

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

**Dokumentation und Würdigung der  
Anhörung zum 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

# Inhaltsverzeichnis

	Seite
<b>Abkürzungsverzeichnis.....</b>	<b>iii</b>
<b>1 Dokumentation der Anhörung.....</b>	<b>1</b>
<b>2 Würdigung der Anhörung.....</b>	<b>2</b>
<b>2.1 Anmerkungen zum Kapitel 2. Ziele der Untersuchung.....</b>	<b>2</b>
<b>2.2 Anmerkungen zum Kapitel 4.1 Konzept der Bewertung / Auswahl der Evidenzebene .....</b>	<b>3</b>
<b>2.3 Anmerkungen zu Kapitel 4.2.1.3 Zielgrößen (Nutzenbewertung) und 4.2.1.4 Studientypen (Nutzenbewertung) .....</b>	<b>3</b>
<b>2.4 Anmerkungen zu Kapitel 4.2.2.3 / 4.2.3.3 Zielgrößen Diagnosestudien.....</b>	<b>5</b>
<b>2.5 Anmerkungen zu Kapitel 4.2.2.4 / 4.2.3.4 Studientypen diagnostische Güte .....</b>	<b>6</b>
<b>2.6 Anmerkungen zum Literaturverzeichnis.....</b>	<b>6</b>
<b>2.7 Sonstige.....</b>	<b>6</b>
<b>2.8 Literaturverzeichnis.....</b>	<b>7</b>
<b>3 Offenlegung potenzieller Interessenkonflikte.....</b>	<b>8</b>
<b>3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen .....</b>	<b>8</b>
<b>3.2 Interessenkonflikte von stellungnehmenden Privatpersonen .....</b>	<b>8</b>
<b>3.3 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige) .....</b>	<b>9</b>
<b>4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....</b>	<b>13</b>
<b>4.1 Teilnehmerliste der wissenschaftlichen Erörterung .....</b>	<b>13</b>
<b>4.2 Tagesordnung der wissenschaftlichen Erörterung .....</b>	<b>13</b>
<b>4.3 Protokoll der wissenschaftlichen Erörterung.....</b>	<b>14</b>
4.3.1 Begrüßung und Einleitung .....	14
4.3.2 Tagesordnungspunkt 1 Zielgrößen in der Nutzenbewertung: Reicht die diagnostische Güte für die Nutzenbewertung aus? .....	16
4.3.3 Tagesordnungspunkt 2 Zielgrößen in Diagnosestudien: Ist die Inter-Rater-Variabilität ein eigenständiger Aspekt der Testgüte? .....	24
4.3.4 Tagesordnungspunkt 3 Verschiedenes .....	31
<b>Anhang: Dokumentation der Stellungnahmen.....</b>	<b>38</b>

**Abkürzungsverzeichnis**

<b>Abkürzung</b>	<b>Bedeutung</b>
G-BA	Gemeinsamer Bundesausschuss
HTA	Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
RCT	Randomised Controlled Trial (randomisierte kontrollierte Studie)
PET	Positronenemissionstomographie
CT	Computertomographie
FDG	Fluor-Deoxy-Glukose
ICD	International Statistical Classification of Diseases and Related Health Problems

## **1 Dokumentation der Anhörung**

Am 02.09.2009 wurde der vorläufige Berichtsplan in der Version Nr. 1.0 vom 20.08.2009 veröffentlicht und zur Anhörung gestellt. Bis zum 30.09.2009 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 3 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 22.10.2009 im IQWiG diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

Eine Würdigung der in der Anhörung vorgebrachten Aspekte befindet sich in Kapitel 2. Im überarbeiteten Berichtsplan sind darüber hinaus Änderungen, die sich durch die Anhörung ergeben haben, zusammenfassend dargestellt. Der überarbeitete Berichtsplan ist auf der Website des IQWiG unter [www.iqwig.de](http://www.iqwig.de) veröffentlicht.

## 2 Würdigung der Anhörung

Die im Rahmen der Anhörung vorgebrachten Aspekte wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des Berichtsplans überprüft. Die wesentlichen Argumente werden im Folgenden diskutiert. Neben projektspezifischen wissenschaftlichen Aspekten wurden auch übergeordnete Punkte, z. B. zu rechtlichen Vorgaben für das Institut, angesprochen. Auf solche Punkte wird im Rahmen dieser projektspezifischen Würdigung der Anhörung nicht weiter eingegangen.

Im Folgenden werden die Argumente aus den abgegebenen Stellungnahmen und aus der Erörterung jeweils unterhalb des betroffenen Berichtsplanabschnitts zusammengefasst. Die Argumente aus den Stellungnahmen und aus der Erörterung werden in Kursiv, die Würdigung der Argumente in Normalschrift dargestellt.

### 2.1 Anmerkungen zum Kapitel 2. Ziele der Untersuchung

*In den Stellungnahmen wurde bemängelt, dass die vorgeschlagene Methodik im Wesentlichen für Arzneimittelbewertungen geeignet sei, die in randomisierten Studien getestet wurden.*

Dazu ist anzumerken, dass randomisierte Studien auch für Fragestellungen zu nicht-medikamentösen Interventionen – also auch diagnostische Maßnahmen – eingesetzt werden [1,2]. Darüber hinaus lässt die hier vorgeschlagene Methodik sowohl für die Nutzenbewertung als auch für die Bewertung der diagnostischen Güte auch nichtrandomisierte Designs zu [3].

*Bezüglich der Prüfintervention wurde darauf hingewiesen, dass eine Differenzierung zwischen alleiniger PET-Untersuchung und Untersuchungen mit einer Kombination aus einem PET-Gerät und einem CT-Gerät nicht zeitgemäß sei.*

Dieser technischen Weiterentwicklung wird mit der vorgeschlagenen Berichtsmethodik insofern Rechnung getragen, als beide Verfahren sowohl einzeln als auch in Kombination mit einem Standardvorgehen verglichen werden und somit auch getrennt ausgewertet werden können.

Außerdem befindet sich in Deutschland aktuell noch eine große Zahl von reinen PET-Geräten im Einsatz [4]. Insofern halten wir die Bewertung des Nutzens und der diagnostischen Güte von Stand-alone-Geräten durchaus für sinnvoll und notwendig.

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

## **2.2 Anmerkungen zum Kapitel 4.1 Konzept der Bewertung / Auswahl der Evidenzebene**

*Mit Hinweis auf das bereits abgeschlossene und veröffentlichte Projekt zur PET-Diagnostik bei malignen Lymphomen (D0601A [5]) wird gefolgert, dass sich eine Bewertung der diagnostischen Güte auf Basis von Sekundärliteratur (wie für das Projekt D0601C unter bestimmten Bedingungen vorgesehen) nicht bewährt habe und somit abzulehnen sei.*

Tatsächlich haben die Autoren des D0601A-Berichts die Probleme bei der Erstellung ausführlich diskutiert. Sie kommen aber nicht zu dem Schluss, dass die Ergebnisse aus diesem Grund in ihrer Validität beeinträchtigt wären – hierzu heißt es in dem Bericht: „Dabei stellte die Berücksichtigung von ausschließlich hochwertigen Evidenzsynthesen ein hohes Qualitätsniveau sicher. Zusätzlich wurden von den Verfassern des vorliegenden Berichts [D0601A] Einschätzungen zur Angemessenheit der Schlussfolgerungen zur Evidenzlage abgegeben.“ Die Erfahrungen aus dem Projekt D0601A fließen in die Methodik des Projekts D0601C ein. Beispielsweise werden durch die Verwendung des Bewertungsinstruments von Oxman und Gyatt [6,7] Mindestanforderungen an die Qualität der Reviews gestellt und es wird zusätzlich die vollständige Ableitbarkeit der diagnostischen Vierfeldertafeln für jede Einzelstudie gefordert, womit die von den Stellungnehmenden erwähnten methodischen Schwierigkeiten minimiert werden können.

*Im Zusammenhang mit dem für dieses Projekt kaum realisierbaren Goldstandard (Biopsie bzw. histologische Aufarbeitung von Operationspräparaten auch für negative Körperbereiche) wird ein „pragmatischer“ Referenzstandard gefordert.*

Ein solcher Ansatz wurde bereits im vorläufigen Berichtsplan [3] skizziert, dort heißt es: „...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.“

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

## **2.3 Anmerkungen zu Kapitel 4.2.1.3 Zielgrößen (Nutzenbewertung) und 4.2.1.4 Studientypen (Nutzenbewertung)**

*In 2 Stellungnahmen wurde kritisiert, dass eine Begrenzung der Nutzenbewertung auf Studien der Evidenzklasse I und II (gemäß G-BA-Evidenzklassifizierung) nicht adäquat sei. Des Weiteren wurde In 2 Stellungnahmen die Verknüpfung von diagnostischer und therapeutischer Strategie zur Bestimmung eines patientenrelevanten Nutzens kritisiert. In 1 Stellungnahme wurde die Verknüpfung sogar grundsätzlich abgelehnt. Der Nutzen einer diagnostischen Methode sei bereits daran erkennbar, ob sie das therapeutische Management bzw. das therapeutische Denken / Handeln beeinflussen könne. Z. B. sei die Änderung der*

*Intention einer Strahlentherapie von „kurativ“ in „palliativ“ alleine ausreichend, um den Nutzen für den Patienten zu belegen, und es bedürfe keiner weiteren randomisierten Studien.*

Inwiefern die diagnostische Güte als Zielgröße für die Nutzenbewertung ausreicht, wurde am 22.10.2009 im Rahmen einer mündlichen Erörterung diskutiert. Mit Hinblick auf die PETAL-Studie [8], in deren Planung auch 2 der an der Erörterung teilnehmenden Personen eingebunden sind, wurde die „grundsätzliche“ Ablehnung der Verknüpfung von diagnostischer und therapeutischer Strategie durch die Stellungnehmer insofern relativiert, als es Ausnahmen von der Regel geben müsse, wie im Beispiel der PETAL-Studie (die Studie ermöglicht einen randomisierten Vergleich zwischen PET-positiven Patienten, die einerseits mit Standardtherapie und andererseits mit einer intensivierten Therapie behandelt werden). *Vonseiten der Stellungnehmer wurde die Ansicht vertreten, dass es nicht die Aufgabe des „Diagnostikers“ sei zu beweisen, dass die Kette aus Diagnostik und Therapie für den Patienten eine Verbesserung bedeute. Aus Sicht des Diagnostikers reiche es aus zu zeigen, dass die diagnostische Methode zum einen „charakterisiert“ (siehe Kapitel 4.3.2) und so präzise wie möglich ist und zum anderen in der Lage ist, den Therapeuten, abhängig vom Ergebnis der Diagnostik, zu einer Behandlungsänderung zu bewegen.*

Diese Sichtweise ignoriert, dass alleine anhand der Testcharakteristik nicht beurteilt werden, ob beispielsweise die durch eine neue diagnostische Methode zusätzlich entdeckte Fälle auch zu einem im Mittel verbesserten Ergebnis für alle positiven Fälle führt. Beispielsweise könnten gerade die zusätzlichen Fälle jene sein, die nicht auf eine entsprechend indizierte Therapie ansprechen (sog. Nonresponder-Selektion) [9] und darüber hinaus können Fälle unentdeckt bleiben, die mit der herkömmlichen Diagnostik entdeckt worden wären und ein Therapieansprechen zu erwarten gehabt hätten, obwohl das neue Verfahren insgesamt mehr positive Fälle als das herkömmliche entdeckt.

*Weiter heißt es in 1 Stellungnahme zur krankheitsstadienadaptierten Therapie: „Demzufolge übt das korrekte Staging beim rezidivierenden kolorektalen Karzinom grundsätzlich einen prognoserelevanten Einfluss auf den jeweiligen Patienten aus, selbst wenn dieser noch nicht direkt nachgewiesen sein sollte.“*

Fokussiert man, wie hier vorgeschlagen, alleine auf die „richtige“ Klassifikation der Fälle, lassen sich Phänomene, wie beispielsweise das sog. Will-Rogers-Phänomen [10,11] nicht ausschließen. Hierbei ist es möglich, ein und dieselbe Gruppe von Patienten mit bekanntem mittlerem Therapieerfolg so auf unterschiedliche Krankheitsstadien umzuverteilen, dass sich anschließend die Behandlungserfolge in allen Stadien verbessern – was zu dem Trugschluss führen kann, das Behandlungsergebnis habe sich insgesamt verbessert.

Zusammenfassend wurde in den schriftlichen Stellungnahmen bzw. in der mündlichen Erörterung nicht überzeugend dargelegt, wie die Bewertung eines patientenrelevanten Nutzens, ohne die Verknüpfung der diagnostischen Güte mit den daraus resultierenden Konsequenzen möglich sein soll.



## 2.4 Anmerkungen zu Kapitel 4.2.2.3 / 4.2.3.3 Zielgrößen Diagnosestudien

*In einer Stellungnahme wurde vorgeschlagen, neben den üblichen diagnostischen Gütekriterien auch die sog. Inter-Observer-Variabilität (auch Inter-Rater- bzw. Zwischen-Beobachter-Variabilität) in die Bewertung eines diagnostischen Verfahrens einzubeziehen. Beispielsweise könnten 2 zu vergleichende Verfahren eine gleiche (mittlere) diagnostische Güte aufweisen, aber eines der Verfahren eine geringere Inter-Observer-Variabilität aufweisen. In einem solchen Fall solle das Verfahren mit geringerer Variabilität als besser beurteilt werden.*

Dieser Aspekt wurde auch in der Erörterung vom 22.10.2009 diskutiert.

Es ist theoretisch vorstellbar, dass die Übertragbarkeit der in Studien (beispielsweise unter Einbezug von Experten) ermittelten Testgüte bei einer geringeren Variabilität zwischen den Bewertern besser gelingt, als wenn diese Variabilität hoch ist. Allerdings ist dieser Zusammenhang nicht zwingend aus einem Variabilitätsunterschied ableitbar. So kann eine größere Variabilität eines Verfahrens auch bedeuten, dass dieses ein größeres Potenzial in sich trägt, genauere Diagnosen zu liefern, als ein Verfahren mit geringerer Variabilität: Beispielsweise könnte bei einer gleichen (mittleren) Sensitivität von 90 % das Verfahren mit höherer Variabilität (z. B. Spanne 80–100 %) gegenüber dem Verfahren mit geringerer Variabilität (z. B. Spanne 75–85 %) als das angesehen werden, was grundsätzlich in der Lage ist, auch eine Sensitivität von 100 % zu erreichen, wohingegen das andere Verfahren über eine Sensitivität von 85 % nicht hinauskommt.

Es ist weiterhin vorstellbar, dass bei gleicher Sensitivität (und Spezifität) die höhere Variabilität des einen Verfahrens gegenüber einem zweiten Verfahren zu unterschiedlichen Zuordnungen zu Krankheitsstadien führt. Sofern eine Studie nur die (binäre) Zuordnung in krank und nicht krank untersuchte bzw. über alle Krankheitsstadien hinweg nur die richtig und falsch zugeordneten Patienten untersuchte (jeweils bzgl. eines Goldstandards), wären bei ebenfalls gleicher Testgüte unterschiedliche therapeutische Konsequenzen mit unterschiedlichen Therapieergebnissen möglich. Allerdings wäre auch hier eine eindeutige Bewertung von *weniger* Variabilität als *besseres* Therapieergebnis (oder auch vice versa) keinesfalls zwingend.

