastasen
- M1 – Fernmetastasen

Die Kategorie M1 kann je nach den betroffenen Geweben spezifiziert werden (meist Leber und Lymphknoten, gefolgt von Peritoneum, Lunge, seltener Skelett, Nebennieren oder Gehirn).

Die aufgrund klinischer Untersuchungen erstellte TNM-Klassifikation wird postoperativ durch die histopathologische pTNM-Klassifikation ergänzt (durch histopathologische Untersuchungen des Primärtumors, der regionalen Lymphknoten und ggf. von Fernmetastasen).

R – Residualtumor (postoperativ):

- R0 – kein Residualtumor
- R1 – mikroskopischer Residualtumor
- R2 – makroskopischer Residualtumor

G – Histopathologischer Differenzierungsgrad („Grading“):

- GX – Differenzierungsgrad kann nicht bestimmt werden
- G1 – gut differenziert
- G2 – mäßig differenziert
- G3 – schlecht differenziert
- G4 – undifferenziert

Anhand der TNM-Angaben kann eine weitere Stadieneinteilung vorgenommen werden. Diese basiert auf Studien zu prognostischen Aussagen. Die entsprechende Einstufung einer Tumorerkrankung erlaubt daher prognostische Aussagen und bestimmt häufig auch die weitere Therapie. Folgende Stadien werden unterschieden:

Ia: Beschränkung der Tumordinfiltration auf die Mukosa und Tela submucosa (Dukes A)

Ib: Beschränkung der Tumordinfiltration bis in die Tunica muscularispropria (Dukes A)

II: T3 oder T4 ohne Lymphknotenmetastasierung (Dukes B)

III: Lymphknotenmetastasierung (Dukes C)

IV: Fernmetastasen (Dukes D)

## 1.6 Diagnostische Verfahren

Der endgültige Nachweis eines rKRC erfolgt durch eine histologische Untersuchung des Gewebes, das durch eine Biopsie, Operation oder eine Obduktion entnommen wird.

Der Stellenwert der verschiedenen weiteren Verfahren zur Diagnostik eines Tumorrezidivs bei der Nachsorge nach der Entfernung des kolorektalen Karzinoms wird in der im Auftrag der Deutschen Gesellschaft für Verdauungs- und Stoffwechselkrankheiten und der Deutschen Krebsgesellschaft erstellten S3-Leitlinie „Kolorektales Karzinom“ [1] wie folgt definiert:

- Symptomorientierte Anamneseerhebung und körperliche Untersuchung sind grundsätzlicher Bestandteil der Nachsorge
- Die Bestimmung des karzinoembryonalen Antigens (CEA) wird alle 6 Monate für mindestens 2 Jahre empfohlen. Ein erhöhter CEA-Wert erfordert eine weitere Diagnostik, berechtigt aber nicht zum Beginn einer systemischen Chemotherapie bei Verdacht auf ein metastasiertes Tumorstadium.
- Die routinemäßige Bestimmung von Laborwerten im Rahmen der Nachsorge ist nicht zweckmäßig.
- Die Untersuchung auf okkultes Blut im Stuhl ist für die Nachsorge nicht geeignet.
- Die Sonographie ist zur Erkennung von Lebermetastasen technisch geeignet. Ihr Routineeinsatz ist aufgrund der Datenlage nicht gesichert. Die Expertenkommission bewertet die Sonographie als einfachstes und kostengünstigstes Verfahren und schlägt daher ihre Nutzung zur Diagnostik von Lebermetastasen vor.
- Die Endosonographie ist zur Erkennung von Lokalrezidiven beim Rektumkarzinom insbesondere in Kombination mit der Endosonographie-gesteuerten Biopsie geeignet. Für den routinemäßigen primären Einsatz in der Nachsorge kann derzeit keine Empfehlung gegeben werden.
- Die Computertomographie ist zur Erkennung von Lebermetastasen, von Lokalrezidiven im kleinen Becken sowie von Lungenmetastasen technisch geeignet. Die Datenlage spricht gegen einen routinemäßigen Einsatz der Computertomographie in der Nachsorge.
- Zum Stellenwert der Röntgenuntersuchung des Thorax in der Nachsorge war keine Konsensfindung möglich.

- Die Koloskopie ist geeignet, Lokalrezidive oder Zweittumoren zu erkennen. Alle Patienten sollten prä- oder innerhalb von 6 Monaten postoperativ eine komplette Koloskopie erhalten. Eine Koloskopie ist nach 3 Jahren und anschließend alle 5 Jahre erforderlich, um metachrone Karzinome oder Polypen zu erkennen.
- Die Sigmoidoskopie ist geeignet, Lokalrezidive und Zweittumoren im eingesehenen Bereich zu erkennen.
- Die starre Rektoskopie ist geeignet, beim Rektumkarzinom Lokalrezidive und Anastomosenveränderungen zu erkennen. Sie kann als alternatives Verfahren zur Sigmoidoskopie angewandt werden.
- Kolonkontrasteinlauf, die virtuelle Kolonographie und die Positronen-emissionstomographie (PET) haben in der programmierten Nachsorge keinen Stellenwert.

80 % der Rezidive treten in den ersten beiden Jahren nach der Operation eines KRK auf, nach 5 Jahren werden praktisch keine neuen Rezidive mehr festgestellt. Die Nachsorge nach der Tumorentfernung soll deshalb für den Zeitraum bis zum 5. postoperativen Jahr ausgedehnt werden. Es werden bei der Nachsorge Untersuchungsintervalle von 6 Monaten in den ersten 2 Jahren empfohlen. Über den Zeitpunkt 5 Jahre hinaus sollten lediglich Koloskopien zum Ausschluss eines Zweitkarzinoms erfolgen [1].

### **1.7 Derzeit übliche Therapiepraxis**

Die Therapie von Rezidiven des kolorektalen Karzinoms lässt sich entsprechend der S3-Leitlinie „Kolorektales Karzinom“ [1] wie folgt beschreiben:

Beim Rektumkarzinomrezidiv sollte, soweit noch keine Vorbehandlung (Radio[chemo]therapie) im Rahmen der Primäroperation erfolgte, diese nun primär favorisiert werden. Bei bereits vorbehandelten Patienten und der Möglichkeit einer R0-Resektion kann primär operiert werden. Bei bereits vorbehandelten Patienten und fraglicher R0-Resektibilität eines Rezidivs muss individuell und unter Berücksichtigung der Intensität der Vortherapie bezüglich eines multimodalen chirurgischen Vorgehens entschieden werden.

Beim lokoregionären Lymphknotenrezidiv des Kolonkarzinoms sollte – sofern möglich – der Versuch einer kurativen kompletten Resektion erfolgen.

Resektable Lungenmetastasen sollen reseziert werden. Bei primärer Irresektabilität soll eine systemische Chemotherapie erfolgen.

Auf die Leber beschränkte R0-resektable Metastasen sollen reseziert werden. Bei primär irresektablen Lebermetastasen soll eine systemische Therapie begonnen werden.

