

## **Kurzfassung**

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) wurde vom Gemeinsamen Bundesausschuss (G-BA) beauftragt, eine Nutzenbewertung der Positronenemissionstomographie (PET und PET/CT) bei Kopf- und Halstumoren durchzuführen.

## **Fragestellung**

Die vorliegende Untersuchung verfolgte 2 Ziele:

### **1: Ermittlung des patientenrelevanten Nutzens der PET bzw. PET/CT**

Primäres Ziel war die Beschreibung des patientenrelevanten Nutzens der Bildgebung mit der PET bzw. PET/CT bei Kopf- und Halstumoren bzw. bei unbekanntem Primärtumor, d. h. insbesondere bei metastatischer zervikaler Adenopathie unbekanntem Ursprungs. Die untersuchten Indikationen für den Einsatz waren a) die Bestimmung des Tumorstadiums („Staging“), b) das Ansprechen des Tumors auf die Behandlung („Residual Disease Evaluation“ / „Re-Staging“) und c) der Nachweis eines Rezidivs bei begründetem Verdacht. Darüber hinaus wurde untersucht, ob sich beim unbekanntem Primärtumor 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 wurden 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.

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

Aufgrund des Mangels an aussagekräftigen Primärstudien zur Ermittlung des patientenrelevanten Nutzens (Ziel 1) wurde zusätzlich eine systematische Bewertung der diagnostischen und prognostischen Wertigkeit der PET bzw. PET/CT vorgenommen (Ziel 2). Dabei ging es primär um die Frage, inwieweit die PET bzw. PET/CT den diagnostischen Standardverfahren ohne PET überlegen ist. Das bedeutet: Erhöht sich durch den Einsatz der PET bzw. PET/CT die Anzahl der korrekten Zuordnungen zum jeweiligen Stadium mit den unterschiedlichen prognostischen Konsequenzen, die Anzahl der korrekten Beurteilungen des Therapieansprechens oder die Anzahl korrekter Diagnosen bzw. die Anzahl korrekter Ausschlüsse von Rezidiven?

## **Methoden**

Für die Nutzenbewertung wurden (randomisierte) kontrollierte Vergleichsstudien – Strategie mit vs. ohne PET – mit patientenrelevanten Endpunkten (z. B. verringerte Mortalität / Morbidität) berücksichtigt, für die Beurteilung der Testgüte systematische Übersichtsarbeiten bzw. prospektive Kohorten- und Querschnittstudien.

Ein „Review of Reviews“ diente als Basis für die Beantwortung der zweiten Fragestellung. Dieser wurde durch eine zusätzliche Recherche nach Primärstudien ergänzt, deren Suchzeitraum sich mit jenem der eingeschlossenen systematischen Übersichtsarbeiten überlappt (Ergänzungsrecherche).

## **Ergebnisse**

Eine umfassende systematische Recherche in bibliografischen Datenbanken und anderen Quellen lieferte für die Nutzenbewertung nur 1 vergleichende Nutzenstudie. Den Einschlusskriterien des Berichts entsprachen 4 systematische Übersichtsarbeiten mit insgesamt 69 Primärstudien. Die Frage des Stagings und der Rezidiverkennung wurde jeweils von 1 systematischen Übersichtsarbeit, die Frage der Detektion von unbekanntem Primärtumoren von 2 systematischen Übersichtsarbeiten untersucht. Zur Fragestellung Therapieansprechen gab es keine systematische Übersichtsarbeit oder Meta-Analyse. Zusätzlich zu den systematischen Übersichtsarbeiten wurden 28 relevante Primärstudien identifiziert.

## **Belege für einen patientenrelevanten Nutzen der PET bzw. PET/CT**

Die einzige Vergleichsstudie zur Nutzenbewertung der PET bzw. PET/CT beim Staging untersuchte in einem direkten Vergleich 2 diagnostisch-therapeutische Strategien mit und ohne PET, konnte aber für das rezidivfreie 2-Jahres-Überleben keinen Unterschied zwischen den Gruppen nachweisen. Bei der geringen Anzahl von Rezidiven (6 Ereignisse in der Gruppe mit PET; 4 Ereignisse in der Gruppe ohne PET) in einer Population von 102 Patienten hat die Studie ungenügende Power, um einen möglicherweise existierenden Unterschied zwischen den beiden Strategien nachzuweisen. Mit diesen Studienergebnissen lässt sich ein patientenrelevanter Nutzen der PET weder nachweisen noch widerlegen.