Letztlich kann festgehalten werden, dass die Variabilität Einfluss auf die Übertragbarkeit der diagnostischen Güte bzw. auf die Behandlungsergebnisse, die aus der mit der Diagnose verknüpften therapeutischen Konsequenz resultieren, nehmen kann. Als Konsequenz für den Bericht ergibt sich, dass die Ergebnisse zur Testgüte im Kontext ggf. vorliegender Aussagen zur Inter-Observer-Variabilität diskutiert werden.

*Zum gleichen Berichtsabschnitt wurde in 1 Stellungnahme angemerkt, dass die skizzierte Methodik zur Bewertung des Nutzens bzw. der diagnostischen und prognostischen Güte auf*

*die Stufe 2 nach Fryback & Thornbury fokussiere und die von den gleichen Autoren vorgeschlagenen Stufen 3 und 4 ebenfalls einbezogen werden sollten.*

Die Stufen 3 und 4 nach Fryback & Thornbury beschreiben die Bewertung einer diagnostischen Maßnahme anhand der Surrogatzielgröße Managementänderung. Hierzu ist anzumerken, dass der vorläufige Berichtsplan vorsieht, die Managementänderungen unter gewissen Umständen in die Betrachtung des Berichts einzubeziehen [3].

Somit ergibt sich bezüglich der Berücksichtigung der Zielgröße Managementänderung keine Konsequenz für den Bericht.

*Der Hinweis aus 1 Stellungnahme, dass die prognostische Güte einen relevanten, die Lebensqualität beeinflussenden Faktor darstelle, erübrigt sich insofern, als dass die Lebensqualität als eine patientenrelevante Zielgröße vorgesehen ist.*

## **2.5 Anmerkungen zu Kapitel 4.2.2.4 / 4.2.3.4 Studientypen diagnostische Güte**

*Für die spezielle Situation, in der die PET-Diagnostik als sogenannter Add-on-Test eingesetzt wird (d. h., es werden nur die testnegativen bzw. testpositiven Fälle der Standarddiagnostik mittels der PET-Diagnostik nachuntersucht), wird auf die Möglichkeit hingewiesen, nur die diskordanten Fälle einer Überprüfung durch einen Referenzstandard zu unterziehen.*

Ein solches Vorgehen ist im vorläufigen Berichtsplan bereits vorgesehen, sowohl für die Nutzenbewertung als auch für die Bewertung der diagnostischen Güte. Einschränkend wird lediglich gefordert, dass alle diskordanten Befunde analog zu den Studien zur diagnostischen und prognostischen Güte mit dem gleichen, validen Referenzstandard untersucht worden sind.

Es ergibt sich keine Konsequenz für den Berichtsplan.

## **2.6 Anmerkungen zum Literaturverzeichnis**

*Dem Vorschlag, das Verzeichnis der ein- bzw. ausgeschlossenen Literatur zugunsten eines schnelleren Auffindens der Publikationen eines bestimmten Autors in alphabetischer Reihenfolge zu präsentieren, wird insofern entsprochen, als die ausgeschlossenen Referenzen in alphabetischer Sortierung mit zusätzlicher Nennung des Ausschlussgrunds gelistet werden.*

## **2.7 Sonstige**

*Dem Hinweis, dass in sogenannten S3-Leitlinien systematische Reviews als Grundlage der Empfehlung erstellt oder recherchiert worden sein können, wird insofern entsprochen, als im Rahmen einer orientierenden Leitlinienrecherche versucht werden wird, solche Evidenz zu berücksichtigen, sofern diese aus der Leitlinie selbst oder entsprechend referenzierten Publikationen zu entnehmen ist.*

## 2.8 Literaturverzeichnis

1. 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).
2. Köbberling J, Trampisch HJ, Windeler J. Memorandum zur Evaluierung diagnostischer Maßnahmen. Nuklearmedizin 1989; 28: 255-260.
3. IQWiG. PET und PET/CT bei rezidivierenden kolorektalen Karzinomen. Berichtsplan D06-01C (vorläufige Version 0.1). 2009.
4. PET-e.V. I. PET(/CT)-Einrichtungen [online]. [Zugriff: 3.12.2009]. URL: [http://www.petev.de/pet\\_einrichtungen.html](http://www.petev.de/pet_einrichtungen.html).
5. IQWiG. Positronenemissionstomographie (PET und PET/CT) bei malignen Lymphomen. Vorbericht D06-01A. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). 2009.
6. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44(11): 1271-1278.
7. 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.
8. Universitätsklinikum Essen. Positronen-Emissionstomographie-gesteuerte Therapie aggressiver Non-Hodgkin-Lymphome (PETAL-Studie): Therapieoptimierungsprüfung; Studienprotokoll; Studiennummer 01-2006; Version 1.04 [unveröffentlicht]. 2008.
9. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? Ann Intern Med 2006; 144(11): 850-855.
10. Chee KG, Nguyen DV, Brown M, Gandara DR, Wun T, Lara PN Jr. Positron emission tomography and improved survival in patients with lung cancer: the Will Rogers phenomenon revisited. Arch Intern Med 2008; 168(14): 1541-1549.
11. Golder W. Positron emission tomography and lymphoma therapy. Onkologie 2001; 24(5): 496-498.

### 3 Offenlegung potenzieller Interessenkonflikte

Im Folgenden sind die potenziellen Interessenkonflikte der Stellungnehmenden sowie weiterer Teilnehmer an der wissenschaftlichen Erörterung zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter [www.iqwig.de](http://www.iqwig.de) abrufbar. Für die externen Sachverständigen liegt das „Formblatt zur Darlegung potenzieller Interessenkonflikte“ mit Stand 12/2006 (8 Fragen) vor. Die in den jeweiligen Versionen der Formblätter aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

#### 3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
ACOMED Statistik	Keller, Thomas, Dr.	nein	ja	nein	ja	nein	nein
Deutsche Gesellschaft für Nuklearmedizin e. V.	Bockisch, Andreas, Prof. Dr. Dr.	nein	nein	nein	nein	nein	nein
	Grünwald, Frank, Prof. Dr.	nein	nein	ja	ja	nein	ja
	Kotzerke, Jörg, Prof. Dr.	nein	nein	nein	nein	nein	nein
Deutsche Krankenhausgesellschaft	Brenske, Michael, Dr.	ja	nein	nein	nein	nein	nein
	Krummrey, Gert, Dr.	nein	nein	nein	nein	nein	nein
	Neumeyer-Gromen, Angela, Dr.	nein	nein	nein	nein	nein	nein
	Schlottmann, Nicole, Dr.	nein	nein	nein	nein	nein	nein

#### 3.2 Interessenkonflikte von stellungnehmenden Privatpersonen

Es wurden keine Stellungnahmen durch Privatpersonen abgegeben.

**3.3 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige)**

<b>Name</b>	<b>Frage 1</b>	<b>Frage 2</b>	<b>Frage 3</b>	<b>Frage 4</b>	<b>Frage 5</b>	<b>Frage 6</b>	<b>Frage 7</b>	<b>Frage 8</b>
Schönermark, Matthias, Prof. Dr.	nein	nein	nein	nein	nein	nein	nein	nein
Hagen, Anja, MPH	nein	nein	nein	nein	nein	nein	nein	nein
Gorenoi, Vitali, Dr., MPH	nein	nein	nein	nein	nein	nein	nein	nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt:

*Frage 1:* Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter<sup>1</sup> abhängig (angestellt) beschäftigt? Falls ja, wo und in welcher Position?

*Frage 2:* Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor einen Interessenverband im Gesundheitswesen oder einen vergleichbaren Interessenvertreter direkt oder indirekt beraten? Falls ja, wen und wie hoch ist / war die Zuwendung / das Honorar?

*Frage 3:* Haben Sie abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres oder der 3 Kalenderjahre davor im Auftrag eines Interessenverbands im Gesundheitswesen oder eines vergleichbaren Interessenvertreters Honorare für Vorträge, Stellungnahmen, Ausrichtung und / oder Teilnahme an Kongressen und Seminaren – auch im Rahmen von Fortbildungsveranstaltungen, oder für (populär-)wissenschaftliche oder sonstige Aussagen oder Artikel erhalten? Falls ja, von wem, für welche Tätigkeiten und wie hoch war die Zuwendung / das Honorar?

*Frage 4:* Haben Sie abseits einer Anstellung oder Beratungstätigkeit und / oder hat die Institution<sup>2</sup>, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten? Falls ja, von wem, für welche Tätigkeit und in welcher Höhe?

*Frage 5:* Haben Sie und / oder hat die Institution, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres oder der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Reisekostenunterstützung ohne wissenschaftliche Gegenleistungen) von einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter erhalten? Falls ja, von wem, aus welchem Anlass und in welcher Höhe?

*Frage 6:* Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile (auch in Fonds) von einer Firma oder Institution, die zu einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter gehört? Falls ja, von wem und welchen Wert haben diese aktuell?

---

<sup>1</sup> Dieses Formblatt erfasst finanzielle Beziehungen zu Interessenverbänden im Gesundheitswesen oder vergleichbaren Interessenvertretern, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie.

<sup>2</sup> Sofern Sie in einer ausgedehnten Institution tätig sind, ist es ausreichend, die geforderten Angaben auf Ihre Arbeitseinheit (z. B.: Klinikabteilung, Forschungsgruppe etc.) zu beziehen.

Im „Formblatt zur Darlegung potenzieller Interessenkonflikte“ (Stand 12/2006) wurden folgende 8 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb der letzten 3 Jahre bei einer Person, Institution oder Firma<sup>3</sup> abhängig (angestellt) beschäftigt, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut<sup>4</sup> finanziell profitieren könnte?

Frage 2: Beraten Sie oder haben Sie innerhalb der letzten 3 Jahre eine Person, Institution oder Firma direkt oder indirekt<sup>5</sup> beraten, die von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

Frage 3: Haben Sie abseits einer Anstellung oder Beratungstätigkeit innerhalb der letzten 3 Jahre im Auftrag einer Person, Institution oder Firma, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte, Honorare für Vorträge, Stellungnahmen, Ausrichtung und/oder Teilnahme an Kongressen und Seminaren – auch im Rahmen von Fortbildungsveranstaltungen – oder für (populär-) wissenschaftliche oder sonstige Aussagen oder Artikel erhalten?<sup>6</sup>

Frage 4: Haben Sie und/oder die Einrichtung, für die Sie tätig sind, abseits einer Anstellung oder Beratungstätigkeit innerhalb der letzten 3 Jahre von einer Person, Institution oder Firma, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte, finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten?

Frage 5: Haben Sie und/oder die Einrichtung, bei der Sie angestellt bzw. beschäftigt sind, innerhalb der letzten 3 Jahre sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Reisekostenunterstützung ohne wissenschaftliche Gegenleistung) von einer Person, Institution oder Firma erhalten, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile (auch in Fonds) von einer Firma oder Institution, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

---

<sup>3</sup> Mit solchen „Personen, Institutionen oder Firmen“ sind im Folgenden alle Einrichtungen gemeint, die direkt oder indirekt einen finanziellen oder geldwerten Vorteil aus dem Ergebnis Ihrer wissenschaftlichen Arbeit für das Institut ziehen könnten. Hierzu gehören z. B. auch medizinische Einrichtungen, die eine zu bewertende medizinische Intervention durchführen und hierdurch Einkünfte erhalten.

<sup>4</sup> Mit „wissenschaftlicher Arbeit für das Institut“ sind im Folgenden alle von Ihnen für das Institut erbrachten oder zu erbringenden Leistungen und / oder an das Institut gerichteten mündlichen und schriftlichen Recherchen, Bewertungen, Berichte und Stellungnahmen gemeint.

<sup>5</sup> „Indirekt“ heißt in diesem Zusammenhang z. B. im Auftrag eines Institutes, das wiederum für eine entsprechende Person, Institution oder Firma tätig wird.

<sup>6</sup> Sofern Sie von einer Person, Institution oder Firma im Verlauf der letzten 3 Jahre mehrfach Honorare erhalten haben reicht es aus, diese für die jeweilige Art der Tätigkeit summiert anzugeben.

*Frage 7:* Haben Sie persönliche Beziehungen zu einer Person, Firma oder Institution bzw. Mitarbeitern einer Firma oder Institution, welche von den Ergebnissen Ihrer wissenschaftlichen Arbeit für das Institut finanziell profitieren könnte?

*Frage 8:* Gibt es andere bislang nicht dargestellte potenzielle Interessenkonflikte, die in Beziehung zu Ihrer wissenschaftlichen Arbeit für das Institut stehen könnten?



#### 4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

##### 4.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma /privat
Bockisch, Andreas, Prof. Dr. Dr.	Deutsche Gesellschaft für Nuklearmedizin e. V.
Gorenoi, Vitali, Dr., MPH	Institut für Epidemiologie, Sozialmedizin und Gesundheitssystemforschung
Großelfinger, Robert, Dr.	IQWiG
Hagen, Anja, MPH	Institut für Epidemiologie, Sozialmedizin und Gesundheitssystemforschung
Keller, Thomas, Dr.	ACOMED Statistik
Kotzerke, Jörg, Prof. Dr.	Deutsche Gesellschaft für Nuklearmedizin e. V.
Krummrey, Gert, Dr.	Deutsche Krankenhausgesellschaft e. V.
Kulig, Michael, PD Dr.	IQWiG
Lange, Stefan, PD Dr.	IQWiG (Moderation)
Scheibler, Fülöp, Dr.	IQWiG
Skipka, Guido, Dr.	IQWiG
Welter, Stefan	Protokollant (Sitzungsdokumentarischer Dienst, Landtag NRW)

##### 4.2 Tagesordnung der wissenschaftlichen Erörterung

<b>Begrüßung und Einleitung</b>	
<b>TOP 1</b>	<b>Zielgrößen in der Nutzenbewertung:</b> Reicht die diagnostische Güte für die Nutzenbewertung aus?
<b>TOP 2</b>	<b>Zielgrößenin Diagnosestudien:</b> Ist die Inter-Rater-Variabilität ein eigenständiger Aspekt der Testgüte?
<b>TOP 3</b>	<b>Verschiedenes</b>

### 4.3 Protokoll der wissenschaftlichen Erörterung

- Datum: 22.10.2009, 14:00 bis 15:05 Uhr
- Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG),  
Dillenburger Straße 27, 51105 Köln
- Moderation: PD Dr. Stefan Lange

#### 4.3.1 Begrüßung und Einleitung

**Moderator Stefan Lange:** Guten Tag, meine Dame, meine Herren! Ein Kollege von uns kommt gleich noch. Aber wir fangen trotzdem schon einmal an.

Bis auf Herrn Keller kennen unsere Sachverständigen das Prozedere; aber ich muss es trotzdem noch einmal kurz erläutern. Wir zeichnen diese Veranstaltung auf und haben dazu vorher die folgende Information gegeben: Wer damit nicht einverstanden ist, müsste diese Erörterung verlassen. – Aber das ist offensichtlich nicht der Fall, sodass wir jetzt starten können.

Sie haben, so hoffe ich, vorher die Tagesordnung zugesandt bekommen. Das sind die für uns offenen Fragen. Sinn und Zweck dieses Gespräches ist es, für uns aus den Stellungnahmen nicht ganz klare Punkte noch einmal zu besprechen, sie erläutert zu bekommen, anzudiskutieren und dann gegebenenfalls entsprechend berücksichtigen zu können. Das heißt nicht, dass wir die Punkte, die hier nicht angesprochen sind, nicht berücksichtigen bzw. diskutieren werden, sondern die angesprochenen Punkte sind für uns unklar geblieben.