## 1.8 Positronenemissionstomographie

Die PET ist ein nicht invasives diagnostisches Verfahren, das die nach dem Zerfall positronenemittierender radioaktiver Substanzen und durch die anschließende Kollision von Positronen und Elektronen freigesetzten Photonen registriert, in ein Bild umwandelt und dadurch geeignet ist, metabolische Prozesse abzubilden. Das PET-Gerät enthält ringförmig um den Patienten angeordnete Photonendetektoren. Aus der räumlichen Verteilung der registrierten Photonen wird auf die räumliche Verteilung der radioaktiven Substanz im Körperinneren geschlossen.

Im Gegensatz zu den klassischen bildgebenden Verfahren, bei denen Informationen über die anatomische Struktur und die Lokalisation von Läsionen erhoben werden (sogenannte morphologische Bildgebung), kann die PET komplementär dazu die Funktion, den Metabolismus und die biochemischen Prozesse der Organe charakterisieren (sogenannte metabolische Bildgebung). Da eine bösartige Neoplasie, darunter auch ein rezidivierendes kolorektales Karzinom, im Vergleich zu gesundem Gewebe einen erhöhten Glukosemetabolismus aufweist, wird bei bösartigen Neoplasien in der Regel das Glukoseanalogon Fluor-Deoxy-Glukose (FDG) als Tracer (d. h. künstliche radioaktiv markierte Substanz, die nach Einbringung in den lebenden Körper am Stoffwechsel teilnimmt) eingesetzt.

In der Anfangszeit der PET wurden Nichtvollringsysteme hergestellt. Die Empfindlichkeit der später entwickelten Vollring-PET-Systeme war den ursprünglichen so stark überlegen, dass nur diese Systeme eine breite Anwendung fanden. Die ersten Vollring-PET-Systeme hatten nur einen Detektorring; bei modernen Systemen werden heute mehrere Ringe nebeneinander angeordnet, wodurch die Sensitivität des Gesamtsystems noch gesteigert wurde [4].

Die Befundung ergibt sich aus dem Vergleich zwischen der Aktivität eines bestimmten Areals und der Aktivität des umliegenden „normalen“ Gewebes. 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, kann der Glukosemetabolismus als *Standardized Uptake Value* (SUV) wiedergegeben werden. Allerdings ist ein positiver FDG-PET-Befund nicht malignomspezifisch; vielmehr können sich auch andere Prozesse mit erhöhtem Glukoseumsatz wie z. B. Infektionen dahinter verbergen.

Die Aktivitätsanreicherungen bei der PET lassen sich anatomisch jedoch nicht immer gut lokalisieren, da in PET-Bildern in erster Linie Stoffwechselprozesse aufgezeigt werden; hinzu kommt die begrenzte Ortsauflösung von etwa 4 bis 6 mm. Eine technische Weiterentwicklung stellt eine Kombination aus PET und CT in einem Untersuchungsgerät (PET/CT) dar. Hier wird der Patient in einem Untersuchungsgang durch die beiden Detektorringe 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 dann farbig

überlagert wird. Mit dieser Methode wird die hohe Auflösung der CT ( $< 1$  mm) mit den hoch sensitiven Stoffwechsellinformationen der PET verknüpft. Die kombinierten PET/CT-Geräte sind seit dem Jahr 2001 auf dem Markt und verdrängen seitdem reine PET-Scanner bei den Neuanschaffungen [5].

Da noch ältere, reine PET-Scanner im Einsatz sind, wird auch mittels moderner Software eine Überlagerung (sogenannte Softfusion bzw. Koregistrierung) von CT-Bildern und PET-Daten berechnet. Dies geschieht mithilfe gemeinsamer Referenzpunkte wie etwa Knochenstrukturen oder anhand externer Positionsdaten. Da das für die Bildfusion verwendete CT-Bild nicht zur Rekonstruktion der PET-Daten (eines reinen PET-Scanners) herangezogen wurde, ist die Bildqualität des resultierenden PET/CT-Bildes schlechter als die eines kombinierten PET/CTs-Geräts.

## 2 Ziele der Untersuchung

Das Ziel der vorliegenden Untersuchung ist die Bewertung des patientenrelevanten Nutzens und Schadens der PET bzw. der PET/CT beim rKRK. Ist dieses Ziel nicht erreichbar, erfolgt die Bewertung der diagnostischen und der prognostischen Güte der PET bzw. der PET/CT beim rKRK .

Unter rKRK wird nicht nur ein bereits diagnostiziertes rKRK, sondern auch ein begründeter Verdacht auf ein rKRK (an der Ursprungslokalisation bzw. Metastase) impliziert. Unter einem begründeten Verdacht auf rKRK wird dabei ein Hinweis auf ein rKRK durch zumindest eine diagnostische Methode verstanden, sofern weder ein PET-Scan noch eine histologische Untersuchung involviert ist.

Da heute nur die Vollring-PET-Systeme eine breite Anwendung finden, wird die Bewertung auf diese Systeme eingeschränkt. Die Bewertung erstreckt sich sowohl auf die reine PET als auch auf kombinierte PET/CT-Systeme.

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

In erster Linie geht es um die Bewertung des patientenrelevanten Nutzens und Schadens, der durch den Einsatz der PET bzw. PET/CT bei begründetem Verdacht auf ein rKRK erwartet werden kann.

Als Indikationen für den Einsatz der PET bzw. der PET/CT werden betrachtet:

- der Nachweis eines rKRK bei begründetem Verdacht
- die Bestimmung des Stadiums eines rKRK

Folgende Technologievergleiche werden betrachtet:

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

Unter „patientenrelevantem Nutzen und Schaden“ (im Weiteren kurz als „Nutzen und Schaden“ bezeichnet) werden Veränderungen verstanden, die für den Patienten fassbare Konsequenzen haben, wie z. B. die Auswirkung auf die Mortalität, Morbidität (darunter auch

Rate erneuter Rezidive und Lebensqualität). Mit dem Begriff „Nutzen“ werden kausal begründete positive Effekte, mit dem Begriff „Schaden“ kausal begründete negative Effekte einer medizinischen Intervention auf patientenrelevante Endpunkte (siehe 4.2.1.3) 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 (Methodenpapier des IQWiG, Version 3.0).

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

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

- Verbessert sich durch den Einsatz der PET bzw. der PET/CT bei Patienten mit begründetem Verdacht auf ein rKRR die Diagnostik (Nachweis bzw. Ausschluss) von Rezidiven?
- Verbessert sich durch den Einsatz der PET bzw. PET/CT die korrekte Zuordnung der Patienten mit rKRR zu den jeweiligen Krankheitsstadien?
- Verbessert sich durch den Einsatz der PET bzw. PET/CT bei Patienten mit rKRR die Vorhersage der patientenrelevanten Zielgrößen?

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

### **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 Bewertung der *Positronenemissionstomographie (PET und PET/CT) bei rezidivierendem kolorektalen Karzinom* beauftragt.

In die Bearbeitung des Projekts werden externe Sachverständige eingebunden.

Zur Erstellung des Berichtsplans wurden am 15.05.2009 Patientenvertreter konsultiert. Die vorgeschlagenen Endpunkte bezüglich der Nutzenbewertung finden sich in den unten aufgeführten Zielgrößen des Berichts wieder.

