

Positronenemissionstomographie (PET) und PET/CT bei Ösophaguskarzinom

Dokumentation und Würdigung der Anhörung zum Vorbericht

Auftrag: D06-01H
Version: 1.0
Stand: 24.06.2013

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Positronenemissionstomographie (PET) und PET/CT bei Ösophaguskarzinom

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

21.12.2006

Interne Auftragsnummer:

D06-01H

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8
50670 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

Berichte@iqwig.de

www.iqwig.de

Inhaltsverzeichnis

	Seite
Abkürzungsverzeichnis.....	iii
1 Dokumentation der Anhörung	1
2 Würdigung der Anhörung	2
3 Offenlegung potenzieller Interessenkonflikte	3
3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen	3
3.2 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige, externe Reviewer).....	4
4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....	7
4.1 Teilnehmerliste der wissenschaftlichen Erörterung	7
4.2 Tagesordnung der wissenschaftlichen Erörterung	7
4.3 Protokoll der wissenschaftlichen Erörterung.....	7
4.3.1 Begrüßung und Einleitung.....	8
4.3.2 Tagesordnungspunkt 1: Berücksichtigung der Studien Lordick et al. 2007 (MUNICON I), Zum Büschenfelde et al. 2011 (MUNICON II) und Lorenzen et al. 2011 (HICON).....	9
4.3.3 Tagesordnungspunkt 2: Indikation Primärstaging: Umgang mit N- und M- Kategorien vor dem Hintergrund der Änderung der TNM-Klassifikation (6. vs. 7. Auflage)	15
4.3.4 Tagesordnungspunkt 3: Beurteilung des Verzerrungspotenzials nach QUADAS-Kriterien für Studien zum M-Staging.....	21
4.3.5 Tagesordnungspunkt 4: Verschiedenes	33
Anhang: Dokumentation der Stellungnahmen.....	40

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AEG	Adenokarzinome des gastroösophagealen Übergangs
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
CRD	Center of Reviews and Dissemination
CT	Computertomographie
DGN	Deutsche Gesellschaft für Nuklearmedizin
FDG	2-[¹⁸ F]Fluor-Deoxy-D-Glukose
G-BA	Gemeinsamer Bundesausschuss
HTA	Health Technology Assessment
MSAC	Medical Services Advisory Committee
PET	Positronenemissionstomographie
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
SIGN	Scottish Intercollegiate Guidelines Network
TNM-Klassifikation	Tumorstadieneinteilung nach Tumorausdehnung (T = Tumor); Lymphknotenbefall (N = Nodi lymphatici) und Metastasierung (M = Metastase)
UICC	Union internationale contre le cancer
VOPT	Verification of Only Positive Testers

1 Dokumentation der Anhörung

Am 23.05.2012 wurde der Vorbericht in der Version Nr. 1 vom 15.05.2012 veröffentlicht und zur Anhörung gestellt. Bis zum 26.06.2012 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 2 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 29.08.2012 im IQWiG diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

Eine Würdigung der in der Anhörung vorgebrachten Aspekte befindet sich im Kapitel „Diskussion“ des Abschlussberichts. Im Abschlussbericht sind darüber hinaus Änderungen, die sich durch die Anhörung ergeben haben, zusammenfassend dargestellt. Der Abschlussbericht ist auf der Website des IQWiG unter www.iqwig.de veröffentlicht.

2 Würdigung der Anhörung

Die im Rahmen der Anhörung vorgebrachten Aspekte wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des Vorberichts überprüft. Die wesentlichen Argumente wurden im Kapitel „Diskussion“ des Abschlussberichts gewürdigt. 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.

Der Abschlussbericht ist auf der Website des IQWiG unter www.iqwig.de veröffentlicht.

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 abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen

Organisation/Institution	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Deutsche Gesellschaft für Nuklearmedizin e.V.	Hellwig, Dirk ¹	nein	ja	nein	nein	nein	nein
	Herrmann, Ken ¹	ja	nein	nein	nein	nein	ja
	Kotzerke, Jörg ¹	nein	nein	nein	nein	nein	nein
	Schmidt, Matthias ¹	nein	nein	nein	nein	ja	nein
	Weber, Wolfgang ¹	nein	nein	nein	ja	nein	nein
Deutsche Krankenhausgesellschaft e. V.	Acikgöz, Öslem ¹	ja	nein	nein	nein	nein	nein
	Dippmann, Anja ¹	ja	nein	nein	nein	nein	nein
	Neumeyer-Gromen, Angela ¹	ja	nein	nein	nein	nein	nein
	Rüther, Klaus ¹	ja	nein	nein	nein	nein	nein
	Schlottmann, Nicole ¹	ja	nein	nein	nein	nein	nein

¹ Version 12/2011

3.2 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige, externe Reviewer)

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Ott, Katja ²	nein	nein	nein	nein	nein	nein
Wolff, Robert ³	nein	nein	nein	ja	nein	nein

² Version 12/2011

³ Version 03/2009

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt (Version: 03/2009):

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter¹ 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 Institution², 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?

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

² 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 Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt (Version 12/2011):

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor angestellt bei einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere bei einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor ein Unternehmen, eine Institution oder einen Interessenverband im Gesundheitswesen, insbesondere ein pharmazeutisches Unternehmen, einen Hersteller von Medizinprodukten oder einen industriellen Interessenverband, direkt oder indirekt beraten?

Frage 3: Haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor direkt oder indirekt von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband, Honorare erhalten für Vorträge, Stellungnahmen oder Artikel?

Frage 4: Haben Sie und / oder hat die Einrichtung¹, die Sie vertreten, abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband, finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten?

Frage 5: Haben Sie und / oder hat die Einrichtung¹, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Unterstützung bei der Ausrichtung einer Veranstaltung, Übernahme von Reisekosten oder Teilnahmegebühren ohne wissenschaftliche Gegenleistung) erhalten von einem Unternehmen, einer Institution oder einem Interessenverband im Gesundheitswesen, insbesondere von einem pharmazeutischen Unternehmen, einem Hersteller von Medizinprodukten oder einem industriellen Interessenverband?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile eines Unternehmens oder einer anderweitigen Institution, insbesondere von einem pharmazeutischen Unternehmen oder einem Hersteller von Medizinprodukten? Besitzen Sie Anteile eines „Branchenfonds“, der auf pharmazeutische Unternehmen oder Hersteller von Medizinprodukten ausgerichtet ist?

¹ Sofern Sie in einer ausgedehnten Institution tätig sind, genügen Angaben zu Ihrer Arbeitseinheit, zum Beispiel Klinikabteilung, Forschungsgruppe etc.

4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

4.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma / privat
Acikgöz, Özlem	Deutsche Krankenhausgesellschaft e. V.
Beckmann, Lars	IQWiG
Ernst, Stefan	Protokollant Sitzungsprotokollarischer Dienst des Landtags NRW
Hellwig, Dirk	Deutsche Gesellschaft für Nuklearmedizin e. V.
Kotzerke, Jörg	Deutsche Gesellschaft für Nuklearmedizin e. V.
Menig, Matthias	IQWiG
Neumeyer-Gromen, Angela	Deutsche Krankenhausgesellschaft e. V.
Ott, Katja	Universitätsklinikum Heidelberg
Sauerland, Stefan	IQWiG
Scheibler, Fülöp	IQWiG
Stöcker, Michael	IQWiG
Thomas, Stefanie	IQWiG
Waffenschmidt, Siw	IQWiG
Weber, Wolfgang	Deutsche Gesellschaft für Nuklearmedizin e. V.
Windeler, Jürgen	IQWiG (Moderation)
Wolff, Robert	Kleijnen Systematic Reviews

4.2 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Berücksichtigung der Studien Lordick et al. 2007 (MUNICON I), Zum Büschenfelde et al. 2011 (MUNICON II) und Lorenzen et al. 2011 (HICON)
TOP 2	Indikation Primärstaging: Umgang mit N- und M-Kategorien vor dem Hintergrund der Änderung der TNM-Klassifikation (6. vs. 7. Auflage)
TOP 3	Beurteilung des Verzerrungspotenzials nach QUADAS-Kriterien für Studien zum M-Staging
TOP 4	Verschiedenes

4.3 Protokoll der wissenschaftlichen Erörterung

Datum: 29.08.2012, 14:00 bis 15:00 Uhr

Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG),
Im Mediapark 8, 50670 Köln

Moderation: Jürgen Windeler

4.3.1 Begrüßung und Einleitung

Moderator Jürgen Windeler: Ich begrüße Sie zur zweiten Runde. Die beiden Anhörungen am heutigen Tag führen dazu, dass einige von Ihnen, die heute Morgen schon einmal da waren, sich das, was ich hier zu sagen habe, noch einmal anhören müssen.

Ich fange mit meinen üblichen Erläuterungen an. Ich begrüße Sie zur Erörterung von PET und PET/CT bei Ösophaguskarzinom und weise Sie zu Beginn darauf hin, dass die Erörterung aufgezeichnet wird und dass das Wortprotokoll der Erörterung ins Netz gestellt, also veröffentlicht wird. Ich weise noch einmal darauf hin, dass Sie, indem Sie der Einladung, in der das schon stand, gefolgt und hierhergekommen sind, Ihr Einverständnis geben, dass das alles so passiert. Ich weise Sie zum Schluss darauf hin, dass Sie, wenn Sie sich jetzt überlegen sollten, das doch nicht zu wollen, an dieser Erörterung nicht weiter teilnehmen können.

In diesem Zusammenhang weise ich darauf hin, dass Sie bitte wegen der Aufzeichnung – und um dem Stenografen die Arbeit zu erleichtern – vor jeder Wortäußerung Ihren Namen sagen, damit diese richtig zugeordnet werden können.

Noch einmal der allgemeine Hinweis zum Beginn: Diese Erörterung dient der Klärung von Unklarheiten, die beim IQWiG nach Studium der Stellungnahmen übrig geblieben sind. Wir haben die Stellungnahmen bekommen, sie gelesen und sie – davon dürfen Sie ausgehen – weitestgehend verstanden. Wir thematisieren jetzt hier die Punkte, bei denen wir noch Unklarheiten haben, die wir gern mit Ihnen klären und erörtern möchten.

Wie die Tagesordnung erkennen lässt – das zum Selbstverständnis für diejenigen, die zum ersten Mal dabei sind –, dient diese Erörterung nicht bzw. sicherlich nicht primär dazu, eine Generaldiskussion über die Gesetzeslage, die Beauftragung des G-BA und den Stellenwert von EbM im deutschen Gesundheitssystem oder Ähnliches zu führen. Dass man unter „Verschiedenes“ seine Meinung zu Punkten, die einem am Herzen liegen, sagen kann, ist klar. Dass man manchmal – wir hatten das auch in der vorherigen Erörterung – vom sehr Speziellen zum etwas Allgemeineren kommt, ist auch, glaube ich, leicht verständlich.

Wir haben Ihnen eine Tagesordnung vorbereitet und vorgelegt. Ich frage Sie zunächst: Gibt es dazu Wortmeldungen? Streichungswünsche? Ergänzungswünsche? – Das ist nicht der Fall.

Gibt es darüber hinaus – über die Tagesordnung und alles, was ich jetzt gerade gesagt habe – Klärungsbedarf, Unklarheiten zum Ablauf oder andere Dinge? – Das scheint nicht so zu sein. – Dann, Herr Scheibler, TOP 1.

4.3.2 Tagesordnungspunkt 1: Berücksichtigung der Studien Lordick et al. 2007 (MUNICON I), Zum Büschenfelde et al. 2011 (MUNICON II) und Lorenzen et al. 2011 (HICON)

Fülöp Scheibler: In der Stellungnahme der DGN wird darauf hingewiesen, dass es unklar geblieben sei, weshalb die MUNICON-I-, die MUNICON-II- und die HICON-Studie nicht angemessen berücksichtigt wurden.

An dieser Stelle sei darauf hingewiesen, dass sowohl die MUNICON-II- als auch die HICON-Studie erst in 2011 publiziert wurden und somit erst zum Studienpool der Nachrecherche zählen. Dennoch kann in dieser Erörterung auf beide Studien schon vorab eingegangen werden.

Aus unserer Sicht handelt es sich bei allen 3 Studien um Interventionsstudien ohne Kontrollarm. Damit können sie weder unter Fragestellung 1, also die Nutzenbewertung, noch unter Fragestellung 2, also Diagnose und prognostische Güte, berücksichtigt werden.

Wir haben die Studien aber im Diskussionsabschnitt gewürdigt und sehen durchaus ihre Bedeutung. Wir würden aber gern von Ihnen wissen, wie Sie diese Studien interpretieren und unter welcher der beiden Fragestellungen Sie diese Studien berücksichtigt sehen wollen bzw. in welcher Form wir sie besser würdigen könnten.

Wolfgang Weber: Ich möchte – leider – mit einer Gegenfrage antworten. Sie haben die vorangegangenen Studien aus München zur Therapiekontrolle beim Ösophaguskarzinom mittels PET berücksichtigt. Jetzt sagen Sie, dass Sie die MUNICON-Studie nicht berücksichtigen könnten, weil es keine randomisierte Studie war. Ich verstehe die Logik dahinter nicht. Denn die Patienten wurden genauso wie in den vorangegangenen Studien behandelt: Das PET wurde gemacht, und das wurde dann korreliert. Zumindest bei der Hälfte der Patienten hat sich im Vergleich zu den Dingen, die bei der vorangegangenen Studie gemacht wurden, gar nichts geändert. Insofern: Könnten Sie kurz erläutern, warum diese Studie jetzt nicht mit hineingekommen ist?

Moderator Jürgen Windeler: Herr Scheibler.

Fülöp Scheibler: Tatsächlich besteht der Unterschied zwischen Ihren Vorgängerstudien, der Studie von Frau Ott und der MUNICON-Studie darin, dass in der MUNICON-Studie PET-gesteuert interveniert wurde. Bei den PET-Negativen wurde die neoadjuvante Chemotherapie abgebrochen und unmittelbar operiert. Das ist das Problem. Das ist im Prinzip eine PET-gesteuerte Intervention – aber nur einarmig. Sie haben zwar einen Interventionsarm, aber keinen Kontrollarm. Das heißt, als vergleichende Studie zum patientenrelevanten Nutzen können Sie sie nicht einschließen, und als Beobachtungsstudie können Sie sie auch nicht einschließen. Denn es handelt sich nicht um eine Beobachtung, sondern um eine Intervention. Aufgrund des PET-Ergebnisses, aufgrund des diagnostischen Tests, hat sich die Intervention bei einigen der Patienten im Vergleich zu vorher verändert. Das können wir nicht separieren.

Selbst wenn wir es separieren könnten, könnten wir vielleicht noch eine einigermaßen valide Aussage für die PET-Positiven machen, aber auf keinen Fall mehr für die PET-Negativen, die nach 2 Zyklen ihre neoadjuvante Chemotherapie abgebrochen haben.

Wolfgang Weber: Etwa 50 % sind bei der PET Responder, also PET-positiv, wie Sie gesagt haben.

(Zustimmung von Fülöp Scheibler)

Diese 50 % sind genauso weiterbehandelt worden.

(Fülöp Scheibler: Ja!)

Es ist schon eine wichtige Aussage der MUNICON-Studie, dass die Rate der Patienten, die bezogen auf die Gesamtpopulation histologische Responder waren, sehr genau gleich wie in den vorangegangenen Studien war.

(Fülöp Scheibler: Ja!)

Das bestätigt sehr schön und in einem prospektiven Setting, dass man diese frühe Therapiebeurteilung durchführen kann.

(Fülöp Scheibler: Genau!)