Es ist auch nicht Sinn und Zweck dieses Gespräches, dass wir jetzt unbedingt zu einem Konsens kommen; das behalten wir uns natürlich vor. Wir wollen auch nicht über die grundsätzlichen Aspekte der Zusammenarbeit zwischen Gemeinsamem Bundesausschuss und IQWiG oder über die grundsätzliche Sinnhaftigkeit von evidenzbasierter Medizin als Grundlage für Entscheidungen sprechen, sondern tatsächlich spezifisch auf das Projekt eingehen.

Ich habe vergessen, mich vorzustellen. Mein Name ist Stefan Lange. Ich bin der stellvertretende Leiter dieses Instituts. Wir führen meistens in den jeweiligen Tagesordnungspunkt ein, was für uns diskussionswürdig oder fraglich ist. Dann bitten wir Sie, sich dazu zu äußern. Wir haben einen Zeitrahmen von zwei Stunden vorgesehen. Den sollten wir eigentlich gut einhalten können.

Als letzten Punkt möchte ich Sie bitten, auch wenn wir jetzt eine relativ intime Runde sind, damit es für den Protokollanten einfacher ist, bei einer Wortmeldung zu sagen, wie man heißt.

Wenn es ansonsten von Ihrer Seite noch Fragen zum grundsätzlichen Ablauf gäbe, dann könnten wir sie jetzt an dieser Stelle klären. – Das ist offensichtlich nicht der Fall.

Vielleicht als Letztes noch zur Erläuterung: Unter dem TOP „Verschiedenes“ können wir das eine oder andere, was Ihnen noch besonders am Herzen liegt, besprechen bzw. aufnehmen. Das Ganze wird als Wortprotokoll der Öffentlichkeit zugänglich gemacht.

Gut, dann starten wir. Herr Großelfinger wird in den ersten TOP einführen.

#### 4.3.2 Tagesordnungspunkt 1

##### **Zielgrößen in der Nutzenbewertung: Reicht die diagnostische Güte für die Nutzenbewertung aus?**

**Robert Großelfinger:** Es geht um die Frage, ob diagnostische Güte allein zur Nutzenbewertung ausreicht. Dazu muss ich ein paar Sätze vorausschicken. Dieser Punkt hat im Wesentlichen mit einer Stellungnahme der DGN zu tun, in der es – ich zitiere das wörtlich – heißt:

„Die Verknüpfung von diagnostischer und therapeutischer Strategie zur Bestimmung eines patientenrelevanten Nutzens ist grundsätzlich abzulehnen.“

Nun wurde an vergleichbarer Stelle bei Erörterungen zu anderen Indikationen – auch zur PET – dieser Punkt schon einmal behandelt. Trotzdem ist bei uns diesbezüglich noch eine Frage offengeblieben, insbesondere deshalb, weil sich der Kontext verändert hat. Denn inzwischen haben wir Kenntnis von einem Studienprotokoll. Ich spreche das Studienprotokoll der PETAL-Studie an, die meinem Verständnis nach so angelegt ist, dass eben genau diese Verknüpfung der diagnostischen und der therapeutischen Einheit – in der PETAL-Studie sogar darüber hinaus noch zusätzlich auf patientenrelevante Endpunkte fokussierend – explizit vorgenommen wurde.

Da wäre es für mich schon interessant, von Ihnen zu hören, wie Sie diesen Widerspruch zwischen dem Statement in der Stellungnahme einerseits und andererseits der Durchführung solcher Studien auflösen, an denen Sie in Person beteiligt sind.

**Andreas Bockisch:** Offen gesagt: Ich sehe den Widerspruch sehr konstruiert. Das ist eine ganz spezielle Frage, bei der man herausfinden möchte, ob die PET in der Lage ist, das therapeutische Prozedere zu steuern. Das kann man ja machen; das ist gar kein Problem. Das heißt doch nicht, dass grundsätzlich – darum geht es ja – die Diagnostik nur dann Wert hat bzw. der Nutzen der Diagnostik darüber bewiesen wird, dass am Ende der ganzen Kette die Therapie nachweislich zu einem besseren Überleben führt.

**Moderator Stefan Lange:** Herr Scheibler.

**Fülöp Scheibler:** Wenn Sie sagen, die Frage der PETAL-Studie ist, ob die PET in der Lage ist, die nachfolgende Therapie zu steuern, frage ich mich: Was ist dann das Ziel? Das Ziel ist doch, dass es den Patienten besser geht, wenn sie mittels PET diagnostiziert wurden, und dass es mittels der unterschiedlichen Chemotherapieregime – manchmal auch mit Stammzelltransplantation kombiniert – schließlich zu besseren Ergebnissen für die Patienten führt.

**Andreas Bockisch:** Ja, aber hier geht es doch darum, dass man in einer ganz bestimmten Situation in dieser Studie ganz konkret eine Frage stellt und mit der Studie diese Frage

beantworten möchte. Das kann ja nicht verboten sein. Das kann aber doch auch nicht der Kronzeuge dafür sein, dass man das immer so zu tun hat.

**Moderator Stefan Lange:** Aber Sie haben – Entschuldigung, ich darf das ergänzen – geschrieben: „... ist grundsätzlich abzulehnen.“

**Andreas Bockisch:** Ja.

**Moderator Stefan Lange:** Ich finde schon: „Grundsätzlich“ ist ein starkes Wort.

**Andreas Bockisch:** „Grundsätzlich“ heißt: in aller Regel und: Es gibt Ausnahmen.

**Moderator Stefan Lange:** Ah ja, okay. – Bitte, Herr Kotzerke.

**Jörg Kotzerke:** Vielleicht darf ich zur PETAL-Studie noch ergänzen: Es ist in diesem Falle ja nicht so, dass die PET-Diagnostik im Sinne eines Medikaments auf dem Prüfstand wäre, sondern das ist ja so geklärt gewesen, dass die PET-Untersuchung im Rahmen der Krankenversorgung stattfindet.

(Zustimmung von Fülöp Scheibler)

**Moderator Stefan Lange:** Das habe ich nicht verstanden, ehrlich gesagt. Können Sie das noch einmal erläutern?

**Jörg Kotzerke:** Sie haben die Frage nach dem Nutzen und nach der diagnostischen Güte gestellt sowie das Beispiel der PETAL-Studie angeführt, um zu sagen: Da wird doch genau das gemacht, was ihr andererseits nicht haben wollt.

(Zustimmung von Fülöp Scheibler)

Genau das ist aber so nicht, dass die FDG-PET dort generell auf dem Prüfstand wäre, sondern sie hat, in der Erstdiagnostik durchgeführt, einen Nutzen für den Patienten, weil man stadiengerecht sieht, und sie hat auch beim Restaging einen Nutzen, weil man die Patienten identifiziert, die auf die Therapie unzureichend ansprechen.

**Fülöp Scheibler:** Uns ist schon klar, dass die PETAL-Studie nicht den generellen Nutzen der PET im Rahmen der Diagnostik von Non-Hodgkin-Lymphomen grundsätzlich untersuchen will, sondern dass das eine sehr spezifische Fragestellung ist, die da untersucht wird. Es geht um Patienten, die im Rahmen der Primärtherapie und im Rahmen des Interimstaging auf die Therapie nicht ansprechen. Man überlegt sich, ob man sie möglicherweise mit einem anderen Protokoll frühzeitig weiterbehandelt und dann bessere Ergebnisse erzielt. So haben wir die PETAL-Studie verstanden.

(Andreas Bockisch: Richtig!)

Auch nur diese Fragestellung kann selbstverständlich mit dieser Studie beantwortet werden. Aber innerhalb dieser Fragestellung ist aus unserer Sicht die Diagnostik mittels PET überhaupt die Voraussetzung, um dieses Patientenkollektiv zu identifizieren, dem man dann mit einem veränderten Protokoll helfen kann. Hätten Sie die PET nicht, könnten Sie auch gar nicht entscheiden: Wem lasse ich das neue Protokoll angedeihen und wem nicht?

Ich muss die Patienten zunächst einmal mittels PET identifizieren, nämlich die, die in der Interim-PET nicht ansprechen.

(Zustimmung von Andreas Bockisch)

Nur für diese Gruppe kann ich dann sagen: Ja, mein neues Protokoll funktioniert besser. Hätte ich aber die PET nicht, nützt mir auch mein neues Protokoll nichts. – Insofern haben Sie sozusagen mit dem Nutzen des neuen Protokolls, den Sie in dieser Studie möglicherweise nachweisen, auch automatisch den Nutzen der PET-Diagnostik nachgewiesen.

Wir finden die Studie ganz toll; das möchte ich nur einmal sagen. Wir sind absolut begeistert, dass es solche Studien auch in Deutschland gibt. Es gibt ja sehr wenige dieser Studien weltweit, und es gibt eine solche Studie in Deutschland.

Wir wollen nicht – um Gottes willen, verstehen Sie uns nicht falsch – die PETAL-Studie irgendwie kritisieren. Wir wollen nur sagen: Sie machen ja diese Studien. Wie können Sie dann auf der anderen Seite in der Stellungnahme sagen, das wäre grundsätzlich abzulehnen?

**Andreas Bockisch:** Das ist eine Frage der Betrachtungsweise. Sie sagen völlig zu Recht: Die PET ist die Voraussetzung für die gesteuerte Therapie. Das ist überhaupt keine Frage. Das heißt, als Studienergebnis kann ja nur herauskommen, dass trotz der PET die Therapie eventuell nicht wirkt. Dann wäre im Umkehrschluss für mich auch klar, dass ein PET mit dieser Frage dann keine Indikation mehr für die PET ist, weil dann genau diese Therapie nicht hilft. Trotzdem hat natürlich die PET weiterhin ihren hohen Stellenwert, weil sie Patienten als Nonresponder identifiziert. Das ist ein hoher Wert. Nur haben Sie dann keine Therapie mehr.

**Moderator Stefan Lange:** Herr Scheibler.

**Fülöp Scheibler:** Ich gehe absolut d'accord, wobei ich noch ergänzen möchte: Wenn Sie bei dieser Studie keinen Effekt finden, ist das im Umkehrschluss noch kein Beweis für den Nichtnutzen der PET als diagnostisches Verfahren. Es könnte ja sein, dass ein anderes Protokoll, das später einmal entwickelt wird, aufgrund der PET-Diagnostik funktioniert.

Dieses Design eignet sich aus meiner Sicht für Folgendes: Wenn Sie einen Effekt der neuen Therapie zeigen, haben Sie damit zwei Sachen gleichzeitig bewiesen. Sie haben bewiesen, dass die PET als diagnostisches Verfahren nützlich ist, und Sie haben bewiesen, dass das neue Protokoll wirkt.

Wenn Sie allerdings keinen Zusatznutzen finden, wissen Sie nicht, woran das lag. Ist die PET nicht geeignet als diagnostisches Verfahren? Oder ist sie geeignet, aber die Chemotherapie ist nicht geeignet? Das ist das Problem. Aber wenn Sie Nutzen zeigen, dann haben Sie beides belegt.

**Andreas Bockisch:** Deshalb sind wir der Ansicht, dass die automatische Beweisführung über die gesamte Kette inadäquat ist. Man muss es schrittweise machen. Man hat ja nicht angefangen und gesagt: Wir prüfen einmal, ob PET und die Therapie funktionieren, sondern es steht auch in dem Protokoll, dass die PET eine etablierte Methode ist, die in der Lage ist, Tumurvitalität nachzuweisen und ein Therapieversagen zu erkennen. Das ist sozusagen schon gebongt. Dann kommt der nächste Schritt: Finden wir eine Therapie, die gruppenabhängig Responder/Nonresponder zum Erfolg führt?

**Moderator Stefan Lange:** Vielleicht darf ich da als Moderator einhaken. Ich stimme Ihnen insofern nicht ganz zu, als Sie sagen, das habe dann schon einen Wert an sich. Die Information an sich hat erst einmal gar keinen Wert, insofern sie keine Konsequenz hat. Ich glaube, darüber ist man sich in der Diagnostik auch schon seit vielen Jahren eigentlich einig. Es muss eine Konsequenz geben.

Natürlich kann die Konsequenz sein: Ich kann auch auf Therapie verzichten, um zum Beispiel Lebensqualität durch eine aggressive Therapie nicht einzuschränken. Aber auch das ist natürlich prüfbar. Dafür gibt es übrigens auch Beispiele, die wir jetzt auch bei dem Lymphomprojekt gefunden haben, nämlich Studien, die sich genau diese Fragen stellen: Kann ich auf Therapie verzichten? – Auch in anderen Bereichen wie zum Beispiel im Kopf-Hals-Bereich gibt es entsprechende Studien, die da aufgelegt sind.

Aber grundsätzlich habe ich den Eindruck, dass wir möglicherweise gar nicht so weit auseinanderliegen, sondern dass es unter Umständen auch ein semantisches Problem ist. Aber, wenn ich Sie richtig verstanden habe, sagen Sie schon: Zur Bewertung der PET gehört letztlich auch, was danach folgt. – Oder ist das falsch? Sehen Sie das nicht so? Habe ich Sie da falsch interpretiert?

**Andreas Bockisch:** Wie Sie wissen, ist Medizin äußerst vielschichtig. Ich tue mich äußerst schwer, die Dinge auf so einen Satz zu fokussieren. Ich bleibe dabei: Natürlich ist die Diagnostik als das Ideal der Wahrheit an sich kein Wert, wenn ich es für mich behalte. Wenn ich aber ein Patient bin und man mir sagt „Diese Therapie versagt, und wir haben auch keine Idee, was wir sonst mit dir machen sollen“, ist das aus meiner Sicht ein entscheidender Wert. Dann braucht man nicht mehr diese oder jene Therapie zu versuchen.

Man kann in den Testverfahren nicht ein Glied aus der Kette herausnehmen und an dem festmachen, dass die gesamte Kette funktionieren muss. Sie sagten richtig; das sehe ich ebenso: Wenn man das Ganze synthetisch aufbaut ... – Wir haben eine Diagnostik, die ganz klar die Prognose vorhersagt, was bei den Patienten passiert, und wir wissen dann, dass es

eine Gruppe von Patienten gibt, um die man sich intensiver, auch mit der Akzeptanz von Nebenwirkungen, weiter kümmern muss. Wenn dann in der Tat heute vielleicht herauskommt – das wird herauskommen; ich bin optimistisch –, also falls es keine adäquate Therapie geben sollte, hat man trotzdem die Gruppen aufgetrennt in die, denen man, wenn man klinische Zweifel hat, heutzutage mehr Toxizität zumuten würde. Das braucht man eben nicht.

Es hilft nichts, wenn man den Patienten vor sich hat und sagt: Na ja, eigentlich weiß ich nicht so recht. – In der Medizin ist es ein hoher Wert, wenn der Therapeut eine hohe Sicherheit ausstrahlt.

**Fülöp Scheibler:** Aber auch in dieser Palliativsituation, wenn ich Sie richtig verstanden habe, besteht das Problem, dass man ja möglicherweise mit dem neuen Verfahren falsch positiv diagnostiziert hat, dass man sozusagen im guten Glauben und in einer vermeintlichen Sicherheit dem Patienten unnötige Therapie erspart.