Der vorläufige Berichtsplan in der Version 1.0 vom 20.08.2009 wurde am 02.09.2009 auf der Website des IQWiG veröffentlicht und zur Anhörung gestellt. Bis zum 30.09.2009 konnten schriftliche Stellungnahmen eingereicht werden. Unklare Aspekte aus den schriftlichen Stellungnahmen zum vorläufigen Berichtsplan wurden am 22.10.2009 in einer wissenschaftlichen Erörterung mit den Stellungnehmenden diskutiert. Die Dokumentation und Würdigung der Anhörung zum Berichtsplan ist auf der Website des IQWiG veröffentlicht.

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

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

#### **3.2 Dokumentation der Änderungen des Berichtsplans**

Im Vergleich zum vorläufigen Berichtsplan Version 1.0 haben sich folgende Änderungen bzw. Ergänzungen ergeben:

##### **3.2.1 Inhaltliche Änderungen**

- Aufgrund eines Hinweises während der Erörterung wurde die Definition des Krankheitsbilds korrigiert, sodass bösartige Neubildungen des Anus und des Analkanals (ICD-Code C21) nicht zum Kolorektalkarzinom gerechnet werden (Abschnitt 2.1).

- Aufgrund eines Hinweises in einer Stellungnahme wird die Recherche um eine orientierende Leitlinienrecherche ergänzt, die systematische Übersichten als Grundlage von sog. S3-Leitlinien identifizieren soll.
- Im Abschnitt: 4.2.1.2 Prüf- und Vergleichsintervention (Nutzenbewertung anhand von Primärstudien) wurde für sogenannte Diskordanzstudien die beschränkende Forderung nach einer Vorabfestlegung der Therapiemaßnahme gestrichen.
- Für die Durchführung von Meta-Analysen zu Überlebenszeiten wurde das Effektmaß explizit als Hazard Ratio benannt (siehe Abschnitt 4.5.1.3).
- Als Konsequenz aus der Erörterung zum Aspekt der Inter-Observer-Variabilität ergaben sich zwar keine Änderungen der Berichtsmethodik, aber ggf. werden diesbezügliche Angaben in den eingeschlossenen Studien extrahiert und diskutiert.

### **3.2.2 Redaktionelle Änderungen**

- Im Abschnitt 4.2.1.3 wurden die primären Zielgrößen in Zielgrößen umbenannt, da in diesem Bericht keine sekundären Zielgrößen definiert wurden.
- Im Abschnitt 4.2.1.5 wurde die Forderung nach einer Mindeststudiendauer für die Nutzenbewertung der PET bzw. der PET/CT anhand von Primärstudien explizit verneint. Bisher fanden sich im Berichtsplan hierzu keine Angaben.
- Auf Anregung in einer Stellungnahme wird das Verzeichnis der ausgeschlossenen Literatur alphabetisch, ggf. mit zusätzlicher Nennung des Ausschlussgrunds im Rahmen der Volltextselektion, dargestellt.
- Im Abschnitt 4.2.1.3 Zielgrößen (Nutzenbewertung anhand von Primärstudien) wurden die Kriterien für die Berücksichtigung der Zielgröße Managementänderung klarer formuliert.

## 4 Methoden

Es werden mehrere Hierarchiestufen der Zielgrößen in Evaluierungsstudien zu diagnostischen Testverfahren genannt (siehe Tabelle 2). Patientenrelevante Zielgrößen werden als höchste Stufe des Klassifikationsschemas angesehen. Somit stehen die Konsequenzen, die aus dem diagnostischen Test für den Arzt, den Patienten und die Gesellschaft resultieren, im Zentrum der Betrachtung (siehe Tabelle 2).

Diese Betrachtungsweise findet zunehmend ihren Niederschlag in der klinischen Forschung, die 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. 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 (G-BA) hat sich in seiner Verfahrensordnung von 2006 bei der Bewertung diagnostischer Tests (§ 18 Abs. 7) dieser Sichtweise angeschlossen [6].

In Tabelle 2 sind verschiedene Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren gegenübergestellt. Neben der Einteilung von Fryback und Thornbury [7] und der Evidenzklassifizierung des G-BA (siehe Tabelle 3) wurde auch das Phasenschema nach Köbberling et al. [8] mit aufgenommen.

Die Studien zur prognostischen Güte eines diagnostischen Testverfahrens können in diesen Klassifikationsschemata der entsprechenden Einteilung für Studien zur diagnostischen Güte zugeordnet werden (z. B. zur diagnostischen Genauigkeit bei nicht bekanntem Krankheitsstatus).

Tabelle 2: Klassifikationsschemata für Evaluierungsstudien zu diagnostischen Testverfahren

Zielgröße	Köbberling et al. [8]	Fryback und Thornbury [7]	G-BA [6]
Patientenrelevante Zielgrößen	Phase 4: Wirksamkeit	6. Stufe: Auswirkung auf Systemebene	Evidenzstufe I
		5. Stufe: Auswirkung im Hinblick auf patientenrelevante Endpunkte	
Surrogatzielgrößen (Managementänderung)		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, Spezifität, Likelihood Ratio, prädiktive Werte	Phase 3: Diagnostische Genauigkeit bei nicht bekanntem Krankheitsstatus	2. Stufe: Auswirkung im Hinblick auf Diskriminationsfähigkeit	Evidenzstufe III
Sensitivität, Spezifität, Likelihood Ratio	Phase 2: Diagnostische Genauigkeit bei bekanntem Krankheitsstatus		
Analytische Sensitivität, Spezifität, Reproduzierbarkeit	Phase 1: Technische Voruntersuchungen	1. Stufe: Technische Auswirkungen	

Tabelle 3: Evidenzklassifizierung des G-BA<sup>1</sup> für Unterlagen zu diagnostischen Methoden

<b>Evidenzklasse</b>	<b>Kriterien</b>
<b>I a</b>	Systematische Übersichtsarbeiten von Studien der Evidenzstufe I b
<b>I b</b>	Randomisierte kontrollierte Studien
<b>I c</b>	Andere Interventionsstudien
<b>II a</b>	Systematische Übersichtsarbeiten von Studien zur diagnostischen Testgenauigkeit der Evidenzstufe II b
<b>II b</b>	Querschnitt- und Kohortenstudien, aus denen sich alle diagnostischen Kenngrößen zur Testgenauigkeit (Sensitivität und Spezifität, Wahrscheinlichkeitsverhältnisse, positiver und negativer prädiktiver Wert) berechnen lassen
<b>III</b>	Andere Studien, aus denen sich die diagnostischen Kenngrößen zur Testgenauigkeit (Sensitivität und Spezifität, Wahrscheinlichkeitsverhältnisse) berechnen lassen
<b>IV</b>	Assoziationsbeobachtungen, pathophysiologische Überlegungen, deskriptive Darstellungen, Einzelfallberichte u. Ä.; nicht mit Studien belegte Meinungen anerkannter Experten, Berichte von Expertenkomitees und Konsensuskonferenzen
1 siehe Verfahrensordnung des Gemeinsamen Bundesausschusses [6]	

#### 4.1 Konzept der Bewertung / Auswahl der Evidenzebene

Bei der vorliegenden Analyse wird der Nutzen und Schaden der PET bzw. der PET/CT beim rKRK bewertet. Falls diese Nutzenbewertung nicht erreichbar ist, wird die diagnostische und die prognostische Güte der PET bzw. der PET/CT beim rKRK ausgewertet.