Katja Ott: Dazu muss man sagen: Das Hauptinteresse aus klinischem Aspekt ist die Beurteilung der Non-Responder. Ganz klar hat man im Arm der Responder nachweisen können, dass die Korrelation mit histopathologischer Response ähnlich ist wie in unseren beiden Vorgängerstudien. Aber wir haben etwa 50 % histopathologische Responder. Das heißt, die Sensitivität ist relativ gering, wenn wir ehrlich sind.

Im Arm der Non-Responder haben wir eine Intervention. Deshalb können wir das nicht korrekt vergleichen, weil wir onkologisch davon ausgehen müssen, dass unter Umständen sekundäre Responder drin sind, denen wir die Chance nehmen. Aber in den beiden Studien MUNICON I und MUNICON II war im Endeffekt die Rate der R0-Resektion deutlich oder signifikant geringer als im Responderarm.

Das wird den Studien häufig vorgeworfen, dass nämlich kein Vergleichsarm da ist und dass wir interveniert haben, ohne bewiesen zu haben, dass es besser ist. Wir argumentieren immer mit historischen Kontrollen, die zwar in Ordnung, aber nicht statistisch valide sind.

Moderator Jürgen Windeler: Herr Scheibler.

Fülöp Scheibler: Die historische Kontrolle haben wir im Diskussionsteil aufgegriffen. Aber das ist tatsächlich eine Methode, die wieder mit sehr starkem Verzerrungspotenzial verbunden

ist. Aus historischen Kontrollen leitet das IQWiG eigentlich keine Aussagen zum Nutzen ab. Das war unser Dilemma bei diesen Studien.

Wir sind übrigens davon überzeugt, dass das eine der besten Serien überhaupt ist, die wir bisher zu PET zu Gesicht bekommen haben.

(Zustimmung von Katja Ott)

Denn es kommt sehr selten vor, dass es Vorgängerstudien gibt, die beispielsweise einen Cutoff festlegen, sowie prospektive Studien, die diesen Cutoff validieren, und dann auch noch Studien, die, aufbauend auf diesen Ergebnissen, tatsächlich die Behandlung variieren.

Die logische Konsequenz daraus wäre jetzt, darauf eine randomisiert-kontrollierte Studie aufzusatteln. Wir stehen bzw. standen mit Herrn Prof. Lordick in Kontakt, der das zwar geplant hatte, es aber mangels Finanzierung nicht durchführen konnte, was sehr bedauerlich ist.

Aber aus diesen einarmigen Interventionsstudien kann man tatsächlich weder etwas zum Nutzen noch zur diagnostischen Güte ableiten.

Wolfgang Weber: Der Punkt zur diagnostischen Güte ist meiner Meinung nach nur teilweise korrekt. Denn Sie können sagen: Es gibt eine Reihe von Studien, bei denen die histologische Ansprechrate insgesamt bei etwa 25 % lag.

Jetzt gab es eine PET-Studie, bei der wir die Therapie abgebrochen haben, wenn kein Ansprechen im PET war. Das waren 50 %. Bei denen, bei denen die Therapie weitergeführt wurde, war die Ansprechrate 50 %. Insofern ist das natürlich schon eine sehr schöne Bestätigung, dass das konsistent bleibt, wenn man nicht davon ausgeht, dass es sehr starke Verzerrungen in der Patientenpopulation gegeben hat, was aber in 2 vorangegangenen Studien auch nicht der Fall war. Ich bin mir nicht sicher, ob man den diagnostischen Teil ganz unter den Tisch fallen lassen kann.

Fülöp Scheibler: Ich sehe das genauso wie Sie. Aber Sie argumentieren jetzt wieder über die historische Kontrolle. Sie argumentieren darüber, dass Sie die Ergebnisse der MUNICON-I- bzw. der MUNICON-II-Studie beispielsweise mit der Studie von Ott vergleichen.

Wir sehen das genauso wie Sie, wir wissen aber schlicht und einfach nicht, ob diese Ergebnisse durch irgendwelche Confounder durch die historische Kontrolle oder durch den unterschiedlichen Einschluss entstanden sein könnten. Die Patienten, die in der Ott-Studie und in der MUNICON-Studie eingeschlossen wurden, sind nicht 2 Kollektive, die durch Randomisierung bezüglich wichtiger Confounder vergleichbar sind. Insofern kann es sein, dass Ergebnisse, die Sie daraus ableiten, verzerrt sind.

Wir wissen, dass es relativ homogene Kollektive sind, aber es bewegt sich auf dem Niveau eines nichtrandomisierten historischen Vergleichs.

Moderator Jürgen Windeler: Herr Hellwig.

Dirk Hellwig: Ich möchte formal werden und auf eine Sache hinweisen, die in der Verfahrensordnung des Gemeinsamen Bundesausschusses steht und zwar bezüglich der zulässigen Studien, die für eine Nutzenbewertung herangezogen werden können. In der Verfahrensordnung unterscheidet sich nur die Vorgabe, was bei Therapie- und was bei Diagnosestudien verwendet werden kann, und zwar gibt es dort auf der höchsten Evidenzstufe die „sonstigen Interventionsstudien“. Nach meinem Empfinden ist die MUNICON-II-Studie eine sonstige Interventionsstudie. Es ist nicht gefordert, dass man eine mehrarmige Studie hat. Es ist auch nicht gefordert, dass man randomisiert, sondern der Gemeinsame Bundesausschuss erlaubt die Beurteilung des Zusatznutzens anhand sonstiger Interventionsstudien.

Wir sind uns einig darüber, dass die hier geschilderten Verzerrungspotenziale gegeben sind. Aber das sollte nicht Grund sein, von vornherein die Betrachtung dieser Studien auszuschließen. Sie können dann am besten im Bericht entsprechend beschreiben, dass dieses Verzerrungspotenzial durch den lediglich anhand des historischen Vergleichs möglichen Vergleich vorliegt. Aber es ist schade, wenn diese wichtige Studie schon primär ausgeschlossen wird. Das steht der Verfahrensordnung des Gemeinsamen Bundesausschusses entgegen.

Stefan Sauerland: Sie haben Recht, dass in der Verfahrensordnung die „anderen Interventionsstudien“ für die diagnostische Verfahrensbewertung genannt werden. Allerdings muss man sich auch anschauen, dass dieser Terminus „andere“ nicht die Tür gegenüber jeglichen Studiendesigns öffnet.

(Dirk Hellwig: Das wünschen wir auch nicht!)

Da würden wir daran denken, weil hier gerade die Kombination aus Diagnose und Therapie das Entscheidende ist, dass man sich auch ein wenig daran orientieren darf, was im Bereich Therapie vorgeschrieben wird. Da werden die Studiendesigns aus gutem Grunde relativ fein aufgedröselt. Dort kann man feststellen, dass historische Vergleichsgruppen – das wäre eine retrospektive Vergleichsgruppe aus der Betrachtung der zweiten Studie – in der Evidenzhierarchie deutlich weiter unten sind.

Wenn Sie das so interpretieren wollten, würde sich in dem Fall mit Blick auf die Verfahrensordnung des Gemeinsamen Bundesausschusses nach meiner wirklich nicht maßgeblichen Rechtsauffassung – ich habe keine Ahnung davon – einiges methodisch beißen, wenn man das so interpretieren wollte.

Moderator Jürgen Windeler: Ich glaube, die Antwort auf die Frage ... Habe ich sowohl die Frage als auch die Antwort verstanden, dass Sie sagen: Diese 3 Studien eignen sich, auch wenn sie nur einarmig sind, aufgrund zwar nur historischer, aber denkbarer Kontroll-/Referenzmöglichkeiten sowohl für eine Nutzaussage als auch – ich will das nicht beurteilen, ich referiere das nur, um zu prüfen, ob ich das richtig verstanden habe bzw. ob wir das richtig verstanden haben – für Aussagen zur diagnostischen Güte? – So habe ich Ihre Rückmeldungen wahrgenommen.

Wolfgang Weber: Ich würde gern präzisieren. Mit „Nutzen“ wäre gemeint, dass man bei etwa der Hälfte der Patienten eine Therapie, die nicht zu einem histologischen Ansprechen führt, abgebrochen hat. Das wäre der Nutzen. Wir hatten noch nicht definiert, was damit gemeint ist.

Stefan Sauerland: Aber in der Publikation, in der Sie und Frau Ott Autoren sind, wird in der Diskussion ganz klar gesagt, dass man eine randomisierte Studie braucht, um zu bestätigen, dass es tatsächlich sinnvoll ist, die Chemotherapie bei den Non-Respondern abubrechen.

Insofern würde ich denken, dass Sie – dort auf Ihre wissenschaftliche Basis zurückgeführt – diese Schlussfolgerung nicht zweimal anders ziehen können.

Wolfgang Weber: Ich würde das nicht zurücknehmen wollen. Aber wenn man in diesen Hierarchien denkt, ist es nicht adäquat, die Studie ganz herauszunehmen, sondern sie ist ein notwendiger Schritt.

(Stefan Sauerland: Ja!)

Man kann nicht immer gleich mit randomisierten Studien alles untersuchen, sondern muss eine gewisse schrittweise Reihenfolge haben. Aus den genannten Gründen hat es sich als sehr schwierig erwiesen, für eine derartige nichtmedikamentöse Studie, deren Hauptziel es ist, weniger Medikamente einzusetzen, einen entsprechenden Sponsor zu finden.

Moderator Jürgen Windeler: Frau Ott.

Katja Ott: Man muss Folgendes beim Nutzen definieren: Mit MUNICON kann man ganz korrekt nur den Nutzen für die Responder und nicht für die Non-Responder beurteilen, da die Non-Responder interveniert sind.

Aber wenn man nach dem wirklichen Nutzen fragt, hätten nur die Non-Responder einen Benefit, weil das diejenigen Patienten sind, bei denen ich abbreche und denen ich eine ineffektive Therapie erspare. Das heißt, der Nutzen, den Sie evaluieren wollen, ist in dem Rahmen, in dem ich evaluieren kann, nicht zu evaluieren.

Die diagnostische Güte ist im Endeffekt ebenfalls nur an den Respondern zu evaluieren. Da habe ich das Gesamtüberleben und das rezidivfreie Überleben sowie die R0-Resektionsrate.

Zu den Non-Respondern, die abgebrochen worden sind, kann ich abgesehen vom Gesamtüberleben und der R0-Resektionsrate nichts sagen, weil ich da keine Vergleichsgruppe habe.

Die Studie ist konsequent auf den Vorläuferstudien aufgebaut, kann aber nicht definitiv sagen, dass ich einen Nutzen habe. Denn ich kann die Non-Responder, die uns klinisch am meisten am Herzen liegen, nicht beurteilen. Ich kann mit der historischen Kontrolle vergleichen. Da sehe ich, dass sie ein mediales Überleben von 26 Monaten haben, während wir früher eines von 18 Monaten hatten. Also scheinen sie zumindest keinen Nachteil davon zu haben.

Aber es wäre interessant, das in einer wirklich randomisierten Studie zu validieren, dass sie einen Benefit haben. Das glauben wir alle, aber wir können das nicht beweisen. Das wird auch immer auf allen Kongressen kritisiert.

Moderator Jürgen Windeler: Ich versuche, wieder zusammenzufassen und die Frage zurückzugeben, ob das bei mir bzw. bei uns richtig angekommen ist. Wenn ich Sie, Herr Weber, richtig verstehe, sagen Sie auch im Zusammenhang mit der Äußerung von Frau Ott bzw. mit dem, was Stefan Sauerland gerade vorgelesen hat, die Studien sollten eingeschlossen werden, weil sie relevante Informationen über auch denkbare Stratifizieren erlauben, aber – jetzt kommt das entscheidende „aber“ – sie sind nach Ihrer Auffassung – so habe ich Sie gerade verstanden – noch nicht ausreichend für eine Nutzensaussage.

Wolfgang Weber: Die Frage ist, wie man „Nutzen“ definiert. Ich gebe Frau Ott völlig recht, dass das beim Überlebensvorteil nicht gezeigt worden ist. Aber was die Studien zeigen, ist, dass man Chemotherapiezyklen einsparen kann, wenn man den Vergleich mit der historischen Kontrolle als statthaft ansieht.

Moderator Jürgen Windeler: Das würden wir im Prinzip und abstrakt betrachtet natürlich als Nutzenkriterium anerkennen – immer unter der Voraussetzung, dass den Menschen dadurch keine Nachteile entstehen.

Es nützt nichts zu sagen: Ich habe weniger Medikamente und selbst nebenwirkungsreiche Medikamente eingesetzt. – Ich sollte schon sicher sei, dass es dem Patienten darunter genauso gut geht wie mit den Medikamenten.

Wolfgang Weber: Völlig d'accord. Aber hier kommt der historische Vergleich.

Moderator Jürgen Windeler: Okay.

Wolfgang Weber: Frau Ott hat es gerade gesagt: Da ist im historischen Vergleich das Überleben der Patienten etwas günstiger. Insofern ist es schon zumindest ein Hinweis dafür, dass hier gar kein Schaden angerichtet wird. Der Nutzen ist die Einsparung dieser Chemotherapiezyklen. – Die Frage ist, wie man „Nutzen“ definiert.

Moderator Jürgen Windeler: Okay, wir kommen sonst leicht in die Gefahr, sehr übergreifend zu werden. Ich glaube, die konkrete Frage, die wir geklärt haben wollten, ist durch Ihre Rückäußerungen für uns besser bewertbar und klarer geworden. Gibt es weitere Wortmeldungen dazu? – Das ist nicht der Fall. Dann kommen wir zu Tagesordnungspunkt 2.

4.3.3 Tagesordnungspunkt 2: Indikation Primärstaging: Umgang mit N- und M-Kategorien vor dem Hintergrund der Änderung der TNM-Klassifikation (6. vs. 7. Auflage)

Fülöp Scheibler: Die DGN weist in ihrer Stellungnahme darauf hin, dass die überwiegende Zahl der in Deutschland diagnostizierten Ösophaguskarzinome im Stadium N1 diagnostiziert werde. Deshalb sei die Feststellung eines N1-Stadiums keine Fragestellung von hoher klinischer Relevanz. Außerdem weisen Sie in Ihrer Stellungnahme darauf hin, dass bei der Indikation Primärstaging die Differenzierung der Stadien N1 und M1a entscheidend sei.

Beide Feststellungen können wir nicht ganz nachvollziehen. Könnten Sie uns das näher erläutern?

Wolfgang Weber: Zunächst einmal muss man zwischen der TNM-Fassung unterscheiden. Das Gutachten wurde unter der TNM-Klassifikation 6 geschrieben. Zwischenzeitlich kam Nummer 7 heraus. Beim Ösophaguskarzinom haben sich sehr viele Einteilungen geändert – eben auch N1- und Einteilung, was M1a ist. Was wir ausdrücken wollten, war auf der Grundlage der TNM 6, dass es einen großen Unterschied macht, ob es sich um Lymphknotenmetastasen handelt, die regional sind oder die nicht-regional sind, weil diese nicht-regionalen Lymphknoten in der alten TNM-Einteilung nach M1a klassifiziert worden sind, während sie in der neuen TNM bereits als M1 gesehen werden.

Hier gibt es einen gewissen methodischen Zusammenhang, weil diese unmittelbar häufigen tumoranliegenden, regionalen Lymphknoten in der PET wesentlich häufiger falsch-negativ sind als diese mehr distanten, nicht-regionalen Lymphknoten. Hier war uns in der Durchsicht des Berichtes nicht klar, ob diese Unterscheidung wirklich immer so durchgeführt wurde und ob hier nicht Gruppen gemischt worden sind.