Aber mit der konventionellen Diagnostik wäre dieser Patient nicht positiv geworden. Man hätte ihn weitertherapiert. Möglicherweise hätte nur einer von hundert oder einer von zehn dann doch von der konventionellen Diagnostik plus der konventionellen Therapie profitiert. Aufgrund eines falsch positiven Befundes der PET „erspart man ihm“ – in Anführungszeichen – die weitere Therapie und sagt: Ihnen können wir nicht mehr helfen, da sind wir uns ganz sicher; die PET hat ein positives Ergebnis geliefert. – Ich finde diesen Fall äußerst tragisch.

Wenn man sich zum Beispiel die NOPR-Studie ansieht, stellt man fest: 30 % der Patienten werden aufgrund der PET upgestaged. Sind das wirklich alle Patienten, die dramatisch schlechter dastehen als mit CT? Oder sind vielleicht falsch positive Befunde dabei?

**Andreas Bockisch:** Damit sind wir wieder bei der Frage nach der Qualität der Diagnostik. Da bestimmt man die Kennzahlen: positiv, prädiktiver Wert, negativ, Sensitivität, Spezifität – das kennt man ja. Das ist auch kein Problem der PET.

Im Prinzip ist die entscheidende Aufgabe, dass man als Diagnostiker studienmäßig zeigen muss: Wie ist die absolute Qualität? Und wie ist die Qualität relativ zu anderen verfügbaren Methoden?

(Zustimmung von Fülöp Scheibler)

Da kann man jetzt speziell bei einem Lymphom sagen: Die Wahrscheinlichkeit, dass wir einen falsch positiven Befund in der Interims-PET erheben, ist äußerst klein. – Im Verhältnis zum CT ist es nahezu gewiss, dass Sie einen falsch positiven Befund erheben.

**Robert Großelfinger:** Da möchte ich einmal nachhaken. Irgendwoher muss sich diese Gewissheit speisen. Wenn man bei der von Ihnen angesprochenen diagnostischen Testgüte



stehen bleibt – Sensitivität, Spezifität, prädiktive Werte –, wird man nicht herausbekommen, ob man nicht doch falsch positive Ergebnisse hat. Man muss letztlich die Intervention – in diesem Fall die Nichtbehandlung – bis zum Ende nachverfolgen, oder wie sehen Sie das?

**Andreas Bockisch:** Es gibt reichlich Studien, in denen einfach die Untersuchung gemacht wurde, ohne dass sich nachher etwas geändert hat. Dann sieht man, wer recht hat. Es gibt Überlebenskurven für PET-Positive und PET-Negative.

**Moderator Stefan Lange:** Herr Kotzerke.

**Jörg Kotzerke:** Ich möchte zu dem Aspekt Stellung nehmen, dass üblicherweise immer die Therapie zur Diagnostik mitbetrachtet werden muss. Grundsätzlich sehen wir es auch so, dass, wenn das PET die Therapie nicht mehr beeinflussen kann, weil der Patient sagt, „Ich lehne jede weitere Maßnahme ab“, auch unsere Diagnostik nicht mehr indiziert ist. Daraus speist sich ja die rechtfertigende Indikation. Aber wenn wir ebendiese rechtfertigende Indikation haben, dass von dieser Diagnostik die weitere Behandlung beeinflusst werden kann oder der Patient weitere Informationen über seine Prognose erhält, würden wir diese Untersuchung durchführen.

(Zustimmung von Moderator Stefan Lange)

Ein Aspekt beim Lymphom – dazu sind wir dann gar nicht mehr gekommen – lautet: Wie ist denn das nachher mit der Stammzelltransplantation bei den Patienten, die eben doch einen recht schweren Verlauf der Erkrankung haben, die rezidiert haben und denen dann Stammzellen transplantiert werden? Man weiß, dass diejenigen, die vor der Transplantation noch FDG-positive Residuen haben, einen schlechteren oder einen ungünstigen Verlauf haben. Trotzdem hat man, weil man weiß, dass es sich um Risikopatienten handelt, die Möglichkeit, frühzeitiger wieder zu schauen und möglicherweise ergänzend noch zum Beispiel durch eine lokale Bestrahlung etwas für den Patienten zu tun. Insofern nimmt das FDG-PET immer wieder Einfluss auf den Behandlungsplan.

**Moderator Stefan Lange:** Okay. Ich glaube, das haben wir erst einmal ausgetauscht, obwohl ich nicht ganz sicher bin, ob wir im Gegensatz zu meiner eben geäußerten Einschätzung tatsächlich ein gemeinsames Verständnis haben, wenn ich das noch aufgreifen darf, Herr Kotzerke, was Sie gerade gesagt haben.

Die Frage, ob das PET Einfluss auf den weiteren Behandlungsverlauf nehmen kann, setzt voraus, dass es verschiedene Therapiealternativen gibt und dass diese Therapiealternativen in Abhängigkeit vom PET-Ergebnis unterschiedlich gut funktionieren. Zum Beispiel stellt – das haben Sie gerade genannt – eine zusätzliche Bestrahlung im Rahmen der Stammzelltransplantation eine Intensivierung der Therapie dar. Dann müsste man wissen, ob wohl diese Bestrahlung dann tatsächlich bei diesem Kollektiv wieder mit einem Benefit

verknüpft ist wie längeres Überleben. Das ist letztlich auch in gewisser Weise eine Fragestellung der PETAL-Studie.

Ich erlaube mir trotzdem, noch eine Frage zu stellen. Herr Bockisch, Sie haben gerade gesagt, eigentlich wollten Sie das nicht so gerne auf einen Satz verkürzen. Aber Sie haben das in Ihrer Stellungnahme so getan. Sie lehnten es grundsätzlich ab, schreiben Sie dort. Wie stehen Sie denn dann zu internationalen Empfehlungen von Leitliniengruppen? Da gibt es die GRADE-Gruppe, die im Augenblick eigentlich die weltweit angesehenste Gruppe ist. Sie macht sich Gedanken, wie man nun zu Empfehlungen in Leitlinien kommt. Sie hat sich speziell im diagnostischen Bereich sehr dezidiert geäußert und gesagt: Nein, Accuracy ist ein Surrogat; das reicht nicht. Wir müssen auch bei Diagnostik zeigen können – wie auch immer –, dass patientenrelevante Outcomes verbessert werden. – Das wurde im Jahr 2007 im „BMJ“ veröffentlicht. Das kann man nachlesen.

Wie stellen sich die Deutsche Gesellschaft für Nuklearmedizin und Sie persönlich als Nuklearmediziner dazu, dass eigentlich weltweit gesagt wird: Wir müssen da einen gewissen Paradigmenwechsel vornehmen? – Aber Sie lehnen das grundsätzlich ab.

**Andreas Bockisch:** Zunächst muss ich sagen, dass Sie zwar dieses Zitat haben, das ich nicht anzweifle, ich allerdings mit zwei Tagen Vorbereitung auf die Tagesordnung sicherlich nicht vorbereitet sein kann, um jetzt den ganzen Kontext dieses Zitates zu erfassen.

(Zustimmung von Moderator Stefan Lange)

Aus meiner Sicht ist die Sache zunächst einmal für mich so weit klar: Ich muss dafür sorgen, dass meine Methoden charakterisiert und so präzise wie möglich sind.

(Zustimmung von Moderator Stefan Lange)

Ich gehe auch d'accord, dass aus der Anwendung der Diagnostik etwas folgen muss. Ich gehe nicht d'accord, dass der Diagnostiker beweisen muss, dass, nachdem der Therapeut infolge der Diagnostik etwas unternommen hat, das Endergebnis für den Patienten besser ist.

Für mich ist eine klare Aussage – egal, ob mit oder ohne PET, mit oder ohne CT bzw. Sonografie; es ist völlig egal, was man macht –, dass der Therapeut dann anders agiert: abhängig vom Ergebnis. Das ist das, was wir zeigen müssen.

(Zustimmung von Moderator Stefan Lange)

Ob der Therapeut richtigliegt, das ist ein anderes Thema. Er muss beweisen, dass er richtig therapiert.

**Moderator Stefan Lange:** Gut. Es gibt noch Zitate auch aus Ihren Fachkreisen. Herr Hutchings zum Beispiel hat 2009 sehr dezidiert gesagt: Wir wissen, dass PET beim Lymphom

gute prognostische Informationen liefert. Wir wissen aber nicht, ob das mit einem Benefit für die Patienten verbunden ist. Deshalb brauchen wir klinische Studien, die das zeigen. – Das hat ein Nuklearmediziner sehr aktuell geschrieben.

**Andreas Bockisch:** Ich freue mich, dass Sie die Expertenmeinung so hoch einschätzen.

**Moderator Stefan Lange:** Nein, nicht Expertenmeinung. Ich komme darauf nur aufgrund der Frage der grundsätzlichen Ablehnung. Ich glaube, wir bekommen das nicht so richtig aufgedrösel.

### 4.3.3 Tagesordnungspunkt 2

#### **Zielgrößen in Diagnosestudien: Ist die Inter-Rater-Variabilität ein eigenständiger Aspekt der Testgüte?**

**Robert Großelfinger:** Dieser Tagesordnungspunkt geht auf eine Stellungnahme zurück, in der vorgeschlagen wird, dass für den Teil unseres geplanten Berichts, in dem wir die diagnostische Güte betrachten wollen, zusätzlich auch die Variabilität, mit der die diagnostische Güte erhoben wird, als Kriterium dafür verwendet werden soll, ob ein Testverfahren A besser ist als das Testverfahren B, wenn es beispielsweise eine geringere Inter-Observer-Variabilität aufweist als das andere Verfahren.

Dies war zunächst einmal ein interessanter Gedanke. Unsere Frage lautet in diesem Zusammenhang: Woraus leitet sich denn bei unterschiedlichen Inter-Observer-Variabilitäten der wirkliche Vorteil des einen Testverfahrens gegenüber dem anderen Testverfahren ab, wenn man dabei die Patientensicht vor Augen hat?

**Moderator Stefan Lange:** Vielleicht darf ich ergänzen: Die Frage richtet sich zunächst einmal an Herrn Keller, der das in seiner Stellungnahme formuliert hat. Sie schreiben, zwei Verfahren könnten zwar eine gleiche diagnostische Güte, jedoch eine unterschiedliche Interbeobachtervariabilität aufweisen. Im Falle einer niedrigeren Beobachterübereinstimmung wäre ein Verfahren als schlechter zu beurteilen. Inwiefern spielt tatsächlich bei gleicher diagnostischer Güte – das ist die entscheidende Voraussetzung dabei –

(Thomas Keller: Ja!)

noch die Variabilität eine Rolle? Das haben wir noch nicht so richtig verstanden.

**Thomas Keller:** Kurz zum Hintergrund: Ich habe das aus rein persönlichem Interesse gemacht, weil ich mitbekommen habe, dass sich das IQWiG mit diagnostischen Fragen beschäftigt. Ich bin auf Diagnosestudien spezialisiert, allerdings mehr im IvD-Bereich.

**Moderator Stefan Lange:** In welchem Bereich?

**Thomas Keller:** In-vitro-Diagnostika.

**Moderator Stefan Lange:** Ach so.

**Thomas Keller:** Ich habe aber quasi Diagnosestudien ausgewertet, die sich mit pathologischen Befunden beschäftigt haben. Dabei gibt es eben eine erschreckende Variabilität zwischen den Beobachtern. Aus dem Grund ist mir klar, dass das ein Thema ist.

Mein Fokus findet sich ganz zum Schluss meiner Stellungnahme: Es soll überhaupt die Inter-Observer-Variabilität in dem Bericht auftauchen, weil sie aus meiner Sicht momentan schlicht fehlt. Studien, in denen über die Inter-Observer-Variabilität eine Aussage getroffen wird,

sollen einen höheren Stellenwert bekommen. Das ist das, was ich gewissermaßen angeregt habe.

Als Beispiel hatte ich das Statement abgegeben, nach dem Sie jetzt fragen. Dazu würde ich jetzt Folgendes sagen: Es gibt bei allen diagnostischen Verfahren – ich kann das aus dem Bereich der klinischen Chemie sagen – sehr starke Bemühungen zur Standardisierung. Darauf zielt das letztlich ab.

In dem Augenblick, in dem eben ein Observer zur Aussage kommt, „Der Zustand des Patienten hat sich verschlechtert“, und in dem ein anderer zu dem Schluss kommt, er habe sich nicht verschlechtert, resultieren daraus unterschiedliche Maßnahmen. Je höher und je stärker die Variabilität auseinandergeht, desto stärker werden auch die Maßnahmen variieren.

(Zustimmung von Fülöp Scheibler)

Es gibt übrigens dazu zwei Artikel – ich habe sie in meiner Stellungnahme nicht zitiert, weil ich mich darum noch nicht gekümmert hatte –, die sich wiederum mit der PET bei Lymphomen beschäftigen, und zwar zum Thema Standardisierung der Interpretation. Das ist zum einen der Artikel von Zeilstra 2007 [Buchbeitrag in NUCLEAR MEDICINE AND MOLECULAR IMAGING Annual Report 2007] und zum anderen jetzt ganz aktuell der Beitrag im „Journal of Nuclear Medicine“, was wahrscheinlich auch ein gutes Journal ist, von Hofman aus London, der im Oktober 2009 [Michael S. Hofman, Nigel C. Smeeton, Sheila C. Rankin, Tom Nunan and Michael J. O’Doherty. Observer Variation in Interpreting 18F-FDG PET/CT Findings for Lymphoma Staging Journal of Nuclear Medicine Vol. 50 No. 10 1594-1597] erschienen ist. Dort wird darauf eingegangen. Das zweite Paper ist insofern interessant, als nämlich dort die Inter-Observer-Variabilität der Diagnostik wieder heruntergebrochen wird auf die Schlussfolgerungen. Dann wird gesagt, es gebe zwar die Variabilität innerhalb der Diagnostik, aber bei den Schlussfolgerungen unterschieden die sich eben weniger.

**Moderator Stefan Lange:** Vielleicht trotzdem noch einmal, weil ich Folgendes nicht verstehe: Wie kann es denn sein, wenn tatsächlich so eine unterschiedliche Variabilität besteht, dass die diagnostische Güte die gleiche ist? Ich würde doch annehmen, dass, wenn die Variabilität hoch ist, ich bei einem Verfahren zum Beispiel im Vergleich zu einem anderen auch schlechtere Gütekriterien habe, also eine schlechtere Sensitivität bzw. Spezifität. Dann würde sich doch diese schlechtere Variabilität in der schlechteren Güte ausdrücken.

(Thomas Keller: Ja, das ist richtig!)

Da wir unseren Fokus – vielleicht darf ich das sagen – auf die Güte legen, haben wir in der Bewertung der Güte quasi diesen Aspekt schon mit drin.

**Thomas Keller:** Es kann aber sein, dass ein Verfahren, wenn es zum Beispiel von einem Experten allein durchgeführt wird, eigentlich eine sehr hohe diagnostische Güte hätte. Weil

aber das Expertenwissen sehr stark in diese Diagnostik eingeht, kann es sein, dass, wenn die Inter-Observer-Variabilität einbezogen werden würde, die gesamte diagnostische Güte abfallen würde. In dem Augenblick hätte man dann quasi die gleiche diagnostische Güte.