Um den Nutzen und Schaden der bildgebenden diagnostischen Technologien PET bzw. PET/CT beim rKRK zu bewerten, soll untersucht werden, inwieweit durch das veränderte klinische Management die Mortalität, Morbidität und Lebensqualität der Patienten beeinflusst werden. Eine Änderung lediglich im Patientenmanagement, ohne dass sich daraus unmittelbar Schlussfolgerungen im Hinblick auf die Beeinflussung von Mortalität, Morbidität oder Lebensqualität ableiten lassen, stellt somit keine patientenrelevante Zielgröße im Sinne dieses Berichts dar.

Diese Fragen lassen sich am zuverlässigsten beantworten, wenn die verschiedenen diagnostischen Vorgehensweisen im Kontext von Managementstrategien betrachtet werden, bei denen die Patienten zusätzlich oder alternativ zur konventionellen Diagnostik nach dem Zufallsprinzip (randomisiert) die PET bzw. die PET/CT erhalten oder nicht, und überprüft wird, welche Änderungen sich dadurch für die Patienten ergeben. Die Überprüfung von verschiedenen Managementstrategien erfolgt deswegen im vorliegenden Bericht durch die

Bewertung von randomisierten kontrollierten Studien (RCTs) bzw. von systematischen Übersichten solcher Studien (ggf. im Rahmen eines HTA-Berichts). Finden sich keine bzw. nur eine geringe Zahl von RCTs, anhand derer der Nutzen und Schaden der PET bzw. PET/CT nicht mit ausreichender Ergebnissicherheit geschätzt werden kann, werden auch prospektive nicht randomisierte kontrollierte (Interventions-)Studien bei der Bewertung berücksichtigt.

Voraussetzung für die Einstufung einer Übersicht bzw. eines HTA-Berichts als systematische Übersicht ist, dass es sich um eine wissenschaftliche Analyse handelt, die auf transparente und reproduzierbare Weise versucht, nach vorab festgelegten Kriterien alle Primärstudien zur Fragestellung zu identifizieren, das Verzerrungspotenzial der Ergebnisse dieser Studien zu bewerten und eine stichhaltige Informationssynthese aus den Ergebnissen der Primärstudien abzuleiten.

Vorrecherchen zur Erstellung des Berichtsplans haben gezeigt, dass für eine Nutzenbewertung der PET bzw. der PET/CT kaum qualitativ ausreichende systematische Übersichten bzw. HTA-Berichte vorliegen. Folglich wird die Nutzenbewertung im Vorbericht ausschließlich auf Primärstudien basieren.

Die Bewertung der diagnostischen und prognostischen Güte eines Tests kann ebenfalls durch die Bewertung von diagnostischen bzw. prognostischen Primärstudien bzw. von systematischen Übersichten solcher Studien erfolgen (ggf. im Rahmen eines HTA-Berichts).

Zu der Frage der diagnostischen Güte der PET bzw. PET/CT wurden bei Vorrecherchen mehrere systematische Übersichten und HTA-Berichte mit mehreren eingeschlossenen Primärstudien gefunden. Diese Frage wird deshalb primär auf der Grundlage dieser Publikationen beurteilt, sobald sichergestellt ist, dass diese Publikationen tatsächlich als systematische Übersichten eingestuft werden und die methodischen Voraussetzungen erfüllen (siehe Einschlusskriterien Abschnitt 4.5.2.1).

Fragestellungen, zu denen mindestens eine aktuelle methodisch gute, systematische Übersicht bzw. HTA-Bericht zur diagnostischen bzw. prognostischen Güte der PET bzw. der PET/CT identifiziert werden kann, werden auf Basis dieser Publikationen ausgewertet. Informationen aus den neueren, in diesen Publikationen nicht berücksichtigten Studien werden im Anschluss an die systematischen Übersichten bzw. HTA-Berichte der jeweiligen Fragestellung angefügt. Diese Studien werden ausführlich im Hinblick auf ihr Potenzial, die Ergebnisse der Informationssynthese der systematischen Übersichten bzw. HTA-Berichte infrage zu stellen, ausgewertet.

Fragestellungen, zu denen keine systematischen Übersichten bzw. HTA-Berichte entsprechend obiger Definition einbezogen werden können, werden allein auf der Basis von Primärstudien bearbeitet.

## **4.2 Kriterien für den Einschluss von Studien und systematischen Übersichten bzw. HTA-Berichten in die Untersuchung**

### **4.2.1 Nutzenbewertung der PET bzw. der PET/CT anhand von Primärstudien**

#### **4.2.1.1 Population**

Folgende Patientenpopulationen werden eingeschlossen:

- Patienten mit begründetem Verdacht auf ein rKRK
- Patienten mit diagnostiziertem rKRK

Eingeschlossen werden Studien, bei denen der Anteil der Patienten mit rKRK (wie zuvor definiert) mindestens 80 % beträgt oder bei denen Subgruppenanalysen für solche Patienten durchgeführt wurden.

#### **4.2.1.2 Prüf- und Vergleichsintervention**

Als Prüfintervention gilt eine diagnostische Strategie mit Anwendung einer Vollring-PET bzw. einer Vollring-PET/CT. Als Vergleichsintervention (Komparator) wird eine diagnostische Strategie ohne Anwendung der PET bzw. der PET/CT betrachtet. Ein weiterer Vergleich beinhaltet eine diagnostische Strategie mit Anwendung der Vollring-PET/CT als Prüfintervention und eine diagnostische Strategie mit Vollring-PET als Komparator. 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 Ergebnissen der PET und CT) sind ebenfalls möglich [9].

#### **4.2.1.3 Zielgrößen**

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

- Mortalität

- krankheitsfreies Überleben
- Morbidität, z. B. Auftreten von Rezidiven, Komplikationen
- gesundheitsbezogene Lebensqualität

Da es sich bei dem rKRC um eine potenziell tödliche Erkrankung handelt, erfolgt die Bewertung der Zielgrößen krankheitsfreies Überleben, Morbidität und Lebensqualität in Abhängigkeit von der Zielgröße Mortalität.

Eine Änderung im Patientenmanagement alleine (ohne Verknüpfung zu den oben genannten Zielgrößen) stellt keine patientenrelevante Zielgröße im Sinne dieses Berichts dar. In Ausnahmefällen kann eine Managementänderung alleine Patientenrelevanz entfalten, wenn folgende Bedingungen erfüllt sind:

- Sie wirken sich unmittelbar plausibel auf patientenrelevante Endpunkte aus. Beispiele für solche Managementänderungen sind die Vermeidung von nebenwirkungsreichen Therapien, die Vermeidung von unnötigen, schmerzhaften diagnostischen Verfahren, oder die Zeitdauer bis zur Sicherung der Diagnose.
- Mit der Managementänderung dürfen bezüglich der oben genannten patientenrelevanten Zielgrößen keine inakzeptablen Verschlechterungen einhergehen. Sollte sich beispielsweise zeigen, dass sich mit der PET/CT eine Vermeidung schmerzhafter Behandlungen erreichen lässt, wäre aus dieser Managementänderung nur dann ein patientenrelevanter Nutzen ableitbar, wenn nicht gleichzeitig eine mehr als inakzeptable Erhöhung der Rezidivrate bestünde. Prinzipiell wäre für diesen Fall also eine „Nicht-Unterlegenheits“-Fragestellung gegeben, idealerweise mit der Vorabfestlegung von Akzeptanzgrenzen (im Sinne einer „höchstens irrelevanten Unterlegenheit“).