## **Diagnostische Güte der PET bzw. PET/CT beim Staging**

Insgesamt wurden nur 1 systematische Übersichtsarbeit und 10 zusätzliche Primärstudien eingeschlossen. Hierbei waren 5 Studien auf die Fragestellung Staging ausgerichtet, die systematische Übersichtsarbeit und 3 weitere Studien auf die Teilfragestellung Staging von Lymphknotenmetastasen und 2 Studien auf die Teilfragestellung Staging von Fernmetastasen. Zur Teilfragestellung Staging von Zweitumoren lieferte 1 Studie, die auch Angaben zum Staging von Fernmetastasen machte, Informationen.

Die systematische Übersichtsarbeit zeigt bei den eingeschlossenen Diagnostikstudien erhebliche methodische Schwächen, alle eingeschlossenen Studien waren nur von niedriger bis mäßiger Qualität. Hauptprobleme waren das häufig retrospektive Design und die unklare oder fehlende Verblindung bei der Beurteilung der PET bzw. PET/CT.

Die meisten der zusätzlich identifizierten Primärstudien hatten – bis auf 2 Studien – alle eine sehr niedrige Güte der Evidenz, was vor allem an einer unklaren oder fehlenden Verblindung

bei der Bewertung der Tests und der Referenzstandards, einer unklaren Patientenselektion und vor allem an sehr kleinen Patientenpopulationen lag. Bei der Fragestellung Staging des Primärtumors untersuchten 3 Primärstudien insbesondere die Detektion einer Knocheninvasion, für die ein Test mit einer hohen Spezifität nötig ist. Die Vergleichstests CT und SPECT hatten meist eine tendenziell höhere Spezifität als die PET. Für die Teilfragestellung Detektion der Lymphknotenmetastasen sollte der Test eine hohe Sensitivität haben. Die PET schnitt dabei nicht wesentlich besser ab als die Vergleichstests MRT und CT. In der Subgruppenanalyse nach Lymphknotenstadium hatten die Tests für Patienten mit klinischem Stadium cN0 aufgrund ihrer sehr niedrigen Sensitivität keine Nachweiskraft mehr. Bei der Teilfragestellung Detektion von Fernmetastasen zeigte die Kombination PET+CT verglichen mit der PET keine wesentliche Verbesserung, während bei den Einzeltests die PET eine höhere Sensitivität als die CT hatte. Die Ergebnisse waren jedoch zum Teil sehr ungenau. Alle Technologien zeigten eine vergleichbar hohe Spezifität. Die zum Teil extrem weiten Konfidenzintervalle bei allen Fragestellungen (z. T. 1–99 %) mit breiten Überlappungen zwischen den Technologien erlauben keine zuverlässigen Aussagen über die Überlegenheit der PET bzw. PET/CT verglichen mit den konventionellen Technologien.

### **Diagnostische Güte der PET bzw. PET/CT beim Therapieansprechen**

Die Beurteilung des Therapieansprechens nach verschiedenen Behandlungen wurde von 10 Primärstudien untersucht. Die Studien wiesen alle, bis auf eine Studie mit mäßig hoher Güte der Evidenz, eine niedrige bis sehr niedrige Güte der Evidenz auf. Hauptprobleme waren sehr kleine Patientenzahlen, eine unklare oder fehlende Verblindung bei der Beurteilung der Tests und der Referenzstandards und fehlende Angaben zur Therapie, die möglicherweise zwischen der PET-Untersuchung und dem Referenzstandard Nachbeobachtung stattgefunden hat. Für zahlreiche Vergleiche existierte nur 1 kleine Studie, der häufigste Vergleich (PET vs. CT) beruhte auf 3 kleinen Studien. Weder die PET noch die CT hatte eine gute gepoolte Sensitivität, die Spezifität war sehr gut für die PET und ungenügend für die CT. Tendenziell hatte die PET eine bessere Spezifität als die CT. Entsprechend den geringen Studiengrößen und den methodischen Mängeln ist eine zuverlässige Aussage zur diagnostischen Wertigkeit der PET bzw. PET/CT im Vergleich zu anderen Technologien allerdings nicht möglich.