**Moderator Stefan Lange:** Das habe ich verstanden. – Guido Skipka.

**Guido Skipka:** Dieser Punkt ist wirklich sehr interessant. Ich glaube auch, dass das passieren kann. Nehmen wir jetzt beispielsweise einmal die Sensitivität. Man hat mit dem einen Verfahren eine Sensitivität von 80 % – unabhängig von dieser Indikation jetzt einmal als Beispiel. Das ist auch sehr präzise zwischen den Ratern. Mit einem anderen Verfahren kommt man auch auf 80 %, allerdings mit einer großen Varianz. Dabei würde ich nicht sagen, dass prinzipiell erst einmal das präzisere Verfahren besser ist. Dann als Statistiker ist man immer zufrieden, wenn die Varianzen klein sind. Das ist in der Regel besser.

Nur hier ist die Situation meines Erachtens ein bisschen anders als üblicherweise, wenn man sich die Varianz von Schätzern anschaut, wo man möglichst nahe an der Wahrheit sein will: je präziser, desto besser. Hier ist die Wahrheit sozusagen die Eins. Man liegt ja auf jeden Fall darunter. Man möchte möglichst, um beim Bild zu bleiben, hochkommen.

Wenn ein Verfahren sehr klein, um die 80 %, streut, ein anderes aber stärker, sagt mir doch dieses andere Verfahren: Es gibt offensichtlich Rater, die sehr gut sind, und es gibt Rater, die viel schlechter sind. – Da würde ich sagen: Das ist erst einmal kein Nachteil des Verfahrens, sondern es gibt mir einen Hinweis, einmal nachzuschauen, warum es gute und schlechte Rater gibt. Oder ist das einfach Zufall? Dann wäre es im Mittel wahrscheinlich für den Patienten wieder egal, was man macht.

**Moderator Stefan Lange:** Ich möchte Folgendes noch nachschieben. Wenn man eine hohe Varianz mit Ratern hat, die auch nach oben hin gut abschneiden, also vielleicht sogar tatsächlich eine persönlich höhere Sensitivität entwickeln – nehmen wir als Beispiel 80 %, die dann tatsächlich eine Sensitivität von 90 oder 95 % hätten bei gleich bleibender Spezifität –, würde einen das doch eher dazu bringen zu versuchen zu sagen: Das ist eigentlich ein gutes Verfahren. Ich muss die anderen dazu bringen, innerhalb dieses Verfahrens genauso gut zu werden.

**Thomas Keller:** Meistens ist es natürlich eben nicht so, dass sie die gleiche Spezifität haben.

**Moderator Stefan Lange:** Genau.

**Thomas Keller:** Sondern sie sind praktisch entlang der ROC-Kurve.

**Moderator Stefan Lange:** Aber dann ist die Voraussetzung, die Sie beschrieben haben, die Güte sei gleich, natürlich nicht gegeben.

**Thomas Keller:** Nein, das ist der Overall.

**Moderator Stefan Lange:** Bitte?

**Thomas Keller:** Das ist praktisch der Mittelwert von allen Observern. Darauf bezieht sich das.

**Moderator Stefan Lange:** Klar, das habe ich schon verstanden. Der Patient weiß zunächst einmal nicht, zu wem er geht. Wir haben ein Verfahren X und ein Verfahren Y. Bei gleicher Güte hat Verfahren X die geringere und Verfahren Y die höhere Variabilität. Die Güte ist im Mittel gleich. Dann weiß der Patient nicht, wohin er bei beiden Verfahren geht und ob er zu einem guten oder einem schlechten Rater kommt.

(Thomas Keller: Ja!)

Dann kommt er hinterher mit der Information heraus bei gegebener Sensitivität und Spezifität im Mittel, die zu soundso viel Prozent falsch positiv oder falsch negativ sein kann. Das ist dann für beide Verfahren gleich.

**Thomas Keller:** Vielleicht kann man das auch noch anders interpretieren. Ich vermute, dass sich auch die Radiologen wahrscheinlich eher nicht in ihrer Diagnostik verstanden fühlen, weil wir hier natürlich immer von einer binären Entscheidung ausgehen.

(Zustimmung von Moderator Stefan Lange)

Die Diagnostik ist natürlich, wenn man so will, eine Grauzone oder eine kontinuierliche Diagnostik. Bei einer geringen Variabilität müsste dann die Entscheidung zwischen zwei näherliegenden Therapieoptionen bzw. bei sehr starker Variabilität und sehr weit auseinanderliegenden Unterschieden bzw. weit auseinanderliegenden Therapieentscheidungen gefällt werden. Ich würde annehmen, dass im ersten Fall, den ich beschrieben habe, der Patientennutzen eher herauszufinden ist als im zweiten Fall.

**Moderator Stefan Lange:** Herr Bockisch.

**Andreas Bockisch:** Ich denke, dass das ein sehr interessanter Aspekt ist – egal, welche Konsequenzen es für Ihre Arbeit hat.

(Zustimmung von Moderator Stefan Lange)

Bekannt ist, dass jegliche Studie ein Modell darstellt. Das ist nicht die Realität. Man sucht sich bestimmte Patienten mit Ein- und Ausschlusskriterien. Wer eine Studie macht – davon darf man ausgehen –, ist oberhalb des Durchschnitts, sodass die Ergebnisse, die in einer Studie erzielt werden, nicht zwingend auf die Allgemeinheit übertragen werden können. Das ist natürlich schon entscheidend.

Ich denke, dass die PET in Bezug darauf besonders gut aussieht, wie der Abfall zwischen der Qualität in der Studie im Verhältnis zum Routinebetrieb ist. Das, so denke ich, ohne dass ich mich auf statistische Worte kaprizieren möchte, reflektiert letztlich, was Sie sagen, also: Wo eine hohe Variabilität ist, wird wahrscheinlich der Verlust in der praktischen Anwendung besonders hoch sein.

**Robert Großelfinger:** Mir scheint, dass es dabei zwei Betrachtungsweisen gibt. Ich würde Ihnen recht geben: Wenn ich Testentwickler bin und in meinem Portfolio zwei Test-Kits mit unterschiedlichen Variabilitäten bei gleicher Testgüte habe, kann man schon so argumentieren: Okay, der Test mit der höheren Variabilität hat gegebenenfalls, wenn symmetrisch um die Testgüte verteilt wäre, das größere Potenzial, wenn es gelingt, wie Herr Lange schon angedeutet hat, die Variabilität nach unten, zum Schlechteren hin, irgendwie zu beseitigen.

Aber wenn man zurückkommt auf den Fokus, den wir in unseren Berichten haben – was hat das für Konsequenzen für den Patienten und welche Diagnose bekommt er? –, kommt man an der Tatsache nicht vorbei, dass die mittlere Testgüte bei beiden Verfahren gleich ist und sich für die Patienten letztlich doch nichts ändert.

(Thomas Keller: Binär betrachtet?)

– Binär betrachtet.

**Moderator Stefan Lange:** Guido Skipka.

**Guido Skipka:** Ich meine, dass man diese Betrachtung auch sofort zum Beispiel auf ROC-Kurven übertragen kann, wo man so etwas Quantitatives hat. Man hat ROC-Kurven zwischen den Ratern, die sehr eng beieinanderliegen. Es gibt ROC-Kurven, die auseinanderstreuen. Meiner Meinung nach hat das nichts mit binär oder stetig zu tun, oder?

**Thomas Keller:** Mir ging es um die Entscheidung, die vom Arzt aufgrund der Diagnose getroffen wird. Bei näher zusammenliegenden Ratern wird sich die Entscheidung weniger unterscheiden, als ...

**Moderator Stefan Lange:** Da sehe ich ein Missverständnis, weil Sie sagen, es gäbe keine binäre Entscheidung. Nein, meines Erachtens gibt es immer den Punkt, an dem man sich zwischen zwei Alternativen entscheiden muss. Darauf kann man sich nicht zurückziehen. Entweder wähle ich Strategie A oder Strategie B. Man kann auch sagen: Ich wähle Strategie A oder nicht Strategie A, aber auch das ist eine binäre Entscheidung anhand irgendeines Befundes oder einer Entscheidungshilfe, die ich zugrunde lege.

Es passiert häufig in der Medizin, dass man sagt, die Welt sei nicht schwarz-weiß, sondern so ein bisschen grau. Das wird eher ein bisschen als Rückzugsmöglichkeit genutzt, um sich eben nicht entscheiden zu müssen. Aber es hilft nichts: Irgendwann müssen Sie sich entscheiden.



Die Binarität ist irgendwann da. In der Situation sind wir übrigens auch – davon einmal abgesehen. Aber das ist letztlich unsere Bewertungsgrundlage. So funktioniert auch das Konzept von Testgenauigkeit, über das wir jetzt sprechen, bzw. besteht bei ROC-Kurven natürlich die gleiche Situation. Nur der Test ist eben unterschiedlich grau. Aber die Entscheidung ist klar: Sie ist tatsächlich schwarz oder weiß.

**Thomas Keller:** In dieser einen Arbeit sind quasi direkt die Level bzw. die gemessenen Daten angegeben. 4 verschiedene Stadien werden darin gefunden. Sie variieren zwischen den Observern von 1 bis 4. Ich vermute – ich bin in dem Feld nicht zu Hause –, dass es, wenn das nur zwischen 2 und 3 variieren würde bzw. wenn es zwischen 1 und 4 variieren würde, im Outcome etwas Unterschiedliches ergibt.

**Moderator Stefan Lange:** Da stimmen Ihnen sicherlich alle zu. Aber noch einmal: Ihre Voraussetzung war: Ich habe zwei Verfahren, die im Grunde genommen die gleiche Güte aufweisen, also die Patienten „gleich gut“ – in Anführungszeichen – richtig positiv und richtig negativ identifizieren.

(Thomas Keller: Ja!)

Daher muss sich das in irgendeiner Weise ausgleichen. Ich gestehe, dass ich nicht glaube, dass es eine solche Situation gibt. Wenn das so ist, wie Sie schildern, und wenn die Variabilität so unterschiedlich ist, wird auch die Güte schlechter sein.

(Thomas Keller: Richtig!)

– Gut, dann stimmt aber Ihre Voraussetzung nicht mehr. Das sehen wir dann in der schlechteren Güte.

**Thomas Keller:** Ja, das ist richtig. In der Regel wird es natürlich so sein.

**Moderator Stefan Lange:** Okay.

**Thomas Keller:** In dem Bericht wird am Anfang kurz darauf eingegangen, dass es bei der Interpretation Probleme gibt. Das taucht dann aber praktisch ...

**Moderator Stefan Lange:** Nicht mehr auf.

**Thomas Keller:** ... nicht mehr auf. Aus meiner Sicht müsste das später in dem Teil „Diagnostik“, der ja sowieso nachgestellt ist, dem eigentlichen Nutzteileil, auftauchen.

**Fülöp Scheibler:** Ich glaube, der Aspekt, den Herr Prof. Bockisch vorhin eingebracht hat, ist durchaus relevant für unsere Schlussfolgerung.

**Moderator Stefan Lange:** Klar.

**Fülöp Scheibler:** Nehmen wir folgenden Fall an: Es gibt zwei Tests, die bei gleicher Sensitivität und Spezifität unterschiedliche Variabilität haben. Dann wäre eine gute Erklärung für diese unterschiedliche Variabilität tatsächlich die Qualifikation der unterschiedlichen Untersucher.

Wenn wir davon ausgehen, dass diese Studien tatsächlich eher im universitären oder im wissenschaftlichen Kontext durchgeführt wurden, aber wir jetzt eine Empfehlung für den Versorgungsalltag aussprechen müssen, ist die Wahrscheinlichkeit bei einem Test mit höherer Variabilität natürlich größer, dass er im Versorgungsalltag schlechter abschneidet. Deswegen würde ich aufgrund eines solchen Tests unsere Empfehlung etwas vorsichtiger aussprechen.

**Moderator Stefan Lange:** Wie immer bei Variabilität muss man sich ansehen, woran sie liegen könnte. Ich glaube, dass die Frage der Verallgemeinerbarkeit von Studienergebnissen auf die Versorgungssituation eine grundsätzliche Frage ist, die sich bei allen Technologiebewertungen stellt, auch bei der Therapie.

Es ist nicht ganz „ungefährlich“ – das sage ich in Anführungszeichen –, wenn man generell infrage stellt, ob das so ist und ob man das verallgemeinern kann, denn dann bräuchte man diese Studien eigentlich gar nicht zu machen. Dann ist es in gewisser Weise ein Totschlagargument.

Es wird übrigens auch, wenn ich das sagen darf, nicht schon dadurch gelöst, dass man sehr breit versucht, die Versorgungssituation abzubilden, und alles zulässt, denn dann habe ich ein internes Übertragbarkeitsproblem. Dann weiß ich auch nicht, ob ich das, was ich intern gesehen habe, auf alle möglichen Situationen, die ich zugelassen habe, übertragen kann.

Nichtsdestotrotz ist es natürlich ein valider Punkt. Wenn es so ist, dass bestimmte Verfahren nur von bestimmten Experten durchgeführt werden können oder eine starke, große und lange Lernkurve erforderlich ist, gehört es sicherlich zur Bewertung eines Verfahrens dazu. – Haben wir die Unklarheit, die wir bei diesem Punkt hatten, geklärt?

**Robert Großelfinger:** Ich denke, ja.

**Moderator Stefan Lange:** Dann sind wir relativ zügig mit den ersten beiden Tagesordnungspunkten durchgekommen.

#### 4.3.4 Tagesordnungspunkt 3 Verschiedenes

**Moderator Stefan Lange:** Haben Sie noch etwas, das Sie unter „Verschiedenes“ ansprechen möchten? – Herr Krummrey.

**Gert Krummrey:** Nur kurz eine Verständnisfrage: Auf Seite 1, wo Sie über Epidemiologie und Krankheitslast sprechen, tauchen auch die ICDs für Anus und Analkanal auf, die eigentlich nicht zum kolorektalen Karzinom gehören. Das ist ein bisschen verwirrend. Ich gehe nicht davon aus, dass Sie das untersuchen wollen.

**Moderator Stefan Lange:** Das prüfen wir noch einmal. Das können wir checken. Vielen Dank für den Hinweis.

**Gert Krummrey:** Darf ich Ihnen noch eine Verständnisfrage stellen?

**Moderator Stefan Lange:** Selbstverständlich.

**Gert Krummrey:** Ändert die Tatsache, dass die PET bei dieser Indikation im Bereich der funktionellen Bildgebung – Sie nennen das morphologische Bildgebung – ein Alleinstellungsmerkmal hat, weil es zum Beispiel beim malignen Lymphom nicht viele andere funktionelle Untersuchungen gibt, zum Beispiel die Szintigrafie, irgendetwas an der Bewertung? Oder sagen Sie: Weil man jetzt keinen Komparator aus derselben Diagnoseklasse, also aus derselben Gruppe der funktionellen morphologischen bildgebenden Verfahren ...

**Moderator Stefan Lange:** Da frage ich unsere Sachverständigen und unseren Mitarbeiter, Herrn Scheibler.

**Fülöp Scheibler:** Aus meiner Sicht ist das grundsätzlich nicht der Fall, denn sonst könnte man sich irgendein Verfahren ausdenken, das wieder irgendwie eine andere Dimension berücksichtigt. Völlig unabhängig von dessen Nutzen und diagnostischer Güte würde man sagen: Das ist ein Alleinstellungsmerkmal, weil wir zum Beispiel aufgrund von – ich denke mir irgendetwas aus – Aurafotografie das Rezidiv messen. Das gab es noch nie auf der ganzen Welt. Deswegen muss das unbedingt einen höheren Nutzen als alles andere haben. – Das ist jetzt überspitzt formuliert. Insofern sehe ich zunächst einmal aufgrund dieses Alleinstellungsmerkmals noch keinen Zusatznutzen per se.