#### 4.2.1.4 Studientypen

Randomisierte klinische 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. Der Berichtsplan sieht daher primär den Einschluss von RCTs in die Bewertung vor.

Finden sich keine bzw. nur eine geringe Zahl von relevanten RCTs, anhand derer der Nutzen und Schaden der PET bzw. PET/CT nicht mit ausreichender Ergebnissicherheit geschätzt werden kann (z. B. geringe Präzision durch kleine Fallzahl bzw. Ereignisrate, hohes Verzerrungspotenzial durch methodische Mängel), werden gegebenenfalls auch verfügbare eindeutig prospektiv geplante, nicht randomisierte kontrollierte (Interventions-)Studien eingeschlossen. Solche Studien können zwar die Aussage von Ergebnissen aus qualitativ belastbaren RCTs mit nicht ausreichender Ergebnissicherheit in der Regel nicht qualitativ

ändern, diese aber ggf. substanziieren. Dieses Vorgehen scheint für die zugrunde liegende Fragestellung gerechtfertigt, da sich erst sehr langsam der Paradigmenwechsel in der Bewertung diagnostischer Verfahren von der alleinigen Analyse der diagnostischen Güte hin zu vergleichenden Studien mit patientenrelevanten Endpunkten entfaltet. Weiterhin erlauben möglicherweise nicht randomisierte kontrollierte (Interventions-)Studien wertvolle Hinweise auf die Ausgestaltung aussagekräftiger(er) Studien.

#### 4.2.1.5 Sonstige Charakteristika

Zumindest eine Vollpublikation zur Studie soll auf Englisch, Deutsch, Französisch oder Spanisch verfügbar sein.

Als Vollpublikation gilt 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 [10] oder des TREND-Statements [11] genügt und somit eine Bewertung der Studie ermöglicht.

Eine Mindestdauer für Studien wird nicht festgelegt.

#### 4.2.1.6 Tabellarische Zusammenfassung der Einschlusskriterien

Es werden für die Bewertung die Studien herangezogen, die alle Einschlusskriterien aus Tabelle 4 erfüllen.

Tabelle 4: Einschlusskriterien für Primärstudien zur Nutzenbewertung

E1a	Patienten mit rKRK bzw. mit begründetem Verdacht auf ein rKRK wie in Abschnitt 4.2.1.1 beschrieben
E2a	Prüf- und Vergleichsinterventionen wie in Abschnitt 4.2.1.2 beschrieben: diagnostische Strategie mit Anwendung der PET bzw. der PET/CT (Prüfintervention) versus diagnostische Strategie ohne Anwendung der PET bzw. der PET/CT (Vergleichsintervention) oder diagnostische Strategie mit Anwendung der PET/CT versus diagnostische Strategie mit Anwendung der PET
E3a	Patientenrelevante Endpunkte (Mortalität, Morbidität, Lebensqualität) und Änderungen im klinischen Management der Patienten hinsichtlich potenziell relevanten Endpunkten wie unter 4.2.1.3 beschrieben
E4a	Randomisierte kontrollierte Studien (und ggf. prospektive nicht randomisierte kontrollierte Studien) wie in Abschnitt 4.2.1.4 beschrieben
E5a	Zumindest eine Vollpublikation zur Studie soll auf Englisch, Deutsch, Französisch oder Spanisch verfügbar sein

## **4.2.2 Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT anhand von systematischen Übersichten bzw. HTA-Berichten**

### **4.2.2.1 Population**

Die Ergebnisse für folgende Populationen müssen für einzelne Studien aus den systematischen Übersichten bzw. HTA-Berichten extrahierbar sein:

- Patienten mit begründetem Verdacht auf ein rKRK (z. B. zum Nachweis des rKRK)
- Patienten mit diagnostiziertem rKRK (z. B. zur Bestimmung des Krankheitsstadiums)

### **4.2.2.2 Prüftechnologie, ggf. Vergleichstechnologie und Referenzstandard**

Als Prüftechnologie gilt die PET bzw. die PET/CT, ggf. dient als Vergleichstechnologie (Komparator) die Diagnostik ohne Anwendung der PET bzw. der PET/CT. Die Prüf- und ggf. auch die Vergleichstechnologie muss zu einem sogenannten Referenzstandard in Beziehung gebracht werden. In den einzuschließenden systematischen Übersichten bzw. HTA-Berichten muss ein solches Ordnungsprinzip erkennbar und die Ergebnisse aus den Einzelstudien entsprechend dargestellt sein.

Als zuverlässigster Referenzstandard bei der Diagnostik von Tumoren gilt die histologische Untersuchung von Operationspräparaten. Bei negativen Befunden sowie bei mehreren Metastasen ist eine Operation meistens klinisch nicht indiziert und ethisch nicht vertretbar. In diesen Fällen wird als Referenzstandard eine histologische Untersuchung eines Teils des Gewebes (durch Biopsie) und ggf. Verlaufsuntersuchungen verwendet. Als Referenzstandard für die prognostische Güte gilt die Nachbeobachtung.

Weitere Technologievergleiche bei diskordanten Befunden nach vorheriger PET bzw. PET/CT und anderer Diagnostik sind möglich. Diese Studien müssen so konzipiert sein, dass alle diskordanten Befunde analog zu den Studien zur diagnostischen und prognostischen Güte mit dem gleichen, validen Referenzstandard untersucht worden sind.

### **4.2.2.3 Zielgrößen**

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

#### 4.2.2.4 Den systematischen Übersichten bzw. HTA-Berichten zugrunde liegende Studientypen

- prospektive Kohortenstudien
- Querschnittstudien (gemäß Phase 3 nach Köbberling et al. [8])

#### 4.2.2.5 Sonstige Charakteristika

Zumindest eine Vollpublikation zur systematischen Übersicht bzw. zum HTA-Bericht soll auf Englisch, Deutsch, Französisch oder Spanisch verfügbar sein.

Als Vollpublikation gilt in diesem Zusammenhang auch die nicht vertrauliche Weitergabe eines Berichts an das Institut oder die nicht vertrauliche Bereitstellung eines Berichts über die Studien, der den Kriterien von Oxman und Guyatt zur Berichtsqualität (Punkte 1, 3, 5 und 7 sollen sich mit „ja“ beantworten lassen) genügt und somit eine Bewertung der Studie ermöglicht.

#### 4.2.2.6 Tabellarische Zusammenfassung der Einschlusskriterien

Es werden für die Bewertung die systematische Übersichten bzw. HTA-Berichte herangezogen, die alle Einschlusskriterien aus Tabelle 5 erfüllen.