### **Diagnostische Güte der PET bzw. PET/CT bei der Rezidiverkennung**

Die systematische Übersichtsarbeit zur Detektion von Rezidiven bei Patienten mit Larynxkarzinomen hatte nur Studien mit niedriger Qualität eingeschlossen. Ein besonders häufiges Problem war ein inadäquater Referenzstandard. Insgesamt hatte die PET eine gute gepoolte Sensitivität und eine mäßige gepoolte Spezifität. Die 8 zusätzlich identifizierten Primärstudien hatten alle eine niedrige bis sehr niedrige Güte der Evidenz. Für die PET reichten die Resultate von klinisch sehr aussagekräftigen Werten (obere Grenze des Konfidenzintervalls für Sensitivität 96 % und für Spezifität 93 %) bis zu klinisch irrelevanten Werten (untere Grenze des Konfidenzintervalls für Sensitivität 76 % und für Spezifität 62 %).

Dabei lag die Sensitivität der PET höher als die der Vergleichstests. Für alle Vergleiche waren die Konfidenzintervalle überlappend und sehr weit, sodass eine Schlussfolgerung bezüglich der Überlegenheit der PET bzw. PET/CT gegenüber anderen Technologien nicht getroffen werden kann.

### **Diagnostische Güte der PET bzw. PET/CT bei der Detektion eines unbekanntem Primärtumors**

Eine systematische Übersichtsarbeit berichtete über die diagnostische Güte der PET, eine andere über die der PET/CT zur Detektion des Primärtumors bei Läsionen mit unbekanntem Primärtumor. Die Primärstudien in beiden Synthesen liefern aufgrund erheblicher methodischer Mängel nur Evidenz von niedriger Güte. Beide systematischen Übersichtsarbeiten betrachteten ausschließlich die diagnostische Güte der PET und PET/CT ohne Bezug zu einem gängigen Vergleichstest. Die Detektionsrate für die PET (433 Patienten) lag im Median bei 43 % (Range zwischen 8 % und 65 %). Die gepoolte Sensitivität war gut [87% (81-92%)], die gepoolte Spezifität nur mäßig [71 % (64-78 %)]. Ein Vergleich der PET/CT mit PET (3 Studien) fand keinen signifikanten Unterschied. Die gepoolte Detektionsrate der PET/CT lag bei 37 % (Der Range betrug 22 % bis 73 %). Die gepoolte Sensitivität [84 % (78-88 %)] und Spezifität [84 % (78-89 %)] für die PET/CT aus der Beobachtung an 433 Patienten wurde als gut eingestuft. Die kleine Primärstudie aus der Ergänzungsrecherche hatte bei erheblichen methodischen Mängeln eine unzureichende diagnostische Güte, änderte aber nicht die Einschätzung der Ergebnisse der systematischen Übersichtsarbeiten. Beim unbekanntem Primärtumor kann die PET wie auch die PET/CT zusätzliche Primärtumoren nachweisen. Der Zeitpunkt des Einsatzes der PET im Ablauf der Diagnostik (komplette vs. inkomplette Abklärung) beeinflusst die Prävalenz der noch nicht identifizierten Tumoren und damit die Detektionsrate. In den vorliegenden Studien ließ sich kein Unterschied in der diagnostischen Güte für die PET/CT und die CT nachweisen.

### **Managementänderungen aufgrund von PET bzw. PET/CT**

In 8 der eingeschlossenen Primärstudien wurde über Managementänderungen berichtet, die in diesem Bericht nur ergänzend dargestellt wurden. Der Prozentsatz der berichteten Managementänderungen reichte von 2 % bis 100 %.