**Moderator Stefan Lange:** Frage beantwortet?

**Gert Krummrey:** Frage beantwortet.

**Anja Hagen:** D'accord.

**Moderator Stefan Lange:** Damit gehen auch unsere Sachverständigen d'accord, wunderbar. – Herr Keller.

**Thomas Keller:** Es gibt noch einen kleinen Punkt. Unter 4.2.3.2 wird auf den Methodenvergleich kurz eingegangen und gesagt, dass man quasi zwei Methoden vergleichen könne. Es wird zwar im Bereich der Radiologie wahrscheinlich noch keine derartigen Studien geben; aber prinzipiell kann man, wenn man die zwei Methoden vergleicht, auf die Goldstandarddiagnose für eine der beiden Gruppen verzichten, und zwar nach dem sogenannten VOPT-Design, also Verification of Only Positive Tested. Die Größen, die man dann erhält, sind nicht absolute Sensitivitäten und Spezifitäten, sondern relative.

**Moderator Stefan Lange:** Wenn Sie die zwei Verfahren miteinander vergleichen wollen?

**Thomas Keller:** Ja.

**Moderator Stefan Lange:** Völlig klar. D'accord.

**Thomas Keller:** Man kommt damit zu validen Aussagen und kann so das Problem umgehen, dass, wie es hier steht, bei negativen Befunden aus ethischen Gründen kein Goldstandard verfügbar ist.

Also das sollte deswegen vielleicht jetzt nicht hier, aber bei vergleichbaren Papern als Option auftauchen, dass diese relativen diagnostischen Güten angegeben werden können.

**Moderator Stefan Lange:** Wenn ich Sie richtig verstehe, meinen Sie, es reiche aus, im Prinzip die diskordanten zu prüfen.

**Thomas Keller:** Genau.

**Moderator Stefan Lange:** Genau. Zwischen zwei gleichen Verfahren prüft man nur die diskordanten und muss nicht auch die konkordanten im Goldstandard nachuntersuchen.

(Thomas Keller: Ja!)

– Okay. Könnten Sie uns vielleicht, weil Sie dazu jetzt ein Stichwort geliefert haben, eine Literaturstelle nennen? Sie hatten eine Abkürzung genannt. Das hörte sich so an, als ob Sie das mit Literatur belegen könnten.

**Thomas Keller:** In Alonzo und Pepe [Margaret Sullivan Pepe, Todd A. Alonzo. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics*, 2001; 2(3): 249–260.] finden Sie das.

**Moderator Stefan Lange:** Alonzo und Pepe? Wir können ja vielleicht noch einmal auf Sie zurückkommen.

**Thomas Keller:** Ich kenne die englische Aussprache der Namen leider nicht. Aber ich kann Ihnen gern die Zitate schicken.

**Moderator Stefan Lange:** Das wäre sehr nett.

**Robert Großelfinger:** Ich hätte noch eine Rückfrage. Wenn ich es richtig verstanden habe, geht es auch um die Nachverfolgung der Diskordanzen. Wenn es darauf hinausläuft, gebe ich nur den Hinweis, dass Diskordanzstudien im Berichtsplan angelegt sind berücksichtigt werden,

(Thomas Keller: Stimmt!)

falls es darauf abzielen sollte.

**Moderator Stefan Lange:** Wenn Sie uns die Literaturstellen dazu schicken, schauen wir uns gerne an, ob wir da etwas methodisch übersehen haben. – Herr Kotzerke.

**Jörg Kotzerke:** Ich hätte noch einen Wunsch, um die Qualität dieses Evaluationsprozesses ein Stück weit transparenter zu machen. Sie suchen in den unterschiedlichsten Datenbanken nach Informationen. Ich habe mich immer gefragt, welche relevante Information Sie wohl zusätzlich zur Standardrecherche in MEDLINE erhalten. Dieser Aspekt würde mich tatsächlich interessieren.

Wir haben es beim Lymphom gesehen. Als es logging, wurden 400 Artikel identifiziert. Dann wurde immer wieder alles ausgeschlossen. Nachher wurde nur ein einziger Artikel als relevant analysiert. Da hat mich der Verdacht beschlichen, dass entweder die Suchstrategie unglücklich war oder dass nicht das Relevante adressiert werden konnte, wenn ein solches Missverhältnis auch bei diesen Meta-Analysen besteht. Dass Sie das gerne aus HTA-Berichten machen, habe ich jetzt gelernt. Ich würde das ungern aus solchen Berichten machen.

Dass Sie deshalb nach diesen Berichten in anderen Datenbanken forschen, verstehe ich. Aber ich wüsste gerne, ob wir wirklich ein Wissensdefizit haben, wenn wir nur mit unserer Standardliteraturrecherche vorgehen. Ich weiß, dass es bei Ihnen zur Qualität dazugehört, dass Sie das über mehrere Datenbanken machen. Ich hätte gern einmal dargestellt, dass das eben tatsächlich einen zusätzlichen Nutzen hat.

**Moderator Stefan Lange:** Jetzt habe ich Ihre Frage nicht richtig verstanden. Würde Sie interessieren, welche relevanten Informationen man über die Standarddatenbanken wie MEDLINE oder PubMed hinaus durch andere Datenbanken bekommen kann?

**Jörg Kotzerke:** Ja, genau.

**Moderator Stefan Lange:** Okay. Unabhängig von unseren Berichten gibt es tatsächlich Untersuchungen, dass man zum Beispiel durch die Datenbank EMBASE je nach Fragestellung etwa 10 % bis 30 % mehr Treffer und auch relevante Treffer bekommt. Dafür zum Beispiel könnte man das angeben.

Man weiß auch, dass man für einen unerheblichen Prozentsatz durch das Durchsuchen von Referenzen von anderen Arbeiten bzw. systematischen Übersichtsarbeiten auch noch zusätzliche Treffer bekommt, die wiederum je nach Thema unterschiedlich sind, wobei man sich tatsächlich fragen kann, ob das von der Größenordnung her relevant ist. – Ich weiß nicht genau, ob das wirklich Ihre Frage getroffen hat.

**Jörg Kotzerke:** Das hat im Grunde meine Frage getroffen. Dass Sie bei verschiedenen Fragestellungen vielleicht 10 % bis 30 % mehr Informationen haben, dazu kann ich nichts zu sagen. Mich würde mit Blick auf dieses Thema, über das wir miteinander reden, interessieren, wie viele Zusatzinformationen tatsächlich in diesem konkreten Fall angefallen sind oder anfallen werden.

**Fülöp Scheibler:** Wenn Sie wollen, können wir das zum Beispiel für das Lymphomprojekt rückwirkend gern einmal analysieren: Wie viele der gefundenen Treffer zum Beispiel in der Ergänzungsrecherche kamen aus MEDLINE, aus EMBASE oder aus anderen Datenbanken?

Das könnte man zum Beispiel im Rahmen eines Abstracts auf einem Nuklearmedizinikongress tun. Wir haben im letzten Jahr so etwas Ähnliches gemacht. Uns hat Folgendes interessiert: Wenn wir nach HTA-Berichten und systematischen Übersichten suchen, stellen wir fest: Es gibt mittlerweile eine ganze Zahl an Datenbanken, in denen man suchen kann, wie die HTA-Datenbank von Cochrane, das Centre for Reviews and Dissemination oder die HTA-Organisationen, auf deren Homepages man suchen kann. Man kann versuchen, über „Google-Scholar“ zu suchen und so weiter.

Wir waren uns auch nicht sicher. Es ist ein riesiger Aufwand, all diese Quellen zu durchsuchen. Bringt uns das denn tatsächlich so viel mehr relevante Treffer? Wir haben dazu auch letztes Jahr ein Poster publiziert, weil wir zumindest bei den PET-Projekten mittlerweile ziemlich sicher sind, dass wir, wenn wir drei Datenbanken durchsuchen – HTA von Cochrane, DARE und Centre for Reviews and Dissemination –, alle relevanten HTA-Berichte und systematischen Reviews gefunden haben.

So etwas Ähnliches könnte man auch für Primärstudien machen. Das heißt, wir könnten retrospektiv sagen: Welche relevanten Studien haben wir identifiziert? Aufgrund welcher Datenbanksuchen wurden sie denn eigentlich gefunden?

Wir müssen nur aufpassen, denn wir haben auch Duplikate. Wenn Sie 2 Recherchen haben, einmal MEDLINE und einmal EMBASE, und ein und dieselbe Publikation finden, werden diese Duplikate ausgeschlossen. Es ist sozusagen Glücksache: Werfen Sie das Zitat von

EMBASE oder von MEDLINE heraus? Insofern müsste man wahrscheinlich schon die Originalrecherchen miteinander vergleichen. Aber das wäre eine weiterführende wissenschaftliche Fragestellung, die ich auch interessant finde.

**Anja Hagen:** Wobei es beim Rauswerfen meistens so ist, dass man häufig eher die EMBASE-Sachen hinauswirft, weil sie kostenpflichtig bzw. teurer sind, außer wenn wirklich mehr Informationen darin stehen. Es gibt manchmal auch Spezialfälle, in denen das Abstract bei MEDLINE zufälligerweise einmal fehlt. Nach unserer Erfahrung werfen wir eigentlich eher die EMBASE-Sachen heraus. Aber das ist von Institution zu Institution unterschiedlich.

**Moderator Stefan Lange:** Gut. Wir nehmen das gerne als Anregung auf. Herr Kotzerke, vielen Dank. Sie merken natürlich: Das wäre eine zusätzliche Belastung. Wir müssen schauen, wie wir das mit unseren sonstigen Obliegenheiten hinbekommen. Herr Krummrey, als Vertreter einer unserer Träger, wird sicherlich sehr genau zuhören,

(Gert Krummrey: Immer! – Heiterkeit)

dass wir jetzt nicht unseren eigentlichen Aufgaben nachkommen. Wir behalten das im Kopf und können versuchen, das für das abgelaufene und vielleicht auch für die zukünftigen Projekte zu checken.

Das Problem, das Herr Scheibler gerade genannt hatte, kann man adressieren; das bekommt man schon hin. Dann wäre genau die Frage: Wie viel haben wir zusätzlich durch Datenbanken wie EMBASE oder das, was man sonst noch durchsucht, gewonnen? Es ist ein nützlicher Hinweis, dass man das vielleicht einmal beim nuklearmedizinischen Kongress vorstellt. Das ist vielleicht auch nicht verkehrt.

**Robert Großelfinger:** Noch ein Kommentar dazu, weil Sie angesprochen haben, dass dieser Fundus an potenziellen Treffern so unglaublich zusammenschnurrt: Das hat natürlich auch damit zu tun, dass die Suchen eher sensitiv angelegt sind. Man geht so vor, dass man auf keinen Fall etwas übersehen will. Daher legt man die Suche nicht von vornherein spezifisch aus. Aufgrund dessen muss man eben mehr aussortieren, bis man das findet, was man über die Einschlusskriterien vorab formuliert hat.

**Moderator Stefan Lange:** Danke. – Gibt es sonst noch Punkte? – Herr Krummrey.

**Gert Krummrey:** Ganz kurz: Wir hatten uns bei der letzten Anhörung über die Problematik – Sie hatten das auch schon angesprochen – des Referenzstandards unterhalten. Wir haben gebeten – Sie hatten dazu auch nachgefragt –, einen pragmatischen Ansatz zu wählen. Wir hatten den Ansatz des schon zitierten Herrn Hutchings vorgeschlagen: Man kann nicht alles biopsieren. Dann müssen wir halt ein Panel bilden und irgendwie über einen Konsensusprozess sagen: Okay, anhand der Konkordanz oder Diskordanz der Befunde gehen wir im Follow-up einfach davon aus, dass das so ist – wir haben keine Biopsie – oder nicht.

Ich habe es jetzt immer noch nicht ganz verstanden. Würden Sie bei solchen Studien, die so einen Weg wählen, der zum Beispiel von der Cochrane Collaboration oder von Start-Initiative durchaus als gangbar angesehen wird, auch sagen: Das ist ein gangbarer Weg, um die Problematik des Referenzstandards zu umgehen oder einfach damit umzugehen?

**Moderator Stefan Lange:** Damit umzugehen. Man löst sie nicht.

**Gert Krummrey:** Nein, das kann man ja auch nicht. Aber Sie würden die nicht ausschließen, weil Sie sagen „Wir brauchen unbedingt immer eine Biopsie“? Wenn es sie nicht gibt, weil das ethisch oder technisch nicht geht, würden Sie sagen „Das ist ein gangbarer Weg“?

**Moderator Stefan Lange:** Frau Hagen.

**Anja Hagen:** Grundsätzlich ist man damit zwar nicht glücklich, aber man muss einen Kompromiss finden, um die Informationen zu erhalten.

**Moderator Stefan Lange:** Laut Berichtsplan ist das auch so vorgesehen.

**Fülöp Scheibler:** Genau. Wir sagen einfach: der beste vorhandene diagnostische Standard. Ist er aber aus unserer Sicht unzureichend, würden wir das entsprechend diskutieren. Das heißt aber nicht, dass wir die Studien per se ausschließen.

**Moderator Stefan Lange:** Bitte, Herr Kotzerke.

**Jörg Kotzerke:** Ich habe noch eine Bitte. Sie haben vorhin darauf hingewiesen, dass auch Ihre Ressourcen begrenzt sind.

**Moderator Stefan Lange:** Selbst unsere, ja.

(Teilweise Heiterkeit)

**Jörg Kotzerke:** Das sind unsere Ressourcen auch. Wir machen das freiwillig nebenbei. Insofern sind diese Zeitschienen, die Sie uns mitteilen, sehr eng, wenn es darum geht, Werke im Umfang von 200 Seiten zu lesen und adäquat zu bewerten, wofür nur 4 Wochen zur Verfügung stehen. Können wir nicht einen anderen Weg finden, sodass wir frühzeitiger in diesen Prozess eingebunden werden?

**Moderator Stefan Lange:** Da muss ich leider auf den Anfang dieser Erörterung und darauf verweisen, dass wir das nicht grundsätzlich diskutieren können. Wir haben unsere Verfahrensweisen, die im Grunde genommen auch mit anderen Institutionen vergleichbar sind. Sie sind sogar eher großzügig. Sie finden zum Beispiel auch bei der amerikanischen Agency mitunter Kommentierungsfristen von 2 Wochen. Da liegen wir mit 4 Wochen schon ganz gut.



Wir versuchen, die Menschen an zwei Stellen einzubeziehen: beim Berichtsplan wie jetzt und noch einmal bei der Erstellung des Vorberichtes. Ich wüsste nicht, wie dazwischen noch irgendetwas anderes sein sollte. Es ist klar, dass die Durchführungsverantwortung für die Erstellung dieses Berichtes in diesem Institut liegt. Das ist nun einmal so vorgesehen.

Nichtsdestotrotz müssen und wollen wir aber auch die Sichtweisen anderer interessierter Kreise berücksichtigen. Wir haben sogar drei Stellen:

1. Im Vorfeld der Berichtsplanerstellung nehmen wir mit Patientenvertretern bezüglich patientenrelevanter Endpunkte Kontakt auf.
2. Wir machen die Anhörung im Rahmen der Berichtsplanerstellung.
3. Im Rahmen der Vorberichtserstellung führen wir eine Anhörung zum Vorbericht durch.