Tabelle 5: Einschlusskriterien für systematische Übersichten bzw. HTA-Berichte zur Bewertung der diagnostischen und prognostischen Güte

E1b	Patienten mit rKRK bzw. mit begründetem Verdacht auf ein rKRK wie in Abschnitt 4.2.2.1 beschrieben
E2b	Prüftechnologie PET bzw. PET/CT, ggf. Vergleichstechnologie wie in Abschnitt 4.2.2.2 beschrieben
E3b	Zielgrößen (Sensitivität, Spezifität, Likelihood-Quotienten etc.) wie in Abschnitt 4.2.2.3 beschrieben
E4b	Systematische Übersichten und HTA-Berichte auf Basis von Kohortenstudien und Querschnittstudien wie in Abschnitt 4.2.2.4 beschrieben
E5b	Zumindest eine Vollpublikation der systematischen Übersicht bzw. des HTA-Berichts soll auf Englisch, Deutsch, Französisch oder Spanisch verfügbar

#### 4.2.3 Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT anhand von Primärstudien

##### 4.2.3.1 Population

Folgende Patientenpopulationen werden eingeschlossen:

- Patienten mit begründetem Verdacht auf ein rKRK (z. B. zum Nachweis des rKRK)

- Patienten mit diagnostiziertem rKRK (z. B. zur Bestimmung des Krankheitsstadiums)

Eingeschlossen werden Studien, bei denen der Anteil der Patienten mit rKRK (wie zuvor definiert) mindestens 80 % beträgt oder bei denen Subgruppenanalysen für solche Patienten durchgeführt worden sind.

#### **4.2.3.2 Prüftechnologie, ggf. Vergleichstechnologie (Komparator) und Referenzstandard**

Als Prüftechnologie gilt die PET bzw. die PET/CT, als Vergleichstechnologie (Komparator) dient ggf. die Diagnostik ohne Anwendung der PET bzw. der PET/CT. Prüf- und ggf. Vergleichstechnologien müssen zu einem Referenzstandard in Beziehung gebracht worden sein. In den einzuschließenden Studien muss ein solches Ordnungsprinzip erkennbar und die Ergebnisse entsprechend dargestellt sein.

Als zuverlässigster Referenzstandard bei der Tumordiagnostik gilt die histologische Untersuchung von Operationspräparaten. Bei negativen Befunden sowie bei mehreren Metastasen ist eine Operation meistens klinisch nicht indiziert und ethisch nicht vertretbar. In diesen Fällen wird als Referenzstandard eine histologische Untersuchung eines Teils des Gewebes (durch Biopsie) und ggf. die Verlaufsuntersuchungen verwendet. Als Referenzstandard für die prognostische Güte gilt die Nachbeobachtung.

#### **4.2.3.3 Zielgrößen**

- Beschreibung der diagnostischen Genauigkeit mittels Sensitivität / Spezifität (ggf. ROC-Kurven) oder positivem / negativem Likelihood-Quotienten, daraus abgeleitet ggf. prädiktive Aussagen mittels z. B. positivem / negativem prädiktiven Wert. Dabei muss die vollständige Ableitung der Information für eine Vierfeldertafel gewährleistet sein (es gilt als nicht ausreichend, wenn z. B. allein Werte für die Sensitivität angegeben werden). Bei Studien, in denen aufgrund des Designs nicht eine einzige Vierfeldertafel zugrunde gelegt werden kann, müssen alle relevanten Informationen vorhanden sein (für eine bivariate Analyse beispielsweise Sensitivität und Spezifität mit adäquaten Variabilitätsangaben). Ebenso können Studien eingeschlossen werden, die ausschließlich diskordante Befunde bezüglich der PET bzw. PET/CT und anderer Diagnostik darstellen, sofern die Datenlage aus Studien mit vollständigen Angaben zur Vierfeldertafel unzureichend ist. Diese Studien müssen so konzipiert sein, dass alle diskordanten Befunde analog zu den Studien zur diagnostischen und prognostischen Güte mit dem gleichen, validen Referenzstandard untersucht worden sind.
- prognostische Aussagen hinsichtlich zukünftiger klinischer Ereignisse (z. B. Überleben, Rezidive) anhand geeigneter statistischer Maßzahlen (z. B. Kaplan-Meier-Schätzungen)

#### **4.2.3.4 Studientypen**

- prospektive Kohortenstudien

- Querschnittstudien (gemäß Phase 3 nach Köbberling et al. [8])

#### 4.2.3.5 Sonstige Charakteristika

Zumindest eine Vollpublikation zur Studie soll auf Englisch, Deutsch, Französisch oder Spanisch verfügbar sein.

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 [12], TREND-Statements [11] oder STROBE-Statements [13] genügt und somit eine Bewertung der Studie ermöglicht.

#### 4.2.3.6 Tabellarische Zusammenfassung der Einschlusskriterien

Es werden für die Bewertung die Studien herangezogen, die alle Einschlusskriterien aus Tabelle 6 erfüllen.

Tabelle 6: Einschlusskriterien für Primärstudien zur Bewertung der diagnostischen und prognostischen Güte

E1c	Patienten mit rKRK bzw. mit begründetem Verdacht auf ein rKRK wie in Abschnitt 4.2.3.1 beschrieben
E2c	Prüftechnologie PET bzw. PET/CT, ggf. die Vergleichstechnologie wie in Abschnitt 4.2.3.2 beschrieben
E3c	Zielgrößen (Sensitivität, Spezifität, Likelihood-Quotienten etc.) wie in Abschnitt 4.2.3.3 beschrieben
E4c	Kohortenstudien bzw. Querschnittstudien wie in Abschnitt 4.2.3.4 beschrieben
E5c	Zumindest eine Vollpublikation zur Studie soll auf Englisch, Deutsch, Französisch oder Spanisch verfügbar

### 4.3 Informationsbeschaffung

#### 4.3.1 Bibliografische Literaturrecherche

Die systematische Suche nach relevanten Primärstudien und systematischen Übersichten bzw. HTA-Berichten wird in folgenden Quellen durchgeführt:

- Suche nach Primärstudien in den bibliografischen Datenbanken MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials (Clinical Trials)
- Suche nach relevanten systematischen Übersichten bzw. HTA-Berichten in den Datenbanken MEDLINE und EMBASE (parallel zur Suche nach relevanten Primärstudien) sowie 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 relevanten Informationen**

Zusätzlich zur Suche in bibliografischen Datenbanken werden folgende Quellen zur Identifizierung publizierter und nicht publizierter relevanter Informationen herangezogen:

- 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
- Um systematische Übersichten zu identifizieren, die im Rahmen einer Leitlinienerstellung durchgeführt wurden und die möglicherweise nicht durch die oben beschriebene Recherchestrategie identifizierbar sind, wird eine Recherche nach evidenzbasierten Leitlinien in folgenden Datenbanken durchgeführt: Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Guidelines International Networks (G-I-N) und National Guideline Clearinghouse (NGC).

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.

Für die Suche nach Primärstudien zur Beurteilung des Nutzens und Schadens sowie für systematische Übersichten bzw. HTA-Berichte zur Beurteilung der diagnostischen und prognostischen Güte gibt es keine zeitliche Einschränkung. Die Suche nach Primärstudien zur Beurteilung der diagnostischen und prognostischen Güte beschränkt sich auf den Zeitraum, der die Lücke zwischen der aktuellsten Recherche der eingeschlossenen systematischen Übersichten bzw. HTA-Berichte und dem Zeitpunkt der Recherche dieses Berichts schließt.