Es ist nicht einfach, das zu trennen, weil auch die Arbeiten das oft nicht ganz sauber machen. Aber ich glaube, um hier die diagnostische Genauigkeit zu beschreiben, ist es wichtiger, diesen Nachweis von nicht-regionalen Lymphknotenmetastasen in der alten TNM M1a hier aufzunehmen, als wirklich rein diese N0- und N-Plus-Einteilung vorzunehmen.

Moderator Jürgen Windeler: Danke schön. – Herr Scheibler.

Fülöp Scheibler: Dann geht es doch eigentlich um die Differenzierung der Stadien M1a und M1b etc., weil Sie die Stadien nach M1a von den restlichen separat haben wollen.

Wolfgang Weber: Von den „Ns“.

Fülöp Scheibler: Von den „Ns“, aber auch von den restlichen „Ms“.

Wolfgang Weber: In zweiter Linie. Vor allen Dingen ist wichtig, nicht diese ehemaligen M1a-Lymphknoten mit den N1-Lymphknoten zu mischen.

Fülöp Scheibler: Im Moment ist es im Bericht so, dass alles, was M ist, zusammengefasst wurde, weil nach alter Klassifikation vorgegangen wurde. Das liegt aber nicht an uns, sondern an den Studien, die wir ausgewertet haben.

Ich verstehe Ihre Kritik so, dass Sie gern eine Auswertung hätten, wo diese große Kategorie M noch einmal unterteilt ist in M1a versus Rest.

Wolfgang Weber: Nein, es geht darum: Ich hatte die Aussage so verstanden, dass die Einteilung gemacht ist: Lymphknotenmetastasen – keine Lymphknotenmetastasen. Aber es ist ganz entscheidend, wo diese Lymphknotenmetastasen sitzen, weil sie in einem Fall im Augenblick als N1 bis N3 oder früher N1 gezählt werden würden oder als Fernmetastasen M1a, jetzt: M.

Stefan Sauerland: Man kann sagen, dass wir hierbei natürlich zum großen Teil an das gebunden sind, was in den Studien berichtet wird. Erst einmal sind die Studien aus früheren Jahren. Sie werden noch nicht wissen, was in 2 Jahren die nächste Änderung der Klassifikation bringt. Das heißt, sie werden sich alle an der alten Klassifikation ausrichten.

Viele Studien werden eine Unterscheidung machen: Fernmetastasierung – ja oder nein. Dann beziehen sie sich natürlich auf diese Punkte.

Es wird vielleicht Studien geben, wenn wir Glück haben, in denen tatsächlich die Art und Weise der Metastasierung sowohl in den Lymphknoten als auch Fernmetastasen genauer aufgeschlüsselt ist, sodass man dann tatsächlich versuchen kann, die Trennwerte selbst zu ziehen und auf diese Weise die Testgüte nach der jetzt überarbeiteten TNM-Klassifikation ex post zu bestimmen. Aber dass das ein großer Teil der Studien sein wird, wage ich zu bezweifeln.

Wolfgang Weber: Ich glaube, da gibt es durchaus einige Studien, bei denen das so gemacht wird. Das wäre wirklich zu überprüfen.

Katja Ott: Ich denke, es ist ganz wichtig zu definieren: neue und alte TNM-Klassifikation. Denn in der alten TNM-Klassifikation hatten wir das alte 1: das war ein Ösophaguskarzinom. Da gibt es die M1a-Situation mit Lymphknoten und Truncus coeliacus. Für die therapeutisch-klinische Behandlung hatte das keine Konsequenz, weil alle diese Patienten auch neoadjuvant behandelt und weil sie auch alle reseziert wurden. Denn die Lymphknotenmetastasen am Truncus gehen automatisch in die D2-Lymphadenektomie hinein, die beim Magenhochzug erforderlich ist. Also bleibt im Endeffekt die Abgrenzung zu den Organmetastasen.

Jetzt bei der neuen Klassifikation muss man wirklich exakt definieren, weil das wirklich extrem kontrovers diskutiert wird, auch wenn man Arbeiten submittiert. Für AEG II und III, die jetzt Ösophaguskarzinome sind, ist die ehemalige M1a-Region nie M1 gewesen, sondern immer N. Das birgt maximales Konfliktpotenzial. Man ist dabei darauf angewiesen, das klar zu definieren.

Wolfgang Weber: Das ist genau der Punkt. Um diese für die Prognose der Patienten relevante Information zu unterscheiden, ist es wichtig, diese Studie unter diesem Gesichtspunkt aufzuschlüsseln. Das ist zumindest bei einer nicht unerheblichen Zahl von Studien auch durchaus möglich.

Fülöp Scheibler: Wir haben das überprüft. Bei den Studien, die wir bisher eingeschlossen haben, ist es wahrscheinlich bei etwa der Hälfte der Studien möglich, diese 3 Kategorien N, M1a und sonstige M zu trennen. Möglicherweise kommen durch die Nachrecherche noch einige dazu. Das Problem ist nur: Wenn man das so macht, können wir natürlich die anderen Studien, die das nach alter Klassifikation bewertet und nicht separat für jeden Patienten ausgewiesen haben, in unsere Analyse nicht mehr einbeziehen. Weil wir in Metaanalysen nur direkte Vergleiche zwischen der PET und einer anderen diagnostischen Technologie bewerten können, führt das dann dazu, dass wir für einige diese Vergleiche möglicherweise – das kann ich nur prognostizieren – gar keine Metaanalysen darstellen können.

Jetzt haben wir überlegt, denn das ist wie die Wahl zwischen Pest und Cholera: Entweder wir machen es ganz korrekt nach neuer Klassifikation. Dann kann es sein, dass wir gar kein Ergebnis bekommen, weil wir nicht aggregieren können. Oder wir sagen: Wir machen es so, wie es die Studien angegeben haben – zum Teil nach alter Klassifikation –, dann haben wir einen relativ präzisen metaanalytischen Schätzer, wie er im Vorbericht dargestellt ist, der aber möglicherweise mit der klinischen Realität nicht hundertprozentig übereinstimmt. Die Frage ist, ob man nicht trotzdem eine Aussage über die diagnostische Güte der PET machen kann, insbesondere im Vergleich zu konventionellen bildgebenden Verfahren.

Moderator Jürgen Windeler: Herr Scheibler hat gerade wunderbar die beiden Alternativen dargestellt. Im Prinzip ist das der Grund, warum wir mit Ihnen hier sitzen. Denn wir wollen hören: eher links- oder eher rechtsherum?

Katja Ott: Als Kliniker ist es maximal relevant, die Fernmetastasierung zu kennen. Der Nodalstatus hat eine gewisse Relevanz, ist aber in der Entscheidungsfindung nicht so entscheidend, denn ein lokal fortgeschrittener, der zu einem hohen Prozentsatz nodal-positiv ist, wird neoadjuvant therapiert.

Die damals alten M1a sind auch alle ins Neoadjuvantkonzept gegangen. Da wurde niemand ausgeschlossen – auch von der Resektion.

Aber der Benefit des PET – die wirklichen M1b zu erkennen –, ist klinisch absolut relevant, da diese Patienten mit Ösophaguskarzinom nicht in kurative Konzepte gehen, sondern sinnvollerweise palliativ therapiert werden.

Fülöp Scheibler: Noch einmal die naive Frage: Wenn von 2 Studien die eine Studie wirklich nur M1b betrachtet, während die zweite Studie M1a plus M1b betrachtet und das Ganze nicht trennbar ist: Soll ich Studien separat auswerten, oder kann ich das alles in einen Topf werfen? Dann habe ich natürlich eine gewisse Unschärfe, die möglicherweise aber nicht klinisch so relevant ist.

Katja Ott: Das ist eine klinische Antwort im Endeffekt. Die Wahrscheinlichkeit der Lymphknotenmetastasierung am Truncus beim AEG I war relativ gering. Ich kenne keine Studie, die das getrennt evaluiert. Deswegen wäre es sinnvoll für die Auswertung des PET, das zunächst alles in einen Topf zu werfen und dann aber zu diskutieren, dass man das Problem der M1a und M1b in der alten Klassifikation hat. Und wenn es möglich ist zu trennen, sollte man trennen.

Moderator Jürgen Windeler: Herr Weber.

Wolfgang Weber: Der Punkt ist der, dass in einigen Studien etwas als „N plus“ allgemein genannt wurde, was eigentlich „M plus“ war. Diese Sachen muss man herausholen, um den wirklichen klinischen Nutzen der PET nachweisen zu können. Das ist in der jetzigen Auswertung nicht erfolgt, weil sie in dieser Gruppe der „Ns“ verschwunden sind.

Sie haben vielleicht weniger Punkte bei den Lymphknoten, was aber für die klinische Fragestellung nicht so relevant ist, wie Frau Ott gerade gesagt hat. Sie haben dann mehr Patienten bei der Fragestellung von M1 versus nicht M1. Das ist die interessantere klinische Fragestellung. Deshalb der Vorschlag von uns, das noch anders zu unterteilen.

Moderator Jürgen Windeler: Wenn es jetzt – so habe ich Frau Ott verstanden, und Sie, Herr Weber, jetzt auch – darauf hinausläuft, zusätzlich zu bisherigen Darstellungen eine ergänzende differenzierte Darstellung zu versuchen, kann ich das so gut einordnen.

Ansonsten müsste ich zurückfragen. Wenn Sie von „mehr herausholen“ sprechen, so ist dies leicht gesagt. Aber die Frage lautet: Was machen wir, wenn es da nicht drin ist? Das ist die Frage, die Herr Scheibler stellt. Wenn man bestimmte Studieninhalte und -konzepte für relevanter als andere hält und eine solche Hierarchie aufbaut, ist für uns die – jedenfalls denkbare – Konsequenz, dass wir uns prioritär um die einen Studienkonzepte kümmern und nicht mehr um die anderen. Sehr relevant für die Antwort auf die Frage, um welche wir uns kümmern, ist die Einschätzung, welche davon man als relevant betrachtet.

Frau Ott habe ich so verstanden: erst einmal alle in einen großen Topf und möglicherweise da noch differenzieren. – Ich habe Sie, Herr Weber, auch so verstanden, dass das ein Weg ist,

den wir gehen sollen. Denn sonst hätten wir die Situation, die Differenzierung für alle Studien zu machen und damit den Studienpool deutlich zu reduzieren.

Wolfgang Weber: Was ich klarstellen wollte, war: Sie reduzieren den Studienpool dann für die differenzierte Beurteilung von N1 versus N0 bzw. N plus versus N0.

Wenn hier nicht genügend Daten vorliegen, ist diese Fragestellung von untergeordneter Bedeutung, weil sich das chirurgische Vorgehen häufig deswegen nicht ändern wird.

Man gewinnt dadurch aber, dass einige Patientendaten, die bis jetzt in der Auswertung „Lymphknoten“ enthalten sind – N-Stadium –, dann in die Auswertung „M-Stadium“ hinübergehen werden. Deshalb kann man vielleicht mehr in dieser Gruppe der Beurteilung der diagnostischen Wertigkeit der PET zum „M-Stadium“. Das meinte ich damit. Wir kommen zu einer Verschiebung, sodass man durchaus sehen kann, dass für Lymphknoten die Aussagen eher schlechter werden. Aber das ist, wie gesagt, eine untergeordnete Fragestellung – das würde ich völlig wie Frau Ott sehen –, denn es würde durch den Nachweis von Fernmetastasen gewinnen.

Moderator Jürgen Windeler: Danke. – Herr Menig.

Matthias Menig: Ich bin nicht ganz tief in den Studien, aber nach der Logik, wie ich sie verstanden habe, war die Datenerhebung in den Studien auf Basis von TNM 6, sprich: M1a. Das ist heute nach aktueller Anschauung ein N-Stadium. Wenn man die Studien entsprechend in M und N sortiert hat, müsste man doch eher davon ausgehen, dass eigentlich Patienten vom M- in ein N-Stadium wechseln, wenn wir es nach der neuen TNM-Klassifikation sortieren würden. Das heißt, man würde Patienten verlieren, die nicht mehr M (M1a) sind, sondern in N rutschen.

Wolfgang Weber: Ich glaube, dass das nicht unbedingt der Fall sein muss. Denn es kommt immer auf die Höhe an, wo der Tumor saß, ob das AEG 1 oder ein zervikales bzw. ein thorakales Ösophaguskarzinom war. Dann muss man wirklich schauen, welche Lymphknotenregionen als befallen beschrieben sind. Von dort aus muss man zurückgehen und fragen: Wie wird das jetzt eingestuft – als M- oder als N-Stadium?

Deshalb denke ich schon, dass man etwas für die Unterscheidung und den Nachweis M1 versus nicht M1 herausfinden kann.

Matthias Menig: Nach dem, wie die TNM 7 im Vergleich zur TNM 6 ist, wurden zervikal und abdominell die M1-Gebiete zugunsten eines N-Status dezimiert. Das heißt: In allen Fällen, die sich geändert haben, kann es eigentlich nur eine Migration von Studienpatienten in Richtung N-Klassifikation geben.

Patienten, die in einer Studie nach TNM 6 einen M-Status hatten, würden nach der Umsortierung bestenfalls einen N-Status erhalten – aber nicht umgekehrt.

Katja Ott: Das ist korrekt. Das heißt: Wir haben einen Pool, der etwas größer ist, nach der alten TNM-Klassifikation. Der kann durch die Stage Migration in der neuen kleiner werden.

Matthias Menig: Vielleicht könnte man – unabhängig davon – noch einmal klarstellen, ob die diagnostische Güte für das N-Stadium nicht prinzipiell auch relevant ist und allein schon von der Systematik her untersucht würde. Für die chirurgische Entscheidung hat es wahrscheinlich gar keinen oder relativ wenig Einfluss, aber für die multimodale Therapie, Strahlentherapie, Indikation und spätestens Planung ist die Kenntnis vom N-Befall auch sehr wichtig.

Wolfgang Weber: Das ist genau der Punkt, den ich auch machen wollte. Es ist, gerade wenn ein nicht-chirurgisches Verfahren eingesetzt wird, natürlich sehr wichtig zu wissen, welche Lymphknoten befallen sind. Diese distanten Lymphknoten kann man natürlich im PET besser als unmittelbar dem Tumor anliegende Lymphknoten nachweisen. Das ist der Punkt: Man muss trennen. Lymphknoten, die unmittelbar dem Tumor anliegen, sind schwieriger nachzuweisen, aber klinisch weniger relevant, als Lymphknoten, die weiter entfernt liegen, besser nachzuweisen sind, aber für die Therapieplanung größere Bedeutung haben.

Moderator Jürgen Windeler: Herr Scheibler.

Fülöp Scheibler: Ich muss noch einmal nachfragen. Sie schreiben:

„Die Feststellung eines N1-Stadiums ist keine Fragestellung von hoher klinischer Relevanz.“

Sollen wir das aus dem Bericht herauslassen?

(Heiterkeit von Stefan Sauerland)

Denn wir haben es so dargestellt. Jetzt ist die Frage: Ist das ein Fehler, das so darzustellen?

Wolfgang Weber: Es ist kein Fehler, das darzustellen. Es gibt auch viele Studien dazu. Aber es ist eine wichtigere Fragestellung, die Verwirrung von alter und neuer TNM-Klassifikation zu vermeiden. Es ist besser, Lymphknoten, die unmittelbar dem Tumor anliegen und Lymphknoten, die mehr distant vom Tumor sind, zu unterscheiden. Wir wollten betonen, dass das die wichtigere Fragestellung ist. Für sie ist die diagnostische Genauigkeit der PET höher. Für sie ist aber auch die klinische Relevanz höher.