Also viel mehr fällt uns, ehrlich gesagt, dazu nicht ein. Dass die Fristen von vielen als recht knapp angesehen werden, wissen wir. Aber wir können es leider nicht ändern. Wir haben sie schon einmal bei sehr ausführlichen, sehr umfangreichen Berichten tatsächlich verlängert. Aber sie sind international gesehen auch nicht besonders kurz, im Gegenteil. Damit können wir leider Ihrer Bitte nicht entsprechen.

Ansonsten schlage ich vor, die Erörterung für heute zu beenden. Wir haben jetzt Gelegenheit, bei einem Kaffee und einem Stück Kuchen noch andere Punkte anzusprechen, die nicht unbedingt Gegenstand dieser Erörterung sind.

Ich bedanke mich für Ihr Kommen und für die interessante Diskussion. Ich vermute, dass wir uns nicht das letzte Mal in dieser Runde oder in dieser Weise gesehen haben. Wir werden wahrscheinlich nicht umhinkommen, uns immer wieder über diesen einen Punkt auseinanderzusetzen und vielleicht auch zu streiten, um dann zu schauen, ob es nicht doch irgendwann zu einem gemeinsamen oder gemeinsameren Verständnis kommen kann. – In diesem Sinne: auf Wiedersehen.

**Anhang: Dokumentation der Stellungnahmen**

# Inhaltsverzeichnis

	Seite
<b>A 1 Stellungnahmen von Organisationen, Institutionen und Firmen .....</b>	<b>A 2</b>
<b>A 1.1    ACOMED statistik .....</b>	<b>A 2</b>
<b>A 1.2    Deutsche Gesellschaft für Nuklearmedizin e. V. ....</b>	<b>A 4</b>
<b>A 1.3    Deutsche Krankenhausgesellschaft .....</b>	<b>A 9</b>
<b>A 2 Stellungnahmen von Privatpersonen.....</b>	<b>A 14</b>

## **A 1 Stellungnahmen von Organisationen, Institutionen und Firmen**

### **A 1.1 ACOMED statistik**

**Autoren:**

Keller, Thomas, Dr.

**Adresse:**

Dr. Thomas Keller  
ACOMED statistik  
Fockestraße 57  
04275 Leipzig

Sehr geehrte Damen und Herren,

bei der Positronen-Emissions-Tomographie handelt es sich um ein diagnostisches Verfahren, bei dem die Subjektivität des Beobachters die diagnostische Güte beeinflusst. Dieser Aspekt durch die Zwischen-Beobachter-Variabilität beschrieben (andere Begriffe sind Beurteiler-Übereinstimmung, Beobachter-Reliabilität).

Die geeignete Einbeziehung der Inter-Beobachter-Variabilität in Diagnosestudien zur Beurteilung bildgebender Verfahren ist Standard. Sie findet in den Zulassungsverfahren Niederschlag. In der Regel folgen diese Studien dem Multi-Reader-Multi-Case design (MRMC-design). Hierzu gibt es umfangreiche Literatur, es sei z.B. auf (Dorfman, Berbaum et al. 1992), (Obuchowski and Zepp 1996), (Zhou 2002) sowie weitere verwiesen. Mittlerweile gibt es auch entsprechende Literatur zu den binären Schätzern (Sensitivität, Spezifität etc.), auf deren Belfügung ich verzichte, da im Bereich der Bildgebung die ROC-Analyse am meisten verbreitet ist.

Diagnostische Verfahren, in denen der subjektive Einfluss eines Beobachters eine Rolle spielt, können eine gleiche (mittlere) diagnostische Güte aufweisen, jedoch eine unterschiedliche Inter-Beobachter-Variabilität. Im Falle einer niedrigeren Beobachterübereinstimmung wäre ein Verfahren als schlechter zu beurteilen.

Die Inter-Beobachter-Variabilität wird entweder als Übereinstimmungsmaß (z.B. Kappa) oder Varianzanteil geschätzt.

Ich möchte vorschlagen, die Zwischen-Beobachter-Variabilität in die Betrachtung der Anwendung der PET in dem Sinn einzubeziehen, dass eine Studie, die neben diesen Aspekt einbezieht, einen höheren Stellenwert (höhere Evidenz) aufweist. Zudem sollte natürlich auch das Ergebnis selbst in die Beurteilung einbezogen werden.

Mein Hinweis gilt auch für die zahlreichen weiteren Untersuchungen zur Positronen-Emissions-Tomographie, die gegenwärtig vom IQWiG vorgenommen werden, wobei anhand der Erkrankungshäufigkeit entschieden werden sollte, inwieweit dies sinnvoll ist.

Mit freundlichen Grüßen



Dr. Th. Keller  
ACOMED statistik  
Fockestr. 57  
D 04275 Leipzig,  
Germany  
[www.acomed-statistik.de](http://www.acomed-statistik.de)  
[tk@acomed-statistik.de](mailto:tk@acomed-statistik.de)  
tel.: 0341 3910195

Dorfman, D. D., K. S. Berbaum, et al. (1992). "Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method." Invest Radiol **27**(9): 723-31.

Obuchowski, N. A. and R. C. Zepp (1996). "Simple steps for improving multiple-reader studies in radiology." AJR Am J Roentgenol **166**(3): 517-21.

Zhou, X., Obuchowski, NA, McClish, DK (2002). Statistical Methods in Diagnostic Medicine. New York, Wiley Interscience. Chapter 9 (277-306)

Obuchowski, NA (2004). How many Observers are needed in Clinical Studies of Medical Imaging? AJR **182** (867-869)

Obuchowski, NA (2007). New Methodological Tools for Multiple Reader ROC studies. Radiology **243** (10-12)

## **A 1.2 Deutsche Gesellschaft für Nuklearmedizin e. V.**

### **Autoren:**

Bockisch, Andreas, Prof. Dr. Dr.

Grünwald, Frank, Prof. Dr.

Kotzerke, Jörg, Prof. Dr.

### **Adresse:**

Prof. Dr. Dr. Andreas Bockisch

Deutsche Gesellschaft für Nuklearmedizin e. V.

Nikolaistraße 29

37073 Göttingen



DGN e.V. c/o vokativ GmbH • Nikolaistr. 29 • D-37073 Göttingen

**Institut für Qualität und Wirtschaftlichkeit im  
Gesundheitswesen**

- Stellungnahme zum Berichtsplan D06-01C  
Prof. Dr. med. Peter. T. Sawicki  
Dillenburger Straße 27  
51105 Köln

DGN e.V.  
Nikolaistr. 29

D - 37073 Göttingen

Tel. (Mitglieder): +49.551.48857-401  
Tel. (Presse): +49.551.48857-402  
Fax: +49.551.48857-79

e-mail: office@nuklearmedizin.de  
Internet: www.nuklearmedizin.de

Ihre Zeichen

Ihre Nachricht vom

Unser Zeichen

Göttingen

29. September 2009

**Stellungnahme der Deutschen Gesellschaft für Nuklearmedizin (DGN) zum Berichtsplan (vorläufige Version) D06-01C (Version 1.0) des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG): Positronenemissionstomographie (PET und PET/CT) bei rezidivierendem kolorektalen Karzinom**

Sehr geehrter Herr Prof. Sawicki,  
sehr geehrte Damen und Herren,

wir bedanken uns für die Gelegenheit, zu dem genannten Berichtsplan in der vorliegenden vorläufigen Version vom 20. August 2009 Stellung nehmen zu können, den Sie zur öffentlichen Diskussion gestellt haben. Für die weitere Bearbeitung des Berichtsplans bitten wir, die folgende Stellungnahme der Deutschen Gesellschaft für Nuklearmedizin zu berücksichtigen.

Wir bedauern, dass trotz der schriftlich und mündlich angemahnten Änderungen der Methodik bereits zum vorläufigen Berichtsplan D06-01B (Stellungnahme der DGN vom 7. Juli 2009 zum vorläufigen Berichtsplan „Positronenemissionstomographie (PET) und PET/CT bei Kopf- und Halstumoren“) die Vorgehensweise hier in unveränderter Form zur Anwendung gebracht wird, was wir erneut als unangemessen zurückweisen.

Konsequenterweise heben wir somit auch in dieser Stellungnahme auf das im Berichtsplan ausgeführte methodische Vorgehen in zwei wesentlichen Punkten ab (Nummerierung analog Berichtsplan):

**2.1 Nutzenbewertung der PET bzw. PET/CT**

Die Verknüpfung von diagnostischer und therapeutischer Strategie zur Bestimmung eines patientenrelevanten Nutzens ist grundsätzlich abzulehnen. Der Nutzen einer diagnostischen Methode ist bereits daran erkennbar, ob sie das therapeutische Management beeinflussen kann.

Die hier vorgeschlagene methodische Vorgehensweise ist im Wesentlichen zur Bewertung von Arzneimitteln geeignet, die in randomisierten Studien getestet wurden. Ausführungen zur Bewertung von diagnostischen Verfahren sind nicht zu finden.

In diesem Zusammenhang ist der „Nutzen“ für den Patienten ebenfalls durch die Möglichkeit gegeben, Fernmetastasen zu erkennen, was eine Änderung der Intention einer Strahlentherapie von „kurativ“ in „palliativ“ zur Folge hat. Die Reduktion der Toxizität mit einer einhergehenden

Verbesserung der Lebensqualität in dieser letzten Lebensphase ist offensichtlich und bedarf keiner Sicherung durch randomisierte Studien.

Darüber hinaus möchten wir auch an dieser Stelle erneut darauf hinweisen, dass die Forderung nach Evidenzklassifizierung des G-BA Stufe I unangemessen ist. Tricoci et al. (JAMA 301:831-41, 2009) untersuchten dabei die klinischen Leitlinien der AHA (American Heart Association) aus den Jahren von 1984 bis September 2008. Es zeigt sich, dass das, was den Empfehlungen entsprechend qualitativ gute klinische Praxis ist, nur in einem extrem geringen Prozentsatz durch Studien mit hohem Evidenzgrad belegt ist. Um dies in Zahlen zu fassen: Von den 16 aktuellen Leitlinien, die einen Evidenzgrad aufführen, sind nur 11% der Empfehlungen mit Evidenz-Level A (Evidenz von mehreren randomisierten Studien oder Metaanalysen) hinterlegt. 41% der Empfehlungen basieren auf Studien des Evidenzlevels B (einzelne randomisierte Studie oder nichtrandomisierte Studien) und 48% auf dem Evidenzlevel C (Expertenmeinungen, klinische Praxis, Fallberichte, Beobachtungsstudien). Das bedeutet, dass in nur 11% der angestrebten und empfohlenen Krankenversorgung tatsächlich Studien eines hohen Evidenzlevels vorliegen.

Nimmt man die höchste Klasse der klinischen Empfehlung Level I (nützlich und wirksam) steigt zwar der Anteil der Arbeiten mit Evidenzklasse A, erreicht aber letztlich nur 18% Anteil. Die klinische Versorgung, die unstrittig als sehr vorteilhaft bewertet wird, beruht damit ganz überwiegend auf Studien und Informationen der Evidenzklasse B und C.

Bei Betrachtung dieser Zahlen kommen erhebliche Zweifel über die Anwendbarkeit ausschließlich des höchsten Evidenzlevels als einziges objektives Maß für den klinischen Nutzen und den Mehrwert für den Patienten in Hinblick auf Diagnostik und Therapiesteuerung.

Es scheint vielmehr umgekehrt einen Bias in der Praxis der klinischen Forschung zu geben, dass prospektiv randomisierte Studien aufgrund des aufwendigen Designs und dem damit verbundenen Aufwand nicht zu den bevorzugten Studienformen der Wissenschaftler gehören, was über den medizinischen Nutzen eines Verfahrens keinerlei Aussage macht. Ein Fehlen von Studien mit hohem Evidenzgrad zeigt daher in keiner Weise ein Fehlen des klinischen Nutzens an, was häufig suggeriert wird.

In der Zusammenfassung lässt sich damit festhalten, dass der Wert eines Verfahrens für die klinische Praxis und die allgemeine Krankenversorgung am besten zu beurteilen ist, wenn analog zur Erstellung einer klinischen Leitlinie vorgegangen wird: Ein interdisziplinäres Expertengremium bewertet den aktuellen Stand des Wissens unter Berücksichtigung aller Evidenzklassen und erstellt dann graduierte Empfehlungen bei realistischen klinischen Szenarien. Dies würde helfen, auch moderne Verfahren in der Krankenversorgung der Bevölkerung verfügbar zu machen.

Ferner weisen wir nochmals darauf hin, dass die methodische Differenzierung zwischen „PET“ und „PET/CT“ nicht mehr zeitgemäß ist, da heute nur noch Kombinationsgeräte hergestellt werden. Der Vorteil dieser methodischen Fortentwicklung liegt nicht ausschließlich in der Möglichkeit, auch eine diagnostische CT-Untersuchung in einer Sitzung durchführen zu können, sondern vor allem darin, dass neben der auch im Berichtsplan angeführten anatomisch vereinfachten Zuordnung der PET-Information auch die Schwächungskorrektur der primären Emissionsmessung durch die CT bei verkürzter Untersuchungszeit sehr präzise, zeitlich in gleich bleibender Qualität und mit einer dem Körpergewicht des Patienten angepassten Strahlungsbelastung erfolgen kann.

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

Die Bewertung der diagnostischen Güte der PET und PET/CT primär auf Grundlage von systematischen Übersichten (Meta-Meta-Analyse) lehnen wir – wie bereits in unseren früheren Stellungnahmen ausführlich dargelegt – entschieden ab (s. Stellungnahme der DGN vom 7. Juli 2009 zum vorläufigen Berichtsplan D06-01B (Version 0.1) des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG): Positronenemissionstomographie (PET) und PET/CT bei Kopf- und Halstumoren):



Bereits bei der Analyse der Positronenemissionstomographie (PET und PET/CT) bei malignen Lymphomen (D06-01A) hat sich gezeigt, dass sich dieses Vorgehen nicht bewährt hat:

*Während des Bearbeitungsprozesses gestaltete es sich als schwierig, den Aufbau und die interne Strukturierung einer jeden Synthese für den vorliegenden Bericht in ein gemeinsames Sortierungsprinzip zu integrieren. ... Zum anderen betraf das die Darstellung der Evidenztabelle in den Evidenzsynthesen, die sich oft nur unvollständig aufbrechen und einzelnen Primärstudien zuordnen ließen. Häufig fehlte der eindeutige Bezug zwischen den Aussagen im Text und den jeweiligen Primärstudien, ... Oft ließ sich zwischen unvollständiger Datenextraktion und fehlender Information in den Primärstudien nicht unterscheiden. Wegen der begrenzten Möglichkeit, Evidenztabelle und Ergebnisberichte eigenständig zu bewerten, erhielten die Bewertungen der Autoren der Evidenzsynthesen ein relativ hohes Gewicht.*

(Positronenemissionstomographie (PET und PET/CT) bei malignen Lymphomen, Abschlussbericht (D06-01A), Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 31.03.2009, ISSN: 1864-2500, S. 117)

### Weitere Anmerkungen

#### Verwendete ein- und ausgeschlossene Literatur

Wir bitten Sie, das Verzeichnis der verwendeten ein- und ausgeschlossenen Literatur – im Gegensatz zur Vorgehensweise im Vorbericht D06-01A – in alphabetischer Folge anzulegen, fortlaufend nummeriert. So sind die Informationen nachvollziehbar gelistet, was die Arbeit einer späteren Stellungnahme erleichtert.

Wir bedanken uns für die Berücksichtigung unserer Anmerkungen für die weitere Bearbeitung des vorläufigen Berichtsplans.

Mit freundlichen Grüßen



Prof. Dr. med. Dr. rer. nat. Andreas Bockisch  
Präsident der DGN



Prof. Dr. med. Frank Grünwald  
stellv. Präsident der DGN



Prof. Dr. med. Jörg Kotzerke  
Schriftführer und  
Vorsitzender des Ausschusses Positronen Emissions Tomographie (PET) der DGN

## Anlagen:

- Formblatt zur schriftlichen Stellungnahme zum Berichtsplan
- Formblatt zur Offenlegung potenzieller Interessenkonflikte: Prof. Dr. Dr. A. Bockisch
- Formblatt zur Offenlegung potenzieller Interessenkonflikte: Prof. Dr. F. Grünwald
- Formblatt zur Offenlegung potenzieller Interessenkonflikte: Prof. Dr. J. Kotzerke
- Zitierte Literatur: Untersuchung klinischer Leitlinien der AHA (American Heart Association) „Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines“ [Tricoci et al. (JAMA 301:831-41, 2009)]

### **A 1.3 Deutsche Krankenhausgesellschaft**

**Autoren:**

Brenske, Michael, Dr.

Krummrey, Gert, Dr.

Neumeyer-Gromen, Angela, Dr.

Schlottmann, Nicole, Dr.

**Adresse:**

Dr. Angela Neumeyer-Gromen

Deutsche Krankenhausgesellschaft, Dezernat V

Wegelystraße 3

10623 Berlin



DEUTSCHE  
KRANKENHAUS  
GESELLSCHAFT

Institut für Qualität und  
Wirtschaftlichkeit im Gesundheitswesen  
- Stellungnahme zum Berichtsplan <D06-01B>  
Herrn Prof. Dr. med. Peter T. Sawicki  
Dillenburger Str. 27  
51105 Köln

Bundesverband  
der Krankenhausträger  
in der Bundesrepublik  
Deutschland

- vorab per E-mail und Fax -

Unser Zeichen	Telefon Durchwahl +49(0)30 398 01 -	Telefax +49(0)30 398 01 -	Datum
Dez. V/ Dr. Neu	1510	3510	28.09.2009

**Stellungnahme der Deutschen Krankenhausgesellschaft zum vorläufigen Berichtsplan D06-01C „Positronenemissionstomographie (PET und PET/CT) bei rezidivierendem kolorektalen Karzinom“**

Sehr geehrter Herr Professor Sawicki,  
sehr geehrte Damen und Herren,

für die Möglichkeit zu oben genanntem Berichtsplan Stellung nehmen zu können, bedanken wir uns recht herzlich.

Der Gemeinsamen Bundesausschuss (G-BA) hat das Institut gem. § 139a Abs. 3, 1. SGB V am 21.12.2006 mit der Recherche, Darstellung und Bewertung des aktuellen medizinischen Wissenstandes der PET-Diagnostik beim rezidivierenden kolorektalen Karzinom beauftragt. Der spätere Abschlussbericht soll eine Grundlage für die Beratungen des G-BA zur Bewertung der PET bei o. g. Indikation nach § 137c SGB V (Methodenbewertung stationär) und § 135 SGB V (Methodenbewertung ambulant) darstellen.

Gemäß § 137c SGB V hat der G-BA zu prüfen, ob Untersuchungs- oder Behandlungsmethoden, die zu Lasten der gesetzlichen Krankenkassen im Rahmen einer Krankenhausbehandlung angewendet werden oder angewendet werden sollen, für eine ausreichende, zweckmäßige und wirtschaftliche Versorgung der Versicherten unter Berücksichtigung des allgemein anerkannten Standes der medizinischen Erkenntnisse erforderlich sind.

Die Deutsche Krankenhausgesellschaft (DKG) als eine Trägerorganisation des G-BA nimmt zum vorliegenden Berichtsplan des Instituts wie folgt Stellung:

1. Die von Ihnen vorab anhand der Ein- und Ausschlusskriterien erfolgte Einschränkung der zu berücksichtigenden Primärliteratur auf Studien der Evidenzstufen I und II halten wir für nicht sachgerecht und nicht konform mit den Vorgaben der Verfahrensordnung des G-BA. Zu dieser Problematik haben wir bereits mehrfach im Rahmen von Stellungnahmen u. a. zur allgemeinen Methodik des Instituts hingewiesen und diesen Aspekt in der Auftragskonkretisierung (Sitzung der Themengruppe PET vom 21.03.2007) bereits thematisiert.

Postfach 120555 · D-10595 Berlin  
Wegelystraße 3 · D-10623 Berlin  
Telefon +49 (0) 30 398 01 - 0  
Telefax +49 (0) 30 398 01 - 3000  
dkg@mail@dkggev.de · www.dkggev.de

Rue d'Arlon 50  
B-1000 Brüssel  
Telefon +32 (0) 2 282-0594  
Telefax +32 (0) 2 282-0598  
europa@dkggev.de

Deutsche Bank AG Berlin, Kto.-Nr. 666 11 77 (BLZ 100 700 00)  
BIC (SWIFT-Code): DEUTDE33 · IBAN: DE36  
Deutsche Bank AG Düsseldorf, Kto.-Nr. 740 730 7 (BLZ 300 700 10)  
BIC (SWIFT-Code): DEUTDE33 · IBAN: DE79  
UST-IdNr. DE119355528

Die Einschränkungen führen gerade auch in diesem Fall dazu, dass ein Großteil wissenschaftlicher Erkenntnisse von vornherein ausgeschlossen bleibt. Damit kann insbesondere die „Darstellung und Bewertung des aktuellen medizinischen Wissensstandes“, die ausdrücklich Gegenstand des Auftrages vom 21.12.2006 ist, nicht sachgerecht erfolgen.

Wie Ihnen bekannt ist, erfolgten in jüngster Vergangenheit durch den G-BA bereits Bewertungen der PET bei zwei anderen onkologischen Indikationen, dem kleinzelligen (SCLC) und dem nicht-kleinzelligen (NSCLC) Lungenkarzinom. Im Rahmen dieser Bewertungen wurden vom G-BA auch Erkenntnisse als relevant für die Entscheidung eingestuft, welche nicht aus randomisierten und kontrollierten Studien stammen. Wir begrüßen es sehr, dass Sie in Ihrem Bericht nun auch prospektive, nicht randomisierte Vergleichsstudien zur Nutzenbewertung heranziehen möchten. Wir bitten Sie darüber hinaus, alle vergleichenden Studientypen zur Bewertung einzubeziehen, d.h. auch Studien auf Stufe III gemäß Verfahrensordnung des G-BA in Betracht zu ziehen.

Die Darstellungen des vorgelegten Berichtsplanes als Ausgangspunkt nehmend, wird der G-BA hier voraussichtlich wieder selbständig weitere Erkenntnisse ermitteln müssen.

2. Im Weiteren möchten wir bitten, explizite Angaben zum Umgang mit Leitlinien zu machen. Dies betrifft insbesondere qualitativ hochwertige, wie z.B. nach AWMF als S-3 gekennzeichnete Leitlinien. Diese beinhalten methodisch hochwertige systematische Reviews, auf deren Grundlage Empfehlungen im Expertenkonsens unter Berücksichtigung der Versorgungsrealität erarbeitet werden. Da sie sich an transparenten Kriterien evidenzbasierter Medizin orientieren, sollten sie Eingang in die Bewertung finden.
3. Aufgrund heterogener Patientenkollektive mit unterschiedlichen Lokalisationen und prognostischen Faktoren sowie der Anwendung unterschiedlicher Therapieverfahren ist die Attribuierung von Effekten eines diagnostischen Verfahrens wie der PET allein auf den von Ihnen definierten patientenrelevanten Nutzen (Mortalität, krankheitsfreies Überleben, Morbidität und Lebensqualität) nicht ausreichend. Die Forderung nach einem Nutzenbeleg auf Stufe 5 nach Fryback und Thornbury (Auswirkung im Hinblick auf patientenrelevante Endpunkte) erscheint daher problematisch. Wie von uns im Rahmen der Gespräche zur Auftragskonkretisierung bereits dargestellt, ist eine Untersuchung der Auswirkungen auf das therapeutische Denken und Handeln und den damit verbundenen Krankheitsverlauf (Stufe 4) angemessener. So beziehen sich die in der Auftragskonkretisierung benannten zentralen Fragestellungen auf klinische Management-Änderungen gemäß Stufe 4 wie z.B. die Vermeidung unnötiger Metastasenoperationen oder die Erhöhung kurativ behandelbarer Metastasen. Entgegen Ihrem Berichtsplan sollten sie deshalb nicht nur „ergänzend“ sondern den Hauptzielgrößen zugeordnet werden.  
Wir begrüßen es sehr, dass Sie gemäß Auftragskonkretisierung die Genauigkeit des Diagnostikverfahrens, d.h. die diagnostische und prognostische Güte bewerten wollen, soweit Primärstudien und Evidenzsynthesen zum patientenrelevanten Nutzen nicht vorhanden sind. Sie beziehen sich auf das Klassifikationsschema von Fryback und Thornbury, welches die Evidenzstufen von Testgüte, Nutzen und Schaden hierarchisch integriert. Dabei spiegelt die Sicherheit eines Testverfahrens Stufe 1, die Auswirkungen in Bezug auf die Diskriminierungsfähigkeit hinsichtlich diagnostischer Genauigkeit Stufe 2, Auswirkungen auf das diagnostische Denken Stufe 3, Auswirkungen auf das therapeutische Denken Stufe 4, Auswirkungen auf patientenrelevante Endpunkte Stufe 5 und schließlich Auswirkungen auf der Systemebene hinsichtlich Kosteneffektivität Stufe 6 wider. In Ihrem Berichtsplan rekurrieren Sie hinsichtlich diagnostischer und prognostischer Güte auf Stufe 2 und zur Nutzen- und Schadensbewertung auf Stufe 5 dieses Klassifikationsschemas. Angesichts der differenzierten, inhaltlich unmittelbar aufeinander aufbauenden Evidenzsteigerung wäre es inhaltlich sinnvoll und stringent, die Evidenz von Stufe 2 bis 5 ein-

schließlich Stufe 3 und 4 nach Fryback und Thornbury in die Betrachtungen einzubeziehen und darzustellen.

4. Darüber hinaus möchten wir darauf hinweisen, dass auch die prognostische Güte einen patientenrelevanten Faktor hinsichtlich Lebensqualität darstellt, und möchten deshalb anregen, dies in die Betrachtungen aufzunehmen.
5. Wir begrüßen sehr, dass Sie Patientenvertreter bei der Erstellung des Berichtsplans eingebunden haben. Sie stellen in Aussicht, dass zudem externe Sachverständige in die Bearbeitung des Projekts eingebunden werden. Wir regen in Bezug auf das vorliegende Bewertungsverfahren an, bereits ab Erstellung des Berichtsplanes onkologische, nuklearmedizinische und chirurgische Expertise einzubeziehen.
6. Zur Verifikation positiver PET-Befunde gilt als höchste Referenz die Biopsie. Aus ethischen und technischen Gründen sind meist jedoch nicht sämtliche PET-Befunde mittels Biopsie zu sichern. Ersatzweise muss daher ggf. eine definitive Diagnosesicherung mittels ergänzender Bildgebung, bildgebenden Verlaufskontrollen (Regression/ Progress von Herden) bzw. des klinischen Verlaufs berücksichtigt werden. Wir begrüßen ihren Hinweis im Bericht und möchten unterstreichen, dass in jedem Fall auch Studien und Evidenzsynthesen ohne Biopsie als Referenzstandard berücksichtigt werden sollten.
7. Darüber hinaus ist zu bedenken, dass die FDG-PET in der Regel als sogenannter „Add-On“-Test zum bestehenden diagnostischen Work-Up durchgeführt wird, so dass es z. B. beim Re-Staging ausreichend sein kann, nur die Patienten mit negativer Standarddiagnostik und positivem PET zu verifizieren.
8. Im Fazit des Abschlussberichtes des Instituts zur Anwendung der PET bzw. PET/CT bei malignen Lymphomen (D06-01A) heißt es: *„Aufgrund des fehlenden validen Referenzstandards kann der Stellenwert der PET im Vergleich zu den konventionellen bildgebenden Verfahren für alle Fragestellungen dieses Berichts nur mithilfe von kontrollierten (idealerweise randomisierten) Studien definiert werden, die die diagnostisch-therapeutische Einheit als eine Fragestellung untersuchen.“*

Es ist absehbar, dass sich die Problematik des fehlenden Referenzstandards auch bei der vorliegenden Indikation ergeben wird, da auch beim rezidivierenden kolorektalen Karzinom und zu untersuchender Metastasen nicht immer eine Biopsie möglich sein wird. In Anbetracht der diversen, noch folgenden Bewertungsverfahren regen wir an, hier einen pragmatischen Ansatz wie unter Punkt 6 skizziert im Umgang mit den sich daraus ergebenden Unsicherheiten der Bewertung zu finden.

Darüber hinaus möchten wir in Bezug auf den Begriff der therapeutisch-diagnostischen Einheit gemäß Auftragskonkretisierung nochmals darauf hinweisen, dass eine *direkte* Beeinflussung der von ihnen angeführten patientenrelevanten Zielgrößen durch ein diagnostisches Verfahren nicht möglich ist. Eine diagnostische Methode kann daher anhand ihrer Befunde nur *indirekt* in Form einer therapeutischen Konsequenz patientenrelevante Endpunkte beeinflussen: Breiter klinisch-wissenschaftlicher Konsens ist das Ziel eines möglichst exakten Tumorstagings, z.B. nach der TNM-Klassifikation, welches zu einer Risikostratifikation der Patienten führt, wonach wiederum krankheitsstadienadaptierte Therapieansätze oder weiterführende Diagnostik entsprechend dem medizinischen Wissensstand gewählt werden. Demzufolge übt das korrekte Staging beim rezidivierenden kolorektalen Karzinom grundsätzlich einen prognoserelevanten Einfluss auf den jeweiligen Patienten aus, selbst wenn dieser noch nicht direkt nachgewiesen sein sollte.

Wir hoffen, dass wir Ihnen mit unseren Anmerkungen einige Anregungen für die Überarbeitung des Berichtsplans und die weitere Auftragsbearbeitung geben konnten. Im Übrigen behalten wir uns vor, ggf. ergänzende Aspekte in das weitere Verfahren einzubringen.

Mit freundlichen Grüßen

Der Hauptgeschäftsführer  
In Vertretung



Dr. N. Schlottmann  
Geschäftsführerin  
Dezernat Medizin

Im Auftrag



Dr. M. Brenske  
Stellv. Geschäftsführer  
Dezernat Medizin

gez. Dr. G. Krummrey, MSc  
Referent  
Dezernat Medizin



Dr. A. Neumeyer-Gromen, MPH  
Referentin  
Dezernat Medizin

## **A 2 Stellungnahmen von Privatpersonen**

Es wurden keine Stellungnahmen von Privatpersonen abgegeben.