#### **4.3.3 Selektion relevanter Studien und systematischer Übersichten bzw. HTA-Berichte**

In einem zweistufigen Verfahren selektieren 2 Reviewer unabhängig voneinander potenziell relevante Publikationen aus den Referenzen der Literaturrecherche. In einem ersten

Selektionsschritt erfolgt anhand des Titels und ggf. Abstracts das Aussortieren von Publikationen zu Studien oder systematischen Übersichten bzw. HTA-Berichten, die mit Sicherheit die in Abschnitt 4.2 genannten Einschlusskriterien nicht erfüllen.

Referenzen, die einer der beiden Reviewer als potenziell relevant erachtet, werden im Volltext bestellt. In einem zweiten Schritt erfolgt von beiden Reviewern eine unabhängige Einstufung der Volltexte als relevant bzw. nicht relevant für den Einschluss in die Bewertung. Treffen auf eine Publikation mehrere Ausschlussgründe zu, wird einer der von den Reviewern genannten Ausschlussgründe dokumentiert. Diskrepanzen im Selektionsprozess werden durch Diskussion und ggf. Einbeziehung eines dritten Reviewers gelöst.

Die identifizierten Publikationen werden den einzelnen Primärstudien und systematischen Übersichten bzw. HTA-Berichten zugeordnet.

#### **4.4 Bewertung der einzelnen Studien**

Die Dokumentation der Studiencharakteristika und -ergebnisse bzw. der Ergebnisse von systematischen Übersichten oder HTA-Berichten sowie deren Bewertung erfolgen anhand von standardisierten Datenextraktionsbögen und Evidenztabelle. Die Datenextraktion wird von einem unabhängigen Reviewer durchgeführt. Ein zweiter unabhängiger Reviewer überprüft die Vollständigkeit und Korrektheit der extrahierten Daten. Diskrepanzen werden durch Diskussion oder ggf. durch Hinzuziehung eines dritten Reviewers aufgelöst.

##### **4.4.1 Bewertung von Studien zur Nutzenbewertung**

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

Die Bewertung der Studieninformationen hängt stark von der Berichtsqualität der jeweiligen Publikationen ab. Alle für die Nutzenbewertung relevanten Ergebnisse werden hinsichtlich ihrer Ergebnissicherheit, bestehend aus dem Verzerrungspotenzial und der Präzision der Ergebnisse, überprüft.

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 mit großer Wahrscheinlichkeit ausgeschlossen werden kann, dass die Ergebnisse relevant verzerrt sind. Von einem hohen Verzerrungspotenzial wird ausgegangen, wenn nicht 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 (z. B. Ablehnung oder Nichtablehnung einer Nullhypothese, Änderung der Effektrichtung) 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.

Unter „Präzision der Ergebnisse“ wird im vorliegenden Bericht die Breite der 95 %-Konfidenzintervalle verstanden. Bei dramatischen Effekten kann auch ein 99 %-Konfidenzintervall verwendet werden.

#### **4.4.2 Bewertung von systematischen Übersichten und HTA-Berichten zur diagnostischen und prognostischen Güte**

Es wird eine Übersicht über die in den systematischen Übersichten bzw. HTA-Berichten jeweils eingeschlossenen Studien mit den wesentlichen Studiencharakteristika tabellarisch dargestellt. Die Ergebnisse der zugrunde liegenden Primärstudien werden dabei nach den

Indikationen, Technologien (Prüftechnologie, ggf. Komparator, ggf. Referenzstandard) und Studientypen aufgeteilt.

Die Bewertung der allgemeinen Qualität einer systematischen Übersicht bzw. eines HTA-Berichts<sup>1</sup> erfolgt mindestens anhand der Kriterien von Oxman und Guyatt [14,15]. Die Bewertung der systematischen Übersichten bzw. der HTA-Berichte kann bei Bedarf um Kriterien z. B. zum Umgang mit Verifikations- bzw. Inkorporationsbias ergänzt werden.

#### **4.4.3 Bewertung von Primärstudien zur diagnostischen und prognostischen Güte**

Falls die Bewertung von Primärstudien zur diagnostischen und prognostischen Güte erforderlich ist, werden die Studien zunächst nach Indikationen, Technologievergleichen und Studientypen aufgeteilt. Die Angaben zu Studienpopulationen, verglichenen Technologien, methodischen Aspekten und Ergebnissen einzelner Studien werden tabellarisch dargestellt.

Die Qualitätsbewertung der diagnostischen und prognostischen Primärstudien wird mithilfe von QUADAS-Checklisten [16] durchgeführt. Die QUADAS-Checklisten beinhalten Fragen sowohl bezüglich der Berichtsqualität als auch des Verzerrungspotenzials einer Studie.

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

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

#### **4.5 Informationssynthese**

Die Informationen einzelner Studien werden einer Informationssynthese unterzogen. Wenn möglich, werden über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die Methoden der quantitativen Informationssynthese (Meta-Analyse, Sensitivitätsanalyse und Subgruppenanalyse) eingesetzt. Abschließend erfolgt eine qualitative zusammenfassende Gesamtbewertung aller Informationen.

##### **4.5.1 Nutzenbewertung der PET bzw. der PET/CT**

###### **4.5.1.1 Kriterien zum Einschluss in die Informationssynthese**

Eine Einstufung des Verzerrungspotenzials des Ergebnisses für einen Endpunkt als „hoch“ führt nicht zum Ausschluss aus der Informationssynthese der Nutzenbewertung. Die

---

<sup>1</sup> Hierzu zählen auch systematische Übersichten, die ggf. als integraler Bestandteil von sog. evidenzbasierten Leitlinien identifiziert werden.

Klassifizierung dient der Interpretation heterogener Studienergebnisse und beeinflusst die Sicherheit der Aussage.

In bestimmten Fällen werden einzelne Ergebnisse aus den Studien zu einem Endpunkt nicht dargestellt bzw. nicht in die Informationssynthese der Nutzenbewertung 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.

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 [17]. Ein Ausschluss aus der Informationssynthese erfolgt auch, wenn der Unterschied der fehlenden Werte zwischen den Gruppen größer als 15 Prozentpunkte ist.

Zum Einschluss in die Informationssynthese einer nicht randomisierten Studie muss das Problem einer möglichen Strukturungleichheit (unfairer Vergleich) in dieser Studie adäquat in der Planung und Auswertung berücksichtigt worden sein.

#### **4.5.1.2 Gegenüberstellung der Ergebnisse der Einzelstudien**

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

Da es sich bei dem rKRC um eine potenziell tödliche Erkrankung handelt, erfolgt die Bewertung der Zielgrößen Morbidität und Lebensqualität in Abhängigkeit von der Zielgröße Mortalität.

#### **4.5.1.3 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 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 (engl.: random-effects-model) [18]. In begründeten Ausnahmefällen werden Modelle mit festen Effekten (engl.: fixed-effects-model) 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 bzw. näherungsweise bestimmt.

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. In begründeten Ausnahmefällen kommen auch andere Effektmaße zum Einsatz.

Zunächst werden die Effektschätzer und Konfidenzintervalle aus den Studien 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 [19]. 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 (4.5.1.4) und klinische Faktoren, sogenannte Effektmodifikatoren (4.5.1.5).