Fülöp Scheibler: Okay.

Moderator Jürgen Windeler: Dann haben wir Einigkeit hergestellt. Wir kommen nun zu TOP 3.

4.3.4 Tagesordnungspunkt 3: Beurteilung des Verzerrungspotenzials nach QUADAS-Kriterien für Studien zum M-Staging

Fülöp Scheibler: Die DGN betont in ihrer Stellungnahme: „dass die Beurteilung des Verzerrungspotenzials nach den QUADAS-Kriterien insbesondere für Studien zum M-Staging problematisch ist, da eine große Zahl der in QUADAS postulierten Qualitätskriterien notwendigerweise verletzt ist.“

Folgt man ihrer Argumentation, muss man konstatieren, dass es in der Indikation des M-Staging offensichtlich nicht möglich ist, unverzerrte klinische Studien durchzuführen. Wir verstehen allerdings nicht, warum Sie vor dem Hintergrund dieser Feststellung das Bewertungsinstrument QUADAS kritisieren. Könnten Sie uns diesen Zusammenhang noch einmal genauer erklären?

Wolfgang Weber: Der Punkt ist: Es ist völlig richtig, was Sie sagen. Bei QUADAS gibt es keine unverzerrten Studien zum Nachweis von Fernmetastasen. Das ist Punkt 1. Ich kann gern noch im Detail erläutern, warum ich überzeugt bin, warum dem so ist.

(Zustimmung von Fülöp Scheibler)

Das heißt nicht, dass QUADAS ein schlechtes Werkzeug ist. Es gibt viele andere Anwendungsgebiete, wo es ausgezeichnet ist. Nur für dieses spezielle Anwendungsgebiet hat es das Problem, dass eine große Zahl der Kriterien a priori verletzt sind.

Wenn ich Studien zum Nachweis von Fernmetastasen beurteile, hat es unserer Auffassung nach wenig Sinn, ein Tool zu verwenden, das von vornherein all diesen Studien ein hohes Verzerrungspotenzial zuweist. Es ist dann kein Qualitätskriterium mehr, das Sie beurteilen, wenn Sie von vornherein schon sagen können: Das ist verzerrt.

Die Frage ist, ob es nicht sinnvoller wäre, andere Ansätze zu wählen, bei denen es auf den Vergleich von 2 Verfahren und nicht mehr auf die absolute Sensitivität und Spezifität von Fernmetastasen hinausläuft. Denn darauf ist QUADAS ausgerichtet; QUADAS möchte möglichst unverzerrte Schätzer für Sensitivität und Spezifität in absoluten Werten bekommen. Das ist bei Fernmetastasen aus verschiedensten Gründen sehr problematisch und meiner Meinung nach nicht möglich. Ich kann es gern auch noch einmal erläutern, warum ich das glaube.

Daher stellt sich die Frage: Hat es Sinn, diese QUADAS-Kriterien einzusetzen, um die Güte zu beurteilen, wenn man eigentlich von vorneherein sagen kann, dass die Kriterien verletzt sind? Ich würde sogar sagen: Es sind sogar noch mehr Studien verletzt, als Sie in Ihrem Bericht aufgeführt haben.

Moderator Jürgen Windeler: Ich habe ein Verständnisproblem. Sie haben – ich hoffe, dass ich richtig zugehört habe und dass ich das gut wiedergeben kann – gesagt: Für die

Betrachtung von Sensitivität und Spezifität bei der Identifizierung von Fernmetastasen ist eine unverzerrte Schätzung von Sensitivität und Spezifität nicht möglich.

Wolfgang Weber: Ja.

Moderator Jürgen Windeler: Das ist aber kein Problem von QUADAS. Denn QUADAS macht nichts anderes, als genau diese Aussage.

Wolfgang Weber: Ja, richtig.

Moderator Jürgen Windeler: Das ändert doch aber nichts an dem Faktum, das Sie gerade ausgesprochen haben, ohne QUADAS zu erwähnen, dass man Sensitivität und Spezifität nicht unverzerrt schätzen kann. Das ist eben so.

Wolfgang Weber: Ja.

Moderator Jürgen Windeler: Das hat aber mit QUADAS nichts zu tun. QUADAS sagt das nur noch einmal, nur mit anderen Worten.

Wolfgang Weber: Der zweite Punkt ist der: Es besteht doch allgemeine Übereinstimmung, dass die Suche nach Fernmetastasen oder Patienten mit Tumorerkrankungen sinnvoll ist.

(Zustimmung von Moderator Jürgen Windeler sowie von Fülöp Scheibler)

– Da wären wir d'accord.

Moderator Jürgen Windeler: So grundlegend – ja.

Wolfgang Weber: Wenn man das sagt und wenn man auch sagt, dass alle Studien, die vor 20 Jahren gemacht worden sind – mit Thorax-Röntgen, Ultraschall und klinischer Untersuchung –, Verzerrungspotenzial haben, was bedeutet das dann für den aktuellen Standard der Diagnostik? Sollen wir all diese Verfahren nicht mehr anwenden, weil es keine Daten dazu gibt, wie genau diese Verfahren sind?

Moderator Jürgen Windeler: Erst einmal Herr Scheibler.

Fülöp Scheibler: Wie gesagt: Nachdem wir erst einmal aus diagnostischen Gütedaten keinen Nutzen für Patienten ableiten können, ist es egal, ob sie verzerrt oder unverzerrt sind.

(Wolfgang Weber: Nein!)

Dann würde sich unser Fazit nicht ändern.

Wolfgang Weber: Nein, das stimmt nicht. Ich muss jetzt leider etwas spitzfindig werden. Wenn ich es richtig verstanden habe, folgen Sie der Hierarchie von Fryback & Thornbury. In

dieser Hierarchie ist die diagnostische Genauigkeit die unterste Stufe, auf der alles andere aufbaut. Das heißt, nach der Systematik ist das notwendige Voraussetzung für alles andere.

(Fülöp Scheibler: Ja!)

Wenn man also sagt „Es gibt keine unverzerrten Studien zu dem Thema“, muss man eigentlich schlussfolgern, dass man das gar nicht mehr machen darf.

(Fülöp Scheibler schüttelt den Kopf.)

Aber dann kommen wir in die etwas paradoxe Situation, dass man notwendigerweise fordern muss, auch mit jedem anderen Stagingverfahren sofort aufzuhören, weil es keine Daten dazu gibt.

Moderator Jürgen Windeler: Ich möchte versuchen, ein paar Dinge generell anzumerken. Es geht nicht darum, sofort aufzuhören. Es geht einfach banalerweise darum festzustellen, ob Studien verzerrt sind oder nicht. Es ging noch nicht einmal darum, unmittelbar deterministisch daraus abzuleiten, was das für die Ergebnisse bedeutet. Es geht nur darum, das Verzerrungspotenzial festzustellen. Insofern kann ich Ihren Anfangskritikpunkt schlicht nicht verstehen.

Es ist praktisch so: QUADAS ist ein Instrument, das vermutlich wie alle Instrumente seine Stärken und Schwächen hat. Es ist aber in der Lage, das Verzerrungspotenzial von Accuracy-Studien festzustellen.

Bei Ihrer Darstellung und in dieser Situation haben wir durch QUADAS ein Ergebnis, das ein Verzerrungspotenzial nahelegt. Das ist auch so. Insofern kann ich schon im Ansatz nicht verstehen, wo bezüglich dieses Grundsatzproblems die Kritik liegt.

Wir werden für diese Situation wahrscheinlich – vielleicht sogar wirklich; ich kann und will das nicht beurteilen – keine verzerrungsfreien – vielleicht verzerrungsärmere – Studien als diejenigen bekommen, die im Moment existieren. Das hat aber mit den Fragen „Machen wir das?“, „Dürfen wir das machen?“ und mit der Historie, was wir gemacht haben, überhaupt nichts zu tun.

Thornbury & Fryback in allen Ehren: Strenggenommen sind Accuracy-Daten keine Voraussetzung für Nutzen. Es ist ganz nett, sie zu haben. Man freut sich daran. Man hat für Patienteninformationen eine Idee: Macht man viel falsch oder nicht? – Für die Nutzenfragestellung ist die Accuracy keine notwendige Voraussetzung. Wir könnten wunderbar mit randomisierten Nutzenstudien umgehen, ohne etwas über die Accuracy zu wissen – Hauptsache, der Nutzen ist da.

Diese Kondition ist nicht da, schon gar nicht die Forderung, dass wir nur unverzerrte Accuracy-Daten brauchen, meinetwegen noch einer bestimmten Größenordnung; das wäre die

nächste Konsequenz, die man ansprechen müsste, sonst hat das gar keinen Sinn. Ich müsste im Grunde genommen etwas sagen wie: Ich brauche 80 % Sensitivität, um in eine Nutzenstudie gehen zu dürfen. – Das hat alles keinen Sinn. Das ist keine Voraussetzung.

Insofern möchte ich das – ich versuche deswegen, das Ganze ein wenig einzuordnen – etwas runterkochen. QUADAS und wir machen nichts anderes, als das Verzerrungspotenzial von Accuracy-Studien festzustellen. Es mag daran liegen – wir haben das in der vorherigen Erörterung heute schon gehabt –, dass Forscher ihre Studien nicht gut gemacht hatten und dass man sie vielleicht besser machen kann. Dazu versuchen wir, uns möglichst wenig zu äußern. Es mag daran liegen, dass man diese Studien nicht anders und besser machen kann. Aber dann ist das Verzerrungspotenzial nicht weg.

Wolfgang Weber: Wenn man das Verzerrungspotenzial als gegeben ansieht und feststellt, dass diese Studien notwendigerweise ein Verzerrungspotenzial haben, weil das kein guter Goldstandard ist, um Metastasen in einem lebenden Menschen auszuschließen – das ist das Grundproblem, auf das alles andere zurückläuft, wenn man mit indirekten Verfahren arbeiten muss –, dann sollte doch der Vergleich einer neuen Methode auf eine andere Weise erfolgen. Man sollte sie einfach mit anderen diagnostischen Verfahren vergleichen.

Ich verstehe nicht, warum man ein Tool, bei dem man von vornherein weiß, dass es alles als „verzerrt“ ausgegeben wird, überhaupt auf diese Studie anwendet.

Moderator Jürgen Windeler: Herr Scheibler, ich lasse Sie sofort antworten. Eine Antwort kann ich Ihnen aber nicht ersparen. Üblicherweise wird dem IQWiG vorgeworfen, dass es nicht systematisch vorgehe.

(Heiterkeit von Stefanie Thomas)

Hier liegt möglicherweise einfach ein Grund darin, wirklich systematisch vorzugehen und es überall gleich zu machen. – Herr Scheibler.

Fülöp Scheibler: Ich glaube, ich verstehe Sie jetzt. Sie sagen: Wenn ich ein Tool nehme, das von vornherein Deckeneffekte hat, kann ich nicht mehr innerhalb der Studien differenzieren.

(Wolfgang Weber: Richtig!)

– Das kann ich verstehen. Dann kann ich auch keine Sensitivitätsanalysen mehr rechnen usw.

Sie sagen aber selbst: Grundsätzlich ist es nicht möglich, die diagnostische Güte bei Metastaserkennung in Studien verzerrungsfrei zu erfassen. Das wäre für mich – sogar im Gegenteil: umso mehr – eine Indikation dafür, vergleichende Studien zu machen.

Es gibt auch die Empfehlungen von verschiedenen HTA-Organisationen: Wie gehe ich mit diagnostischen Verfahren um, wenn es keinen Goldstandard gibt?

(Wolfgang Weber: Ja!)

Ich mache randomisiert-kontrollierte Studien. – Genauso sehe ich das auch.

Ich glaube tatsächlich, dass es bestimmte Fälle gibt, in denen die diagnostische Güte keinen Sinn hat. Wir haben das zum Beispiel bei der Epilepsie. Wenn die PET sozusagen operationssteuernd ist und wenn die Anfallsfreiheit nach der Operation der Goldstandard ist – beeinflusst durch die PET –, handelt es sich nicht mehr um eine diagnostische Fragestellung.

Es ist nicht so, dass wir verlangen: Machen Sie das Unmögliche. Machen Sie jetzt eine unverzerrte diagnostische Studie. – Wir sagen vielmehr: Für diese Situationen sind Diagnosestudien einfach nicht das geeignete Tool. Man muss vergleichende Studien machen.

(Wolfgang Weber: Ja!)

Genau im Gegenteil zu Ihrer Argumentation: Hier spricht sozusagen ein hohes Verzerrungspotenzial eigentlich dafür, randomisiert-kontrollierte Studien zu machen. Deswegen bin ich auch nie mit dieser Fryback-&-Thornbury-Hierarchie, die eigentlich gar keine sein soll, so ganz happy. Denn es sieht immer so aus, als müsse man die Leiter erklimmen.

Aber es ist natürlich so, dass Sie, wenn Sie schon auf der Ebene der diagnostischen Güte schlechtere oder exakt die gleichen Werte haben, wahrscheinlich beim Nutzen auch keinen Unterschied finden. Aber es ist nicht gesagt, dass man zwangsläufig eine höhere diagnostische Güte voraussetzen muss, damit man überhaupt die nächste Stufe nimmt.

Wolfgang Weber: Vielen Dank für die Erläuterung. Es ist wirklich gut, das zu hören und zu wissen. Aber ich gehe nicht d'accord mit der Schlussfolgerung, dass man dann randomisierte Studien macht.

Fülöp Scheibler: Oder vergleichend, von mir aus.

Wolfgang Weber: Aber diese vergleichenden Studien gibt es ja. Mein Punkt ist, um zu unserem Thema zurückzukommen: Man kann beim Nachweis von Fernmetastasen beim Ösophaguskarzinom sehr gute vergleichende Studien wie zum Beispiel CT versus PET machen, die beide ein enormes Verzerrungspotenzial haben, die aber trotzdem sehr wertvolle Aussagen dazu liefern können, dass die PET ein besseres diagnostisches Verfahren als die CT ist.

Katja Ott: Als Kliniker sollte man da nicht jetzt einfach ... PET/CT ist derzeit Standard. Dann muss ich im Endeffekt konventionelles CT versus PET/CT vergleichen.

Wolfgang Weber: Ja, richtig. Das kommt natürlich von der Frage mit dazu. – Aber ich denke, der Punkt ist der: Wenn man Studien eine wie auch immer geartete Güte aufgrund des

Verzerrungspotenzials zuspricht und sie dann weniger berücksichtigt, können sie trotzdem völlig valide Daten für den Vergleich von Verfahren A mit Verfahren B geben. Das habe ich in diesem Bericht nicht gefunden.

Moderator Jürgen Windeler: Herr Scheibler und Herr Beckmann.

Fülöp Scheibler: Wir haben diese Frage schon mehrfach diskutiert, ob, wenn in einem direkten diagnostischen Vergleich zwischen 2 Technologien für beide das gleiche Verzerrungspotenzial gilt, der Vergleich der beiden immer noch valide ist. Das kommt sehr darauf an. Das hängt davon ab, unter welchem Therapieregime das geschieht. Es ist zum Beispiel so, dass die PET als metabolisches Verfahren vielleicht anders auf eine adjuvante Chemotherapie reagiert als ein morphologisches Verfahren. So grundsätzlich würde ich es nicht unterschreiben, dass das immer gleich sein muss.

Wenn man schon sehr stark verzerrte Daten hat – Sie sagen selbst, dass diese Studien in mehreren Aspekten verzerrt sind –, könnte ein weiterer Aspekt sein, dass dieses Verzerrungspotenzial eben nicht gleichartig, sondern unterschiedlich auf den einen Test und auf den anderen Test wirkt. Dann wird es sehr schwierig.

In den Diagnosestudien ist häufig eine dieser Therapien therapieleitend, die andere nicht. Sie vergleichen das direkt. Sie müssen letztlich entscheiden, aufgrund welcher Bildgebung Sie die Lymphknoten herausnehmen.

Wolfgang Weber: Es gibt aber auch Studien, bei denen ... In den diskrepanten Fällen, die wurden aufgelöst ... Eigentlich sind es nur die diskrepanten Fälle, die den weiteren Vergleich bestimmen werden.

(Fülöp Scheibler: Genau!)

Es gibt durchaus Studien dazu, die das gemacht haben.

(Fülöp Scheibler: Ja!)

Sie können zwar von der absoluten Sensitivität und Spezifität verzerrt sein können– das ist völlig richtig –, aber sie sind notwendigerweise verzerrt. Sie können aber trotzdem durch den Vergleich der diskrepanten Fälle eine Aussage dazu treffen, wie die diagnostische Güte von PET oder PET/CT im Vergleich zu anderen diagnostischen Verfahren ist.

Moderator Jürgen Windeler: Jetzt kommt Herr Beckmann dran.

Lars Beckmann: Herr Scheibler hat meinen Kommentar eigentlich vorweggenommen. Ich schaue mir das aus biometrischer Sicht an. Beim Vergleich zweier Tests muss eigentlich nachgewiesen werden, dass der Vergleich unverzerrt ist.

Ich bin kein Experte im Bereich M-Staging beim Ösophaguskarzinom, aber ich stelle mir das sehr schwierig vor. Ich würde diese Studien auch sehr kritisch sehen.

Wolfgang Weber: Damit es etwas konkreter wird, was ich damit meine: Sie haben einen Patienten mit einem Ösophaguskarzinom im Oberbauch. Er hat im PET einen positiven Lymphknoten supraklavikulär. Er soll jetzt eine Strahlentherapie bekommen. Standardmäßig würden diese supraklavikulären Lymphknoten hier nicht mitbestrahlt werden. Sonst zeigen beide Verfahren keine Metastasen. Jetzt wird dieser eine Lymphknoten gezielt herausgenommen. Dann stellt sich heraus: Er ist entweder positiv oder negativ. – Ist das nicht ein valider Datensatz, um zu vergleichen, wie die diagnostische Genauigkeit im Vergleich ist? Nicht in absoluten Werten, das ist völlig klar. Da können noch weitere Metastasen sein, die mit beiden Verfahren nicht entdeckt worden sind. Aber trotzdem ist das dann doch ein valider Vergleich der beiden Verfahren zum Nachweis von Metastasen.

Moderator Jürgen Windeler: Unmittelbar dazu?

Lars Beckmann: Wenn Sie für die gesamte Stichprobe die Richtung einer Verzerrung im Vergleich angeben könnten, könnte man eine Aussage treffen: Der Test ist konservativ, oder er ist liberal. Das müssten Sie für Ihre gesamte Stichprobe angeben können. Das stelle ich mir sehr schwierig vor.

Wolfgang Weber: Könnten Sie das etwas näher erläutern?

Lars Beckmann: Sie haben ein Beispiel genannt, in dem, wenn ich das richtig verstanden habe, die PET eine Metastase entdeckt, die von einer Standardtherapie oder CT nicht entdeckt wurde.

Wenn das der kausal einzige Fall ist, also „Die PET entdeckt in jedem Fall immer mehr Metastasen als der Komparator“, kann man für den Bias in dem Vergleich eine Richtung angeben. Wenn Sie das aber nicht können, ist der Vergleich meiner Ansicht nach nicht zu ziehen.

Wolfgang Weber: Man würde nicht nur diesen Fall testen, sondern man würde auch den umgekehrten Fall testen, in dem Metastasen nur im CT positiv und im PET negativ sind. Ich würde alle nicht-kongruenten Fälle sichern. Das ist in einigen Studien auch so durchgeführt worden. Nur die Bereiche, die in beiden Verfahren negativ sind, hat man nicht weiter untersucht. Natürlich ist dadurch die Gesamtsensitivität verzerrt – keine Frage –, aber trotzdem ist die Qualität der Studie für den Vergleich nicht eingeschränkt. Man darf bloß die absolute Sensitivität nicht glauben, die dabei herauskommt. Sie schwankt etwas – in einigen Studien auch erheblich –, aber trotzdem erscheint mir der Vergleich der beiden Verfahren doch valide. Das ist eine wesentlich geringere Hürde als zu sagen: Man muss eine randomisierte Studie machen, um das zu belegen. Vor allem: mit welchem Endpunkt?

Moderator Jürgen Windeler: Jetzt kommen Frau Ott, Herr Scheibler, Herr Sauerland und Herr Menig dran.

Katja Ott: Klinische Anmerkung: Wenn ich PET/CT mit dem CT vergleiche, habe ich heutzutage durch das CT keinen Bias mehr. Denn die Radiologen sagen: Ein kombiniertes PET/CT hat dieselbe Qualität wie ein CT.

Bei M1: Das ist klinisch unbeschrieben und eine der relevantesten Fragestellungen überhaupt. Das Problem ist der Goldstandard. Denn wenn ich eine M1-Situation habe und sie dann histologisch sichere, so ist das das Optimum und das Einzige. Aber gerade beim Ösophaguskarzinom gibt es Lokalisationen, wo das nie gesichert wird. Deswegen sind die Studien extrem heterogen mit den Vergleichsparametern. Denn ich habe leider nicht in allen Fällen einen adäquaten histopathologischen Vergleich. Oft habe ich einen Vergleich mit Bias, weil sie entweder chemotherapiert oder radio-/chemotherapiert sind – und erst post hoc definiert. Das ist ein echtes Problem.

Fülöp Scheibler: Ich wollte eigentlich das Gleiche sagen. Grundsätzlich, um diese Diskussion abzukürzen: Unser Berichtsplan ist dick. Da steht alles Mögliche drin.

(Heiterkeit von Jörg Kotzerke und Dirk Hellwig)

Da steht auch etwas zu den Diskordanzstudien und zum VOPT-Design drin; VOPT: verification of only positive testers –, also immer nur die positiven Fälle. Die können Sie orten und dann theoretisch histologisch sichern. Wenn es solche Studien gäbe und wenn wir sie identifiziert hätten, hätten wir sie auch eingeschlossen. Sie sind nach Berichtsplan 4.2.1.4 Teil der diagnostischen Güte.

Das Problem ist tatsächlich, wie Frau Ott sagt, dass man, wenn solche Studien dann noch nicht einmal einen validen Goldstandard einsetzen, dann vergleichen sie 3 diagnostische Verfahren im luftleeren Raum miteinander. Da würden wir sagen: Diese Studien schauen wir uns nicht an.

Moderator Jürgen Windeler: Herr Sauerland.

Stefan Sauerland: Die Idee von Ihnen, Herr Weber, die relative Sensitivität zu berechnen und anhand einer solch inkompletten Abklärung der Patienten, bei der man die Patienten dann einem Referenztest unterzieht, wenn eines der beiden Testverfahren positiv ist oder beide positiv sind, ist mehrfach bereits in der Literatur diskutiert worden. Das Hauptproblem ist, dass ich dann erst einmal nur die relative Sensitivität bestimmen kann. Dann muss man auch wieder fragen: Ist das für die klinische Information maßgeblich?

Zweitens bekommen Sie natürlich Probleme. Diese sind viel größer, wenn es darum geht, die Spezifität – auch nur die relative Spezifität – zu bestimmen. Da würden wir normalerweise –

es mag Ausnahmen in klinischen Spezialsituationen geben – sagen: Wir brauchen eigentlich stets Angaben zu beidem: zu Sensitivität und zu Spezifität.

Moderator Jürgen Windeler: Jetzt ist Herr Menig an der Reihe.

Matthias Menig: Es wurden schon alle Punkte gesagt.

Moderator Jürgen Windeler: Okay. – Herr Weber.

Wolfgang Weber: Zuerst einmal zum letzten Punkt von Herrn Sauerland: Ich glaube, das ist die Realität, mit der man leben muss. Es gibt eben die Problematik, dass gerade die Spezifität und auch die Sensitivität bei diesem Fernmetastasenachweis nur mit diesen Einschränkungen zu bestimmen sind. Man muss mit diesen gegebenen Einschränkungen neue diagnostische Verfahren beurteilen. Es hat keinen Sinn, sie mit Kriterien zu beurteilen, die einfach nicht zu erfüllen sind, weil die Studiendesigns zu kompliziert sind. Das ist der Grundpunkt, den ich hier nennen möchte: Man muss ein Verfahren zur Beurteilung einsetzen, das auch geleistet werden kann. Das ist – so scheint es mir, da stimmen wir, Gott sei Dank, überein – für die absolute Sensitivität und Spezifität aus den genannten Gründen nicht möglich.

Man muss schauen: Was ist das kosteneffektivste Verfahren, diese Angaben zu erreichen, wenn man sagt, man müsse sich auf die Vergleiche der Verfahren oder Reihungen von Sensitivität und Spezifität beschränken? Ich glaube: Auch die Spezifität kann man durchaus bestimmen. Wenn sich die PET-Befunden als falsch und die CT-Befunde als richtig-negativ in den beurteilten Regionen erweisen, kann man das entsprechend berücksichtigen.

Es ist nicht völlig richtig, was Frau Ott sagt, dass nicht immer eine histologische Sicherung möglich ist. Aber dann kann man zumindest diese Regionen im Verlauf systematisch nachbeurteilen, wobei man dann natürlich wiederum den QUADAS-Kriterien entgegenläuft, weil dann der Test ein Teil des Goldstandards wird. Aber auch das ist nicht anders möglich bei den Studien zum Ganzkörperstaging.

Moderator Jürgen Windeler: Herr Scheibler und Herr Wolff. Herr Wolff möchte zuerst.

Robert Wolff: Ich möchte etwas Grundsätzliches zur Bewertung von QUADAS sagen. Ich glaube: Es ist bisher schon deutlich geworden, dass wir uns, wenn wir QUADAS einsetzen, Diagnostic Accuracy Studies anschauen, die in der Hierarchie nicht ganz so hochwertig wie direkte Studien bzw. Studien zum patientenrelevanten Nutzen sind. Trotzdem versuchen wir natürlich, sie ordentlich anzuschauen. Da scheint mir QUADAS – ich glaube, das war auch nicht bestritten – das beste Instrument zu sein, das auf dem Markt ist. Nur als Beispiel: Es wird zum Beispiel von der Cochrane Collaboration und vom Center of Reviews and Dissemination (CRD) als das Instrument empfohlen, was man benutzen sollte.

Sie, Herr Weber, sagten, dass bestimmte Items bei der Bewertung vom M-Staging beim Ösophaguskarzinom automatisch verletzt sein müssen. Ich glaube, heute haben Sie schon ein

paar Items genannt. Insbesondere auch in Ihrem Zeitschriftenartikel haben Sie einige Beispiele genannt. Insgesamt sind sie auf 6 Beispiele gekommen, die verletzt sein müssen, woraus Sie – bitte korrigieren Sie mich, wenn ich das falsch darstelle – im Wesentlichen schlussfolgern: Wenn so viele schon von vornherein verletzt sind, hat es dann überhaupt Sinn, das Instrument so anzuwenden?

Wolfgang Weber: Richtig. Wenn 6 von 14 von vornherein verletzt sind, stellt sich die Frage: Ist das ein Instrument, um die Güte von Studien zu ranken?

Robert Wolff: Genau. Dazu 2 Gedanken: Zum einen möchte ich das aufgreifen, was schon gesagt wurde: Das ist im Prinzip – ein bisschen flapsig gesprochen – ein „Don’t shoot the messenger!“ – Wenn die Botschaft des Instruments ist, dass es da systematisch Verzerrungsprobleme gibt, ist das nicht unbedingt ein Problem vom QUADAS-Instrument, sondern es ist erst einmal ein Problem der zugrundeliegenden Studien.

Wolfgang Weber: Entschuldigung, wenn ich Sie unterbreche. Aber ich glaube, das ist ganz wichtig für die Präzisierung: Es ist kein Problem der zugrundeliegenden Studien; es ist ein Problem der zugrundeliegenden Fragestellung.

Moderator Jürgen Windeler: Ja, das sehe ich auch so. Nur, Entschuldigung, Herr Wolff,

(Robert Wolff: Kein Problem!)

ist die Konsequenz dessen, was wir gerade diskutieren, doch eigentlich – die Fragestellung hätte ich genauso gesehen –: Diese Studien sind für diese Fragestellungen nicht geeignet. Sie beantworten sie auch nicht. Insofern ist die Einschätzung des QUADAS-Instruments für diese Studien völlig richtig. Denn sie sind verzerrt; das ist sozusagen die objektive Ebene. Die subjektive Ebene ist: Sie beantworten diese Frage nicht.

Ich glaube auch, dass Sie vorhin deutlich gemacht haben, dass man diese Frage der Fernmetastasen im Grunde nur im Sinne von relativen Maßen und Regelbetrachtungen beantworten kann, und eben nicht mit diesen Accuracy-Studien.

(Wolfgang Weber: Ja!)

Insofern ist meiner Ansicht nach das Ganze ein völlig schlüssiges Konzept. Die einzige Frage, die wir uns stellen müssen, ist, ob wir diese Studien nicht explizit in dem Sinne – nicht nur ein hohes Verzerrungspotenzial, sondern auch: Sie beantworten die Frage nicht – beschreiben müssen.

Wolfgang Weber: Ich würde darauf antworten, dass ich sage: Sie haben alle Studien als empirische Studien ausgewertet. Aber nicht alle diese Studien sind notwendigerweise als derartige Accuracy-Studien durchgeführt worden.

Moderator Jürgen Windeler: Ich habe Sie vorhin anders verstanden: Ich habe Sie so verstanden, dass man in dieser Situation eigentlich gar keine verzerrungsfreien und damit aussagekräftigen Accuracy-Studien machen kann.

Wolfgang Weber: Ja, richtig. Aber man kann trotzdem einen Vergleich von PET und CT oder sinnvollerweise PET/CT versus PET durchführen. Derartige Vergleiche wurden auch in einigen Studien durchgeführt. Die Studien haben – man kann sagen: bedauernswerterweise – allerdings immer noch diese absolute Sensitivität und Spezifität angegeben, die meiner Meinung nach keinen Sinn hat.

Aber trotzdem gibt es auch durchaus Auswertungen in den Studien, die diese relativen Aspekte berücksichtigen. Das, denke ich, ist die relevantere Information aus den Studien als die Angabe, dass das PET 90 % Sensitivität bei Fernmetastasen hatte, denn diese Zahl kann man nicht beweisen.

Moderator Jürgen Windeler: Jetzt haben wir Herrn Wolff ein bisschen unterbrochen.

(Wolfgang Weber: Entschuldigung!)

Deswegen hat er jetzt das Wort.

Robert Wolff: Ich möchte noch etwas zum zweiten Punkt anbringen, was die 6 Items angeht, die verzerrt sein müssen. Der Aussage stimme ich nicht zu – in dem Sinne, dass ich mir diese 6 Items gezielt angeschaut habe. Ich weiß nicht, wie viel es bringt, das in dieser großen Runde dezidiert durchzugehen. Das können wir aber gern machen.

Das einzige der 6 Items, bei dem wir tatsächlich auch ein Problem sehen würden, ist das QUADAS-Item-6, bei dem es um die differenzielle Verifikation geht. Sie haben den Bericht vor Augen, dass im Prinzip 2 Referenzstandards bewertet wurden, nämlich zum einen die Biopsie und insbesondere für die Test-Negativen die klinische Nachbeobachtung. Natürlich hat man oft die folgende Situation: Wenn da kein Befund ist, kann man auch nichts biopsieren. Die muss man nachbeobachten. Insofern gibt es 2 Referenzstandards, die eingesetzt werden. Das kann man vom Studiendesign her nicht vermeiden.

Aber das ist auch der Punkt, den wir im Bericht entsprechend diskutiert haben. Wir haben auch darauf hingewiesen, dass das nach unserer Meinung einzige Problem mit dem QUADAS-Instrument ist, was wir im Weiteren entsprechend berücksichtigt haben. Mit den anderen 5 Punkten sehe ich schlichtweg keine Probleme.

Moderator Jürgen Windeler: Wir sollten in der Tat jetzt keine Grundsatzdiskussion über QUADAS-Item führen, zumal wir in der Betrachtung bzw. der Referenz zu unserer Arbeit und dem Vorbericht relativ weit gekommen waren, auch was den Einsatz von QUADAS und die Interpretation angeht. Wir sind – das muss man erwähnen – zu einem Konsens gekommen, was den Stellenwert bestimmter Studien im Vergleich zu anderen Studien angeht.

Wir sollten jetzt nicht – das können Sie bi- oder trilateral machen – das QUADAS-Instrument durchexerzieren. Aber Sie hatten sich noch einmal gemeldet.

Wolfgang Weber: Ich wollte eigentlich genau das Gleiche sagen. Nur einen Punkt kann ich mir nicht ganz verkneifen.

Moderator Jürgen Windeler: Bitte.

Wolfgang Weber: Sie hatten es selbst gerade gesagt: Die Patienten müssen dann nachbeobachtet werden. Diese Nachbeobachtung erfolgt mithilfe irgendeiner Art von Bildgebung. Damit haben Sie sofort einen Konflikt mit QUADAS-Item-7,

(Robert Wolff: Nein!)

dass nämlich der Indextest Bestandteil des Referenzstandards wird.

Moderator Jürgen Windeler: Dazu noch Herr Wolff.

Robert Wolff: Wenn der Indextest die PET ist, ist es ein Problem, wenn in der Nachbeobachtung PET eingesetzt wird. Aber es gibt genügend Beispiele, wo in der Nachbeobachtung tatsächlich nur CT oder normale, konventionelle Bildgebung eingesetzt wird.

Wolfgang Weber: Wenn Sie auf der anderen Seite PET und CT vergleichen, haben Sie das Problem, dass das CT bevorteilt wird.

Moderator Jürgen Windeler: Dann schlage ich vor, dass wir die restlichen 4 Punkte, die noch offen sind ...

(Robert Wolff: 5!)

– Okay. – Der entscheidende Punkt ist noch einmal deutlich geworden. Da scheint man sich auch einig zu sein. Für diese spezielle Situation werden die üblichen Accuracy-Studien, die versuchen, Sensitivität und Spezifität unverzerrt zu schätzen, scheitern. Die braucht man auch nicht.

„Ersatzweise“ – ausdrücklich in Anführungszeichen – muss man sich um die relative Sensitivität und Spezifität eigentlich kümmern und schauen, ob man Studien identifizieren und bewerten kann, die diese relativen Informationen liefern.

Das ist das, was ich gerade wahrgenommen habe und bei dem ich den Eindruck habe, dass in diesem Grundverständnis kein grundsätzlicher Konflikt besteht – wie immer man auch jetzt mit dieser Bewertung zurande kommt und zu welchem Ergebnis man kommt. Das weiß ich persönlich erst einmal gar nicht, aber das werden wir dann sehen. Aber in der grundsätzlichen

Vorgehensweise ist das ein Weg, der nachvollziehbar ist. – Weitere Wortmeldungen? – Dann kommen wir zum nächsten TOP.

4.3.5 Tagesordnungspunkt 4: Verschiedenes

Angela Neumeyer-Gromen: In der letzten Woche hat die Guidelines International Network Conference in Berlin stattgefunden, wo auch explizit die Bedeutung und Wichtigkeit von Leitlinien für den G-BA und das IQWiG herausgearbeitet wurden. Sie haben auch auf Leitlinien Bezug genommen und nach diesen gesucht.

Dazu möchte ich zum einen zur Suche fragen. Sie geben verschiedene Zeitpunkte an, bis wann Evidenzsynthesen gesucht wurden, also bis Februar 2011. Dann sagen Sie, darunter könnte man auch Leitlinien fassen –

(Zustimmung von Fülöp Scheibler)

gerade die methodisch hochwertigen, denen systematische Reviews zugrunde liegen.

Dann sagen Sie, die Suche in den Datenbanken der Leitlinienanbieter sei am 10.03.2010 erfolgt. Sie zitieren aber weiter hinten Leitlinien, die 2011 erschienen sind, wie zum Beispiel die S3-Leitlinie zum Magenkarzinom, also diese AWMF-Leitlinie. Darauf möchte ich Bezug nehmen.

Die erste Frage ist: Wie systematisch, transparent, nachvollziehbar oder vollständig ist Ihre Suche in Bezug auf Leitlinien? In diesem Feld gibt es offenbar S3-Leitlinien. Diese Leitlinien sind methodisch sehr hochwertig, weil sie externe Evidenz im Sinne von systematischen Reviews – auch das, was Sie hier im IQWiG tun – mit der klinischen Expertise in Verbindung bringen und daraus Empfehlungen ableiten. Die Leitlinienersteller verstehen sich auch als maßgebliche Einrichtungen der evidenzbasierten Medizin.

Zweite Frage: Die Empfehlungen in diesen Leitlinien divergieren teilweise deutlich von Ihren Schlussfolgerungen. Wir können nicht ganz nachvollziehen, wie das sein kann, insbesondere da es sich um sehr hochwertige Leitlinien handelt. So sagt diese S3-Leitlinie zum Magenkarzinom, dass für Patienten mit Adenokarzinom des ösophagogastralen Übergangs und kurativen Therapieoptionen nach konventionellem Staging mit CT und Endosonografie eine PET/CT für das Staging von lokoregionär und nicht-lokoregionär gelegenen Lymphknoten als auch zum Nachweis oder zum Ausschluss von Fernmetastasen die PET oder PET/CT eingesetzt werden kann.

Es gibt eine SIGN-Leitlinie aus dem Jahr 2006. Wir sitzen in einer Anhörung zum Zeitpunkt 2012. Da ist jetzt ein Review erschienen. Sie überarbeiten diese SIGN-Leitlinie von 2006. Dieses Scottish Intercollegiate Guidelines Network ist eine der international führenden Einrichtungen der evidenzbasierten Medizin, die zu dem Schluss kommt, nachdem sie die neue Evidenz gesichtet hat – sie nennen das „Main conclusions from new evidence“ –:

„Diagnostic performance of FDG-PET for distant metastases was significantly higher than that of CT, ...”

Ihre Schlussfolgerung ist – ich zitiere Sie genau –:

„Es konnte keine Aussage zugunsten eines Verfahrens getroffen werden, ...“

In Bezug auf PET versus CT allein. – Unsere Frage ist: Wie kann es zu solchen doch sehr widersprüchlichen Aussagen von Leitlinien mit einem sehr hohen methodischen Niveau auf der einen Seite und Ihren Schlussfolgerungen auf der anderen Seite kommen?

Moderator Jürgen Windeler: Herr Scheibler, vielleicht können Sie etwas zu dem Recherchevorgang sagen?

Fülöp Scheibler: Das übergebe ich Frau Waffenschmidt.

Moderator Jürgen Windeler: Frau Waffenschmidt.

Siw Waffenschmidt: Ich möchte kurz etwas zur Informationsbeschaffung sagen. Sie stützt sich natürlich auf verschiedene Säulen. Da sind bibliografische Datenbanken, Leitlinienanbieter und andere Quellen drin. So eine Suche erfolgt zu unterschiedlichen Zeitpunkten. Sie wird für den Abschlussbericht noch einmal wiederholt; derzeit ist sie noch nicht drin. Daher haben wir soweit alles abgedeckt. Die Frage ist, ob Sie meinen, dass da noch etwas fehlt?

Angela Neumeyer-Gromen: Das ist meine Frage. Wir sind natürlich nicht in der Lage, das komplett zu replizieren und einen Bericht, für den Sie Jahre brauchen, zu wiederholen. Uns fällt nur auf, dass unterschiedliche Zeitpunkte von Suchzeiträumen angegeben werden. Teilweise werden dann Leitlinien, wie jetzt im Fall der Leitlinien, die eigentlich außerhalb des Suchzeitraums liegen, zitiert. Dann werden wieder nur die zitiert, die innerhalb des Suchzeitraums liegen. Ich kann nicht abschließend beantworten, ob Sie wirklich alle erfasst haben.

Moderator Jürgen Windeler: Jetzt Herr Sauerland, Frau Ott und Herr Scheibler.

Stefan Sauerland: Wir machen zum einen eine systematische Suche nach Evidenzsynthesen – darunter sind auch Leitlinien –, aber wir lassen es uns zum anderen nicht nehmen, dass wir, wenn der Vor- oder der Abschlussbericht in die finale Phase kommt – ich drücke das ganz praktisch aus –, noch einmal zu schauen – dann aber unsystematisch –, ob es nicht gerade bei Leitlinien, von denen wir wissen, dass sie von hochwertigen Leitlinienanbietern sind bzw. es sich um deutsche Leitlinien handelt, etwas Aktuelles gibt, auf das man zum Beispiel in der Diskussion Bezug nehmen kann.

Unbenommen der Systematik der Gesamtrecherche kann es immer und in jedem Bericht vorkommen, dass Studien darüber hinaus in die Bewertung einbezogen werden. Das ist im Fall der MUNICON-Studien auch so gewesen. Diese Studien waren in der normalen Recherche noch nicht aufgetaucht, aber wir sind auf anderen Wegen auf diese Studien aufmerksam geworden.

Moderator Jürgen Windeler: Frau Ott.

Katja Ott: Sie haben die S3-Leitlinie zum Magenkarzinom zitiert. Da wurde sehr um die Formulierungen gerungen:

(Zustimmung von Angela Neumeyer-Gromen)

„Muss“ ist einklagbar, „sollte“ ist evidenzbasiert,

(Angela Neumeyer-Gromen: „Kann“!)

„kann“ ist optional. Ich möchte Herrn Lordick zitieren, der sich dazu geäußert hat. In der aktuell erschienenen Leitlinie zum Magenkarzinom befinden sich bereits Angaben zu Tumoren des gastroösophagealen Übergangs, die nach den aktuellen UICC-Klassifikationen zumindest in Teilen auf Ösophaguskarzinome angewandt werden können. Die darin enthaltenen Aussagen zum Einsatz der PET sind sehr zurückhaltend.

(Angela Neumeyer-Gromen: So, wie ich es vorgelesen habe, war es als Empfehlung formuliert!)

„Kann“ ist das, was wir als gut empfinden, aber nicht beweisen können. Das ist beim PET eigentlich so, dass alle Studien hypothesengeneriert sind, dass es potenziell einen Benefit hat, aber es gibt wenig, wo wir sagen können, dass es de facto so ist.

Moderator Jürgen Windeler: Herr Scheibler.

Fülöp Scheibler: Eine kurze Antwort zum zweiten Teil Ihrer Frage: Wir haben regelhaft in unseren Berichten, im Diskussionsabschnitt immer den Vergleich der Empfehlungen der wissenschaftlichen Fachgesellschaften zum Einsatz von PET und PET/CT mit unseren Ergebnissen. Genau auf diese Fragen gehen wir ein.

Man findet manchmal Leitlinien mit sehr optimistischen Empfehlungen. Dann prüft man, worauf diese Empfehlungen basieren, und versucht, dann die Evidenzgrundlage der jeweiligen Leitlinie zu finden. Da verläuft ganz häufig der Pfad im Sande. Jeder, der an einer Leitlinienerstellung beteiligt war, weiß, dass es in diesem Zusammenhang meist eine breite Fragestellung gibt. Dann gibt es für einzelne Punkte in dieser Fragestellung sehr gute Evidenz aus systematischen Reviews usw. Für andere Punkte gibt es sehr starke Opinionleader in so

einer Leitliniengruppe. Deswegen kommen in die Leitlinien manchmal starke Empfehlungen, ohne dass eine starke Evidenz dahinter liegt.

Wir prüfen systematisch für vorhandene internationale Leitlinien, ob sie den Empfehlungen des IQWiG-Berichts ent- oder widersprechen. Wenn sie widersprechen, sind wir natürlich auch daran interessiert herauszufinden, woran das liegt.

Moderator Jürgen Windeler: Frau Neumeyer-Gromen.

Angela Neumeyer-Gromen: Mich verwundert das SIGN-Review, das erschienen ist, sehr. Das ist sozusagen ein Kurzreview, das demnächst bestimmte Empfehlungen überarbeitet werden. Wenn die von „Main conclusions from new evidence“ sprechen und dann sagen, dass die diagnostische Performanz tatsächlich signifikant höher ist, ist das in der Regel nicht einfach nur ein Handheben, um herauszufinden, welche Meinung die Experten haben, sondern dann scheint es da härtere Evidenz zu geben. Das erstaunt uns doch sehr.

Das ist aktuell in Überarbeitung, aber auf den SIGN-Internetseiten verfügbar, dass sie genau diese Empfehlungen zum Staging aufgrund dieser Evidenzlage überarbeiten wollen.

Moderator Jürgen Windeler: Herr Weber.

Wolfgang Weber: Ich darf etwas in die gleiche Kerbe schlagen. Das ist nicht auf die deutsche Leitlinie bezogen, sondern es gibt seit 2008 von der australischen Behörde MSAC ein HTA zum Thema von PET bei Ösophaguskarzinomen. Das kommt zu der Schlussfolgerung, dass die Sensitivität und die diagnostische Genauigkeit der PET zum Nachweis von Fernmetastasen signifikant höher als die der CT ist.

Ich glaube, um auf unsere vorherige Diskussion zurückzukommen, dass diese Unterschiede, die Sie gerade angesprochen haben und die auch zum australischen Bericht entstanden sind, darauf beruhen, dass sie gerade in der Methodik liegen: Wie beurteile ich Studien, und welche schließe ich aufgrund von Verzerrungspotenzial möglicherweise nicht ein? Oder fokussiere ich auf einen Vergleich der Methoden und nicht so sehr auf die absolute Sensitivität und Spezifität, worauf QUADAS fokussiert?

Moderator Jürgen Windeler: Das haben wir jetzt geklärt. Diesen Punkt nehmen wir mit. Abgesehen davon glaube ich, dass es sein kann, dass nach der Recherche, die für diesen Bericht gemacht worden ist, noch Daten dazugekommen sind.

Ich will allerdings auch sagen, dass mich die Frage ein bisschen wundert. Eigentlich ist doch jedem klar – das habe ich in der vorherigen Erörterung gesagt –: Das hier hat mit Leitlinien und insbesondere mit Leitlinienempfehlungen insofern nichts zu tun, als das völlig verschiedene Aufgaben sind.

Insofern Leitlinien zu anderen Empfehlungen als wir kommen, ist das so. Das wird auch öfters mal so sein. Ich habe vorhin versucht, die Gründe zu erläutern. Daher habe ich Verständnis dafür, dass man einen Abgleich macht, wie wir es gerade beschrieben haben, mit Berichten anderer HTA-Organisationen oder meinetwegen mit – allerdings dann bitte mit qualitativ genauso guten – Berichten, die den Leitlinienevidenzberichten, die den Leitlinienempfehlungen zugrunde liegen. Das finde ich völlig in Ordnung.

Aber Leitlinienempfehlungen haben eine völlig andere Klasse und eine ganz andere Zielsetzung. Da sich die AWMF in ihren Leitlinienberichten auch sehr damit rühmt, dort das Konsensprinzip zu verfolgen – das spielt bei uns bei der Bewertung keine Rolle –, ist das eine völlig andere Qualität. Ich glaube Leitlinienempfehlungen mit Nutzenbewertungen zu vergleichen, ist die völlig falsche Ebene.

Angela Neumeyer-Gromen: Beides sind Einrichtungen der evidenzbasierten Medizin.

Moderator Jürgen Windeler: Nein, das ist ein völlig anderer Typus.

Angela Neumeyer-Gromen: Sie sind eine Einrichtung, die Leitlinienersteller sind auch Einrichtungen der evidenzbasierten Medizin. Da gibt es viele Überlappungspunkte. Sie nehmen selbst Bezug auf die Leitlinien; Ihnen ist das auch nicht ganz egal. Sie prüfen das.

(Zustimmung von Fülöp Scheibler)

So ist es für die leistungsrechtliche Bewertung am Ende, wozu wir Ihren Bericht auch als eine der Grundlagen zugrunde legen, doch wichtig, das in die Diskussion zu stellen. Es geht um evidenzbasierte Medizin.

Moderator Jürgen Windeler: Aber, Frau Neumeyer, lesen Sie den Bericht. Sie haben doch gerade gehört: Es steht unter „Diskussion“ drin. Ich verstehe die Aufregung nicht.

Angela Neumeyer-Gromen: Es widerspricht sich.

Moderator Jürgen Windeler: Wir haben dem an dieser Stelle den Stellenwert gegeben, den es braucht. Wir sagen: Die Nutzenbewertung ist so. – Wir schauen uns um, was die Leitlinienempfehlungen machen, und gleichen das miteinander ab. Das tun wir.

Dass Sie jetzt aber hingehen und sagen: Das steht in den Leitlinien drin, das sollte bzw. kann gemacht werden, nachdem wir – übrigens für mich überhaupt nicht überraschend – hören, welche schwierigen Diskussionsprozesse das nach sich zieht, um das dann mit Aussagen im Nutzenbericht ... Das hat nun wirklich nichts miteinander zu tun. – Herr Weber.

Wolfgang Weber: Ich denke, das sind wirklich 2 unterschiedliche Dinge. Das ist das eine. Was ich gesagt habe, ist eine Aussage zum Vergleich der diagnostischen Genauigkeit von 2 Verfahren.

Moderator Jürgen Windeler: Gut, ganz anders.

Wolfgang Weber: Das ist eine wissenschaftliche Fragestellung. Das muss man schon diskutieren. Es kann nicht sein, dass die Lichtgeschwindigkeit 400.000 km/s ist, während sie in Australien 200.000 km/s beträgt. Derartige Unterschiede muss man versuchen aufzulösen. Darauf hatten Sie hingewiesen.

(Angela Neumeyer-Gromen: Genau!)

Moderator Jürgen Windeler: Herr Scheibler.

Fülöp Scheibler: Herr Prof. Weber, Sie haben natürlich recht, dass Unterschiede zwischen verschiedenen Evidenzsynthesen zum Beispiel dadurch zustande kommen können, dass man unterschiedliche Einschlusskriterien, Methoden usw. wählt. Die Evidenzsynthese der MSAC kennen wir natürlich sehr genau. Wir stehen in intensivem Austausch mit den Kollegen. Der wesentliche Unterschied besteht darin, dass sie eigentlich keine Metaanalyse machen. Sie fassen die Ergebnisse von qualitativ hochwertigen systematischen Übersichten und HTA-Berichten narrativ zusammen und aktualisieren dies noch anhand von Recherchen in der Primärliteratur.

Das machen wir auch. Das ist die gleiche Methodik, die wir auch nutzen – mit dem einzigen Unterschied, dass wir dann aus den Evidenzsynthesen die Ergebnisse, die 4-Felder-Tafeln der einzelnen Primärstudien herausholen und aggregiert Metaanalysen rechnen. Das fehlt dort. Dort gibt es eine qualitative Beschreibung der Ergebnisse vorhandener Evidenz und dann eine Empfehlung. Dass diese Empfehlung dann aufgrund dieser unterschiedlichen Methode zu einem anderen Ergebnis kommen kann, liegt in der Natur der Sache.

Wolfgang Weber: Die Empfehlung verstehe ich. Aber es werden, wie gesagt, Aussagen dazu gemacht, wie die diagnostische Genauigkeit ist. So etwas muss man schon etwas genauer analysieren, um herauszufinden, woran das dann liegt.

(Angela Neumeyer-Gromen: Ja!)

Das kann man alles anders beurteilen und werten; das ist gar keine Frage. Aber erst einmal: Die diagnostische Genauigkeit ist ein mathematischer Wert. Er sollte überall gleich sein.

Moderator Jürgen Windeler: Das ist, glaube ich, Theorie.

Wolfgang Weber: Zumindest das Ziel sollte es sein. Man sollte es, wenn es nicht der Fall ist, diskutieren.

Moderator Jürgen Windeler: Noch einmal – Herr Scheibler hat es gesagt, ich wiederhole es gern –: Die Auseinandersetzung mit diesen abweichenden Ergebnissen anderer

Evidenzsynthesen – egal, ob Leitlinien oder systematischen Reviews woanders – findet statt. Sie findet explizit, diskutierend und erörternd statt.

Wolfgang Weber: Für den australischen Bericht habe ich sie nicht gefunden.

Fülöp Scheibler: Das ist richtig. Ich habe das geprüft. Wir können ihn noch in den Abschlussbericht aufnehmen.

Moderator Jürgen Windeler: Okay, gut. – Sie findet statt, heißt das. – Herr Hellwig.

Dirk Hellwig: Ich kann auch gern noch zur Vollständigkeit, um es hier erwähnt zu haben, einen italienischen HTA-Bericht aufführen, der in 2011 publiziert wurde: „Criteria for appropriate use of FDG-PET in esophageal cancer“. Da kann ich Ihnen gern noch den Internet-URL schicken.

Moderator Jürgen Windeler: Okay. – Gibt es weitere Punkte unter „Verschiedenes“? – Das scheint nicht der Fall zu sein. Dann möchte ich mich für diese sehr konzentrierte Erörterung herzlich bedanken. Ich bedanke mich für alle, die heute auch die beiden Erörterungen mit uns abgewickelt haben. Wir haben eine Reihe von auch für uns wichtigen Rückmeldungen bekommen, die wir verwerten können. Ich danke Ihnen nochmals für die Diskussion und wünsche Ihnen einen guten Heimweg.

(Beifall bei den Anzuhörenden)

Anhang: Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
A 1 Stellungnahmen von Organisationen, Institutionen und Firmen.....	A 2
A 1.1 Deutsche Gesellschaft für Nuklearmedizin e. V.	A 2
A 1.2 Deutsche Krankenhausgesellschaft e. V.....	A 12

A 1 Stellungnahmen von Organisationen, Institutionen und Firmen

A 1.1 Deutsche Gesellschaft für Nuklearmedizin e. V.

Autoren:

Hellwig, Dirk

Herrmann, Ken

Kotzerke, Jörg

Schmidt, Matthias

Weber, Wolfgang



DGN e. V. • Nikolaistraße 29 • D-37073 Göttingen

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

– IQWiG Stellungnahme zum Vorbericht D06-01H –
Prof. Dr. med. Jürgen Windeler
Im Mediapark 8
50670 Köln

Geschäftsstelle
Nikolaistraße 29
D-37073 Göttingen

Tel. (Mitglieder): +49 (0)551 48857-401
Tel. (Presse): +49 (0)551 48857-402
Fax: +49 (0)551 48857-79

Email: office@nuklearmedizin.de
Homepage: www.nuklearmedizin.de

Ihre Zeichen

Ihre Nachricht vom

Unser Zeichen

Göttingen

20. Juni 2012

Stellungnahme der Deutschen Gesellschaft für Nuklearmedizin zum Vorbericht (vorläufige Nutzenbewertung) des IQWiG zum Auftrag D06-01H: Positronenemissionstomographie (PET) und PET/CT bei Ösophaguskarzinom, Version 1.0 vom 15. Mai 2012

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

für die Möglichkeit, zu oben genanntem Vorbericht vom 15. Mai 2012 Stellung nehmen zu können, bedanken wir uns recht herzlich.

Grundsätzlich ist die Kürze der Stellungnahmefrist zu bemängeln. Ihr Institut benötigt zwei Jahre, um die Ergebnisse der Literaturrecherche zu bewerten, während unserer wissenschaftlichen Fachgesellschaft lediglich vier Wochen zugestanden werden, um Mängel zu erkennen und Hilfestellung zu geben. Daher konnten wir die inhaltliche Korrektheit des Berichts nur stichprobenartig überprüfen. Leider fand sich selbst in dieser Stichprobe eine Vielzahl von Fehlern, so dass wir über die Validität des Berichts besorgt sind. Vor allem aber erheben wir Einwände gegenüber der bei der Erstellung des Berichtes angewandten Methodik.

Zusammenfassung

Die Ziele des Berichtes sind die Ermittlung des patientenrelevanten Nutzens der PET bzw. PET/CT und die Bewertung der diagnostischen und prognostischen Güte der PET bzw. PET/CT. Leider hat der umfangreiche Bericht beide Ziele aus methodischen und inhaltlichen Gründen verfehlt. Methodisch halten wir die angewandte Definition eines „patientenrelevanten Nutzens“ für willkürlich und nicht adäquat für ein diagnostisches Verfahren. Selbst wenn man die Definition des IQWiG akzeptiert, sind die geforderten Studien zum Nachweis dieses Nutzens sowohl ethisch problematisch als auch extrem kostenaufwändig und deshalb nicht Erfolg versprechend. Bezüglich der diagnostischen und prognostischen Güte der PET ist die Auswahl der primären Literatur sowie der Evidenzsynthesen nicht nachvollziehbar. Eine – in der Kürze der Zeit nur stichprobenhaft durchgeführte – Überprüfung ergab, dass wesentliche Literatur (wie ein australischer HTA-Bericht mit sehr ähnlicher Fragestellung) nicht aufgenommen bzw. nicht einmal erwähnt sind. Die Bewertung des Verzerrungspotentials der ausgewerteten Publikationen erfolgte anhand von Kriterien, die prinzipiell nicht für Studien zum Nachweis von Fernmetastasen (eine der wichtigsten Indikation für die PET- und PET/CT-Diagnostik) geeignet sind. Schließlich sind die Kriterien zur Bewertung des Verzerrungspotentials widersprüchlich zu den Kriterien zur Beurteilung der Übertragbarkeit der Ergebnisse auf die Patientenversorgung in Deutschland. Inhaltlich zeigen Stichproben, dass die Primärliteratur inkorrekt und ohne den nötigen klinischen Sachverstand exzerpiert wurde. Aufgrund dieser Mängel ist der Bericht unserer Meinung nach als Grundlage für

iqwig_d06-01h_vorbericht_stellungnahme_dgn_120620.doc

die Beratung und eine Beschlussfassung durch den Gemeinsamen Bundesausschuss (G-BA) ungeeignet.

1. Stellungnahme zum „patientenrelevanten Nutzen“

Bezüglich der Beurteilung des „patientenrelevanten Nutzens“ stellen wir fest, dass kein internationaler oder nationaler Konsensus darüber besteht, wie ein derartiger „Nutzen“ definiert werden soll (siehe Diskussion in Vach et al. J Nucl Med 2011). Die in dem Bericht angewandte Definition „Veränderungen, die kausal auf den Einsatz der PET bzw. PET/CT zurückzuführen sind und für den Patienten fassbare Konsequenzen haben“ (Seite V, Absatz 2) ist willkürlich und nicht „evidence based“.

Die im Vorbericht formulierte Reduktion des Wertes einer diagnostischen Maßnahme auf die Wahrscheinlichkeit, mit der sie das „patient outcome“ optimiert, widerspricht der Jahrhunderte langen ärztlichen Erfahrung, dass Patienten und ihre Angehörigen Klarheit über das Stadium einer Erkrankung und ihre Prognose haben wollen, selbst wenn die therapeutischen Möglichkeiten begrenzt sind.

Ignoriert man diesen Gesichtspunkt und akzeptiert die IQWiG-Definition von „patientenrelevanten Nutzen“, sind die geforderten Methoden zum Nachweis eines derartigen Nutzens für das Tumorstaging nicht adäquat, sondern ethisch problematisch und mit nicht vertretbar hohen Kosten verbunden. Nach dem Vorbericht (Seite 9, letzter Absatz und Seite 15 letzter Absatz) werden randomisierte Studien gefordert, um einen patientenrelevanten Nutzen nachzuweisen (vgl. auch IQWiG Allgemeine Methoden, Version 4.0 vom 23.9.2011, Seite 9, Absatz 2: „Andere Studientypen als RCTs sind in der Regel für einen Kausalitätsnachweis nicht geeignet“). Derartige randomisierte Studien erscheinen für das Staging von Ösophaguskarzinomen kaum durchführbar, da verschiedene internationale Leitlinien, vor allem die NCCN guidelines (der internationale Standard für die Behandlung von Tumorpatienten) den Einsatz der PET/CT zum Staging empfehlen¹. Der für randomisierte Studien erforderliche „state of equipoise“ kann also für diese Fragestellung nicht mehr angenommen werden.

Selbst wenn man (mit viel Selbstbewusstsein) diesen internationalen Standard ignoriert, erscheinen die Kosten für eine randomisierte Studie unverhältnismäßig hoch, um zu belegen, dass eine Methode mit höherer diagnostischer Genauigkeit zum Nachweis von Fernmetastasen einen „Nutzen“ für Patienten mit Ösophaguskarzinom darstellt. Die Morbidität und Mortalität der lokalen Therapie von Ösophaguskarzinomen ist mit die höchste aller Therapien in der Onkologie. Es ist ein onkologisches Grundprinzip, dass eine lokale Tumorthherapie in kurativer Intention nicht sinnvoll ist, wenn (nicht erkannte) Fernmetastasen unbehandelt bleiben. Das Vermeiden von derartig schwerwiegenden Eingriffen, wie der Ösophagektomie oder der definitiven Radiochemotherapie von Ösophaguskarzinomen bei Patienten mit systemisch metastasierter Erkrankung erscheint uns somit als ein unmittelbar einsichtiger „Nutzen“ für den Patienten. Uns ist dabei natürlich bewusst, dass man theoretische Überlegungen anstellen kann, unter welchen Umständen dieser „Nutzen“ möglicherweise nicht zutreffen könnte. So kann man z. B. postulieren, dass Patienten mit nur in der PET bzw. PET/CT nachgewiesener Metastasierung eine bessere Prognose haben als Patienten, bei denen eine Metastasierung auch mit anderen Verfahren nachgewiesen wurde. Möglicherweise könnten diese Patienten deshalb von einer lokalen Tumorthherapie in kurativer Intention profitieren. In der bisherigen Verfeinerung der diagnostischen Methoden, z. B. von der planaren Röntgenthoraxaufnahme zum CT des Thorax wurde ein derartiger Effekt allerdings nie festgestellt, allerdings auch nicht in randomisierten Studien ausgeschlossen. Im Falle der Sorge vor derartigen unerwünschten Effekten einer verbesserten Diagnostik, müssten konsequenterweise auch randomisierte Studien zum Staging von Ösophaguskarzinomen mit CT vs. Röntgen-Thorax gefordert werden. Dabei dürfte man es allerdings nicht belassen. Vielmehr müsste als erstes eine randomisierte Studie durchgeführt werden, die Staging alleine mit körperlicher Untersuchung vergleicht mit Staging durch körperliche Untersuchung und Röntgen-Thorax.

¹ In diesem Zusammenhang ist es auch inkorrekt, wenn im Vorbericht ausgeführt wird:

„Im Zuge der Recherche wurde eine Leitlinie der Deutschen Krebsgesellschaft und der Deutschen Gesellschaft für Chirurgie (AWMF-Registernr. 032 / 008, Stand: 10 / 2003) identifiziert. Diese interdisziplinäre Leitlinie „Ösophaguskarzinom“ ist zum Zeitpunkt der Entstehung des vorliegenden Berichts nicht mehr aktuell und somit nicht mehr verfügbar.“

Die Leitlinie für die Versorgung ist sehr wohl noch verfügbar, wenn auch keine Überarbeitung innerhalb des von den Autoren angestrebten Zeitraums erfolgte.

Darüber hinaus hatte die DGN bereits im Stellungnahmeverfahren zum vorläufigen Berichtsplan auf die Problematik der Durchführung von prospektiven Studien und insbesondere RCTs bei Fragestellungen im Zusammenhang mit seltenen Erkrankungen hingewiesen (s. Stellungnahme der DGN zum vorläufigen Berichtsplan vom 25. März 2010). Ihrerseits war in der Würdigung unserer Stellungnahme dieses Problem heruntergespielt worden und keine Änderung des Berichtsplans erfolgt. Die Ihrerseits in der Würdigung der Stellungnahme angeführten Zahlenbeispiele (40.000 Patienten pro Jahr) sind für das Ösophagus-Karzinom um den Faktor 10 zu hoch angesetzt (siehe aktuellen Zahlen des Robert-Koch-Instituts, Krebs in Deutschland, 8. Ausgabe, 2012 www.krebsdaten.de) und stehen im Widerspruch zu der im Vorbericht dokumentierten Häufigkeit des Ösophaguskarzinoms in Deutschland. Der methodische Rahmen ist folglich unpassend für die zu bearbeitende Problemstellung.

2. Stellungnahme zur Nutzenbewertung anhand der diagnostischen und prognostischen Güte der PET bzw. PET/CT

2.1 Methodische Fehler bei der Auswahl der Literatur

Die Auswahl von Evidenzsynthesen und Primärliteratur für den Bericht ist nicht nachvollziehbar. So findet der HTA-Bericht der australischen Gesundheitsbehörde (Medical Services Advisory Committee, MSAC) keine Erwähnung, der im Internet frei verfügbar ist (ISBN (Online) 1-74186-706-1). Aufgabe des MSAC ist es „to provide advice to the Minister for Health and Ageing on the strength of evidence available on new and existing medical technologies and procedures in terms of their safety, effectiveness and cost-effectiveness“ (Seite 2 des HTA-Berichtes). Auf der Grundlage der gleichen Datenbasis wie das IQWiG und unter Benutzung ähnlicher Instrumente (z. B. QUADAS) kommt dieser Bericht zu folgenden Schlussfolgerungen: „<FDG-PET increases> the detection of distant metastases from primary oesophageal cancer“, „<leads to> the avoidance of surgical morbidity and mortality in patients who avoid oesophagectomy“ (Seite 14 des Berichtes). Beide Aussagen sind konträr zu den wesentlichen Schlussfolgerungen des IQWiG-Berichtes. Dies wirft zunächst die Frage auf, warum dieser Bericht eines Institutes mit ähnlichem Auftrag wie der des IQWiG ignoriert werden konnte. Selbst wenn man hier von einem Versehen ausgeht, bleibt die grundlegende Frage, warum eine – nach Aussage des IQWiG – objektive Methodik zur Beurteilung der vorliegenden Evidenz zu derart konträren Ergebnissen führen kann.

Bezüglich der Primärliteratur lassen sich die Suchstrategien in Anhang A nicht nachvollziehen, da die Suche anscheinend sehr viele Tumortypen, nicht nur Ösophaguskarzinome, einschließt. Dies macht es praktisch unmöglich in der Kürze der Zeit für die Stellungnahme die Literatursuche nachzuvollziehen.

Das Ausschluss-Kriterium „30 % fehlende Angaben“ für die Primärliteratur erschließt sich uns nicht. Es bleibt intransparent, wann Arbeiten wegen fehlender Angaben ausgeschlossen werden (z. B. Ott et al. weil ein Teil nicht initial FDG-avid war, während bei anderen Arbeiten nur 24 von 58 Patienten ausgewertet wurden wie bei Liberale et al. 2004; Wieder et al. 2004, weil ein Teil kein frühes Verlaufs-PET hatte). Laut Wieder et al. wurden 38 Patienten eingeschlossen, davon haben 27 ein PET bereits nach 2 Wochen erhalten, wohl aber nach Abschluss der Therapie. Aus unserer Sicht sollten Patienten, die kein frühes Verlaufs-PET erhalten haben, nicht als „fehlende“ Werte gewertet werden. Selbst wenn man dieser Argumentation folgen sollte, entsprechen 11 von 38 Patienten 29,0 %, und damit ist der Anteil nicht größer als 30 %. Hier bedarf es einer klaren Dokumentation der Sachverhalte, die als Begründung für einen Ausschluss angeführt werden, um nicht beliebig zu wirken.

Unklar bleibt, weshalb die MUNICON-Studien und die HICON-Studie nicht angemessen berücksichtigt wurden (Lordick et al. 2007, Zum Büschenfelde et al. 2011, Lorenzen BMC Cancer 2011). Hierbei zeigt u. a. die Arbeit von Lorenzen et al., weshalb FDG-PET beim Ösophagus-Karzinom eingesetzt wird. Wenn ein metabolisches Ansprechen nicht nachzuweisen ist, ist die Prognose so schlecht, dass dies die Evaluation neuer innovativer Therapiekonzepte rechtfertigt. Morphologische Parameter reagieren viel langsamer, so dass durch den Einsatz der PET versucht wird, Strategien zu entwickeln, das Überleben der Patienten zu verbessern.

Es ist auch nicht nachvollziehbar, aufgrund welcher Kriterien einige Literaturquellen der weiteren Analyse zugeführt oder zur Wertung Ihrer Analyseergebnisse herangezogen werden und welche nicht. Insbesondere die methodische Vorgehensweise der Analyse von Meta-Analysen wird abgelehnt, sie ist in besonderem Maße intransparent. Durch die Beschränkung der Literatur-Analyse auf einerseits Meta-Analysen und andererseits Primärliteratur, welche nicht in diesen Meta-Analysen enthalten ist, kommt es zu einer unzulässigen Begrenzung der eingeschlossenen Primärliteratur. Das liegt daran, dass die Meta-Analysen Dritter durchaus unter anderen

Fragestellungen als bei Ihnen durchgeführt wurden. Durch Ihren grundsätzlichen Ausschluss von Primärliteratur, welche bereits in eingeschlossenen Meta-Analysen erfasst ist, nehmen Sie sich selbst die Möglichkeit, eine Informationssynthese anhand der Daten aus der Primärliteratur im Hinblick auf Ihre Fragestellung durchzuführen. Damit werden Sie dem Auftrag des G-BA, einen themen- und fragestellungsbezogenen Nutzenbericht zu erstellen, nicht gerecht.

Der Vorbericht hat den Mangel der fehlenden Nachvollziehbarkeit der Zuordnung von Ausschlussgründen der Literatur zu Themengebieten des zu bearbeitenden Auftrages. Da Sie auftragsgemäß mehrere Fragestellungen zu der beauftragten Tumorentität zu bearbeiten hatten, muss die Darstellung der ein- und ausgeschlossenen Literatur anhand mehrerer Diagramme wie in Abbildung 1 Ihres Berichtes illustriert werden und es müssen separate Auflistungen vorgenommen werden. In der vorliegenden Form ist nicht nachvollziehbar, warum welche Arbeiten für welche Fragestellung akzeptiert bzw. ausgeschlossen wurden. Wir fordern daher konkret, dass eine Aufstellung nach den Themenbereichen erfolgt, die im Berichtsplan genannt sind.

2.2 Inhaltliche Fehler bei der Zusammenfassung der ausgewählten Literatur

Eine Stichprobe für zwei Anwendungsgebiete ergab, dass die Ergebnisse der klinischen Studien aufgrund fehlenden klinischen Sachverständnisses missverstanden wurden, so dass die „Evidenzsynthesen“ auf inkorrekten Daten basieren.

Das erste Beispiel ist die diagnostische Genauigkeit der PET bzw. PET/CT beim N-Staging von Ösophaguskarzinomen. Die überwiegende Zahl der in Deutschland diagnostizierten Ösophaguskarzinome wird im Stadium N1 festgestellt. Deshalb ist die Feststellung eines N1-Stadiums keine Fragestellung von hoher klinischer Relevanz. Wichtig ist dagegen der Nachweis von distanten Lymphknotenmetastasen, die das weitere therapeutische Vorgehen beeinflussen. Die Differenzierung der Stadien N1 und M1a ist dabei entscheidend. Die verschiedenen Arbeiten müssen auf die Aufteilung in zervikale, thorakale und abdominale Lymphknotenstationen überprüft werden, wie sie bei Kato et al. 2008 zu finden ist (Warum gerade diese Arbeit wegen 30 % fehlender Daten ausgeschlossen wurde, konnten wir nicht nachvollziehen, so dass hierzu eine ergänzende Klärung zur Schaffung von Transparenz nötig ist). In der Arbeit von Choi et al. (2000) ist die Separierung in N- und M1a möglich, da die Rohdaten der Patienten in der Vollpublikation patientenbezogen dokumentiert sind (Table 2 in der Arbeit).

In Ihrer Referenz 81 (Flamen et al. 2000) wurden distante LN-Metastasen separat ausgewertet von lokalen LK; auch Stage IV wurde separat ausgewertet. Sie kommentieren aber nur das N-Staging (Tabelle 20; Seite 67), weil angeblich die Referenztests radiologische Verfahren beinhalten. Diese Bewertung ist aber für das M1a Stadium nicht korrekt. Ebenso ist in Ihrer Referenz 82 (Lerut et al. 2000) die Datenbasis des M1a-Staging erkennbar, wird aber nicht erwähnt.

Ein weiteres Beispiel für Verständnisprobleme und inkorrekte Darstellungen im Vorbericht ist das Re-Staging von Ösophaguskarzinomen. Hier zeigt der Bericht mehrere Fehler. Zum einen wird das Restaging auf den Nachweis des lokalen Tumorsprechens reduziert. Ebenso wichtig ist der Nachweis oder Ausschluss von unter der Therapie aufgetretenen Fernmetastasen (siehe Bruzzi et al. Cancer 2007). In dem Bericht wird global vom „histologischen Ansprechen“ als Referenzstandard gesprochen. Leider ist das „histologische Ansprechen“ ein Sammelbegriff, der von unterschiedlichen Autoren gänzlich anders definiert wird.

Diese Problematik wird bei den in Abbildung 6 zusammengefassten Studienergebnissen deutlich. Bei Bruecher et al. und Wieder et al. wurde histologisches Ansprechen als ein Mandard-Regressionsgrad von 1 und 2 definiert, bei Flamen et al. als eine Abnahme des T-Stadiums von der prätherapeutischen Bildgebung, bei Higuchi als Grad 2 oder 3 Ansprechen nach der Klassifikation der Japanischen Gesellschaft für Erkrankungen des Ösophagus, bei Kroep et al. und Roedl et al. ein Anteil von weniger als 10% vitaler Tumorzellen, bei Song et al. als keine oder mikroskopische Tumorreste und bei Cerfolio et al. kein Nachweis von Tumorgewebe. Aufgrund dieser deutlich unterschiedlichen Definitionen des Referenzstandards ist der Wert der Zusammenstellung der Sensitivität und Spezifität dieser Studien in dem „Forrest Plot“ in Abbildung 6 nicht zu erkennen.

In diesem Zusammenhang muss klargestellt werden, dass der „Goldstandard Histopathologie“ (6.1.3; S. 151) und seine klinische Relevanz bei der Beurteilung des Ansprechens auf eine Induktionsbehandlung fraglich sind. Sie schreiben hierzu: „[...] kann das Ergebnis der Histologie nach Resektion als Goldstandard für die Differenzierung zwischen Patienten, die von einer NAC(R) nicht profitieren bzw. profitieren können, betrachtet werden“. Dieser Standpunkt ist derzeit sehr umstritten und sollte hinterfragt werden (siehe u. a. Javeri et al. Cancer 2009).

Unabhängig von diesen inhaltlichen Verständnisproblemen wurden Daten offensichtlich inkorrekt wiedergegeben und widersprüchlich beurteilt. So wird die Studie von Ott et al. abgewertet (Tabelle 40), weil das verwendete PET-Gerät nicht angegeben ist. Dieses Gerät wird aber im Text, durch Referenz auf eine vorangegangene Studie, eindeutig identifiziert. Es werden auch fehlende Angaben zur Entscheidungsfindung bemängelt, obwohl die Entscheidungsfindung anhand eines prospektiven, quantitativen Schwellenwertes in dem Artikel beschrieben ist. Es werden auch fehlende Angaben zur Vorbehandlung bemängelt, obwohl aus dem Artikel klar hervorgeht, dass es sich um eine neoadjuvante Therapie von neu diagnostizierten Tumoren handelt. Auf Seite 120 wird kritisiert, dass sich die Patienten in einem unterschiedlichen Krankheitsstadium befanden. Tatsächlich wurden in die Studie nur Patienten im Stadium III eingeschlossen. Schließlich wird bemängelt, dass die Art der Chemotherapie nicht in der multifaktoriellen Analyse berücksichtigt wurde, was nicht korrekt ist.

Aufgrund der Kürze der zur Verfügung stehenden Zeit konnten wir nur in Stichproben das korrekte Exzerpieren von klinischen Studien überprüfen. Dennoch können wir nur schlussfolgern, dass die Datenerhebung von Untersuchern durchgeführt wurde, die nicht mit der Literatur zur PET bzw. PET/CT und zum Ösophaguskarzinom vertraut sind. Die Ergebnisse der Datenerhebung wurden offenbar auch nicht von klinischen Experten mit medizinischem Sachverstand in der Versorgung des Ösophagus-Karzinom adäquat durchgesehen und korrigiert.

2.3. Methodische Probleme bei der Beurteilung der ausgewählten Literatur

Selbst bei einer Unterstellung, dass diese Fehler bei der Auswahl der publizierten Literatur und der Zusammenstellung der darin beschriebenen Daten nur unwesentliche Aspekte betreffen, bleibt die formale Analyse aus folgenden Gründen mangelhaft.

Die Beurteilung des Verzerrungspotentials nach den QUADAS-Kriterien ist für Studien zum M-Staging problematisch, da eine große Zahl der in QUADAS postulierten Qualitätskriterien notwendigerweise verletzt sind (Weber WA, 2011). Im Vorbericht wird dies für das QUADAS-Item 6 (differentielle Verifikation) eingeräumt (Seite 82, Absatz 1). Es bleibt aber offen, warum dieses Item trotzdem bewertet wurde. Schwerwiegender ist, dass (aus ähnlichen Gründen) bei Studien zum M-Staging mit bildgebenden Verfahren auch die QUADAS Items 3,4,5 und 7 verletzt sind (Weber WA, 2011). Mitarbeiter des IQWiG haben zu der Kritik an der Anwendung von QUADAS beim Ganzkörperstaging Stellung genommen (letter to the editor, in press). Die Kritik an den Items 3-7 wurde dabei nicht beanstandet. Dennoch wurden alle diese Items