Eine Managementänderung per se, kann nicht als Beleg für den Nutzen eines diagnostischen Tests gewertet werden, da nicht alle Managementänderungen mit positiven Auswirkungen für die Patienten verbunden sein müssen. So führte z. B. in der Studie von **Yen 2005** die PET bei 3 Patienten (6 %) zu einer unnötigen Ausdehnung der Neck-Dissection. Eine solche rückblickende Beurteilung der Managementänderungen fand in den Studien häufig nicht statt und kann bestenfalls nur Hinweise auf den Wert der Managementänderungen geben. Um zu belegen, dass die Managementänderungen, die sich aus den Ergebnissen eines diagnostischen Tests ergeben, auch zu einem Nutzen für den Patienten führen, sind in den meisten Fällen vergleichende Interventionsstudien, möglichst mit randomisiertem Design erforderlich.

## **Fazit**

Für das primäre Staging, das Lymphknotenstaging und die Diagnostik von Fernmetastasen helfen weder die einzige Nutzenstudie noch die aufgefundenen Studien zur diagnostischen Güte, einen möglichen Zusatznutzen der PET und PET/CT im Kontext der vorhandenen Testverfahren in der Versorgung nachzuweisen. Obwohl eine große Zahl von Testgütestudien den Stellenwert der PET und PET/CT im primären Lymphknotenstaging untersucht hat, ließ sich weder für die Sensitivität noch die Spezifität eine deutliche Verbesserung gegenüber anderen diagnostischen Verfahren erzielen.

Bei der Bewertung des Therapieansprechens hatte die PET in den Einzelstudien tendenziell eine höhere diagnostische Güte als die anderen Technologien mit erheblicher Unsicherheit in den geschätzten Werten. Selbst ein Poolen der 3 Studien mit dem CT-Vergleich ermöglichte keine Aussage zur Überlegenheit der PET oder einer anderen Technologie.

Die PET zum Nachweis von Rezidiven erreichte in der eingeschlossenen Meta-Analyse eine hohe Sensitivität mit einer guten Detektion von Rezidiven. Die mäßige Spezifität bedeutet gleichzeitig eine höhere Anzahl falsch positiver Befunde mit weiterem Abklärungsbedarf. Sind die sonstigen Untersuchungsbefunde negativ, verbleibt bei Arzt und Patient eine Restunsicherheit über die Existenz eines Rezidivs und das beste weitere therapeutische Vorgehen. Ob die PET der konventionellen Diagnostik (CT / MRT) wirklich überlegen ist, kann anhand der Ergebnisse nicht zuverlässig beantwortet werden. Inwieweit ein frühzeitiger Rezidivnachweis auch Vorteile für den Patienten bringt, muss noch untersucht werden.

Beim unbekanntem Primärtumor spielt die PET eine mögliche Rolle, wenn der Tumor nach einer vollständigen Untersuchung mit allen verfügbaren anderen diagnostischen Verfahren nicht lokalisiert werden konnte. Der zu erwartende Benefit ist wegen der heterogenen Studienpopulation und der unterschiedlichen Abklärungsstadien der Patienten (komplette versus inkomplette Abklärung vor der PET) in den Studien nicht (ausreichend) gut fassbar. Ob die Untersuchung mit PET oder PET/CT für die Detektion von unbekanntem Primärtumoren mit einem patienten-relevanten Nutzen verbunden ist, verbleibt gegenwärtig unklar.

## **Forschungsbedarf**

Generell gilt, dass bei der Planung, der Durchführung und dem Reporting von diagnostischen Studien unverändert ein erheblicher Verbesserungsbedarf besteht. Bei den Tumoren des Kopf-Hals-Bereichs besteht für die PET/CT möglicherweise eine Rolle in der Rezidivdiagnostik. Gleiches gilt für die Diagnostik des unbekanntem Primärtumors. Vergleichsstudien mit dem aktuellen diagnostischen Standard sind dabei unerlässlich. Aufgrund der weiten Verbreitung von PET/CT sollten solche Studien mit der fusionierten Technologie durchgeführt werden. Ob eine bestehende höhere diagnostische Güte der PET/CT bzw. PET auch mit einem

patientenrelevanten Nutzen verbunden ist, muss jedoch in randomisiert-kontrollierten Studien untersucht werden.

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