#### **4.5.1.4 Sensitivitätsanalysen**

Zur Einschätzung der Robustheit der Ergebnisse sind Sensitivitätsanalysen hinsichtlich methodischer Faktoren geplant. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ und „niedrig“ wird hierfür verwendet. Weitere methodische Faktoren können aus den im Rahmen der Informationsbewertung aufgefallenen Merkmalen abgeleitet werden, z. B. die Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder die Wahl des Effektmaßes.

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

#### **4.5.1.5 Subgruppenmerkmale und andere Effektmodifikatoren**

Die Ergebnisse werden hinsichtlich potenzieller Effektmodifikatoren – d. h. Faktoren, die je nach Ausprägung die Effektstärke unterschiedlich beeinflussen – untersucht. Dies können direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Behandlungen sein. Im Gegensatz zu den in Abschnitt 4.5.1.4 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 Homogenitätstest (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

Sollten aufgrund der 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.

## **4.5.2 Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT**

### **4.5.2.1 Kriterien zum Einschluss in die Informationssynthese**

Eine Voraussetzung für die Berücksichtigung einer systematischen Übersicht bzw. eines HTA-Berichts in der Informationssynthese ist das Erreichen eines Mindestpunktwertes von 5 für den Index von Oxman und Guyatt [14,15]. Die genaue Klassifizierung über den Mindestpunktwert von 5 anhand der Kriterien von Oxman und Guyatt [14,15] dient der Interpretation heterogener Studienergebnisse bei der Informationssynthese und beeinflusst die Sicherheit der Aussage.

Die Einstufung der Primärstudien zur diagnostischen und prognostischen Güte nach ihrer Qualität führt nicht automatisch zum Ausschluss dieser Studien aus der Informationssynthese. Die Einstufung dient primär der Interpretation heterogener Studienergebnisse bei der Informationssynthese und beeinflusst nur die Sicherheit der Aussage.

### **4.5.2.2 Gegenüberstellung der Ergebnisse der einzelnen systematischen Übersichten bzw. HTA-Berichte sowie der Primärstudien**

Die Ergebnisse der systematischen Übersichten bzw. der HTA-Berichte zur diagnostischen und prognostischen Güte werden zunächst vergleichend beschrieben. Danach erfolgt durch Gegenüberstellung dieser Ergebnisse eine Abschätzung der Heterogenität der Ergebnisse innerhalb der jeweiligen Fragestellung.

Außerdem werden Ergebnisse aus den neueren Studien, die in den (in die Informationssynthese einbezogenen) systematischen Übersichten bzw. HTA-Berichten nicht berücksichtigt sind, den Ergebnissen aus den systematischen Übersichten bzw. HTA-Berichten gegenübergestellt und im Hinblick auf ihr Potenzial ausgewertet, die Ergebnisse der Informationssynthese aus systematischen Übersichten bzw. HTA-Berichten infrage zu stellen.

Beim Vorliegen inkonsistenter Ergebnisse aus mehreren Arbeiten zu einer Fragestellung wird nach möglichen Erklärungen für die beobachtete Heterogenität gesucht.

#### **4.5.2.3 Meta-Analysen**

Eine Meta-Analyse zur Bewertung der diagnostischen und prognostischen Güte der PET bzw. der PET/CT erfolgt, sofern die Studien hinsichtlich der Fragestellung und relevanter Charakteristika vergleichbar sind, analog dem in Abschnitt 4.5.1.3 beschriebenen Vorgehen. Dabei werden jedoch nur die Ergebnisse der Primärstudien aus der Ergänzungsrecherche in Meta-Analysen zusammengefasst und den ggf. vorhandenen Meta-Analysen aus den systematischen Übersichten / HTA-Berichten gegenübergestellt.

Bei Diagnostikstudien wird für Testgütekriterien die Meta-Analyse anhand der wichtigsten diagnostischen und prognostischen Zielgrößen (z. B. Sensitivität, Spezifität, Likelihood Ratios) in einer bivariaten Analyse durchgeführt [20,21].

#### **4.5.2.4 Sensitivitätsanalysen**

Zur Einschätzung der Robustheit der Ergebnisse sind Sensitivitätsanalysen geplant. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse anhand der Kriterien des QUADAS-Instruments [16] (Qualität „niedrig“ oder „hoch“) wird hierfür verwendet. Weitere methodische Faktoren zur Überprüfung in der Sensitivitätsanalyse können aus den im Rahmen der Informationsbewertung aufgefallenen Merkmalen abgeleitet werden, z. B. verblindete Beurteilung der Prüf- und Vergleichstechnologie vs. unverblindete Beurteilung oder konsekutive vs. nicht konsekutive Patientenrekrutierung.

#### **4.5.2.5 Subgruppenmerkmale und andere Effektmodifikatoren**

Die Auswertung der Ergebnisse hinsichtlich der Subgruppenmerkmale und Effektmodifikatoren bei der Bewertung der diagnostischen und prognostischen Güte der PET bzw. der PET/CT erfolgt analog dem in Abschnitt 4.5.1.5 beschriebenen Vorgehen.

## 5 Literaturverzeichnis

1. Schmiegel W, Reinacher-Schick A, Arnold D, Graeven U, Heinemann V, Porschen R et al. S3-Leitlinie "Kolorektales Karzinom": Aktualisierung 2008; Ergebnis einer evidenzbasierten Konsensuskonferenz (8.–9. Juni 2007). Z Gastroenterol 2008; 46(8): 799-840.
2. Batzler WU, Giersiepen K, Hentschel S, Husemann G, Kaatsch P, Katalinic A et al. Krebs in Deutschland 2003 - 2004: Häufigkeiten und Trends. Berlin: Robert Koch-Institut; 2008. URL: <http://www.ekr.med.uni-erlangen.de/GEKID/Doc/kid2008.pdf>.
3. Sobin LH, Wittekind C (Ed). TNM classification of malignant tumours. New York: Wiley-Liss; 2002.
4. 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: 26.09.2007]. URL: <http://www.g-ba.de/downloads/40-268-238/2003-05-23-Abschluss-PET.pdf>.
5. Deutsche Forschungsgemeinschaft PET-CT [online]. [Zugriff: 29.10.2009]. URL: [http://www.dfg.de/forschungsfoerderung/wissenschaftliche\\_infrastruktur/wgi/download/pet\\_c t\\_stellungnahme\\_050209.pdf](http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/wgi/download/pet_c t_stellungnahme_050209.pdf).
6. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses [online]. 18.04.2006 [Zugriff: 22.08.2007]. URL: [http://www.g-ba.de/downloads/62-492-83/VerfO\\_2006-04-18.pdf](http://www.g-ba.de/downloads/62-492-83/VerfO_2006-04-18.pdf).
7. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991; 11(2): 88-94.
8. Köbberling J, Trampisch HJ, Windeler J. Memorandum zur Evaluierung diagnostischer Maßnahmen. Nuklearmedizin 1989; 28: 255-260.
9. Lijmer JG, Bossuyt PMM. Various randomized designs can be used to evaluate medical tests. J Clin Epidemiol 2009; 62(4): 364-373.
10. 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.
11. 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.

12. 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.
13. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.
14. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44(11): 1271-1278.
15. 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.
16. 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.
17. 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.
18. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3): 177-188.
19. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
20. 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.
21. Egger M, Davey Smith G, Altman DG (Ed). *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books; 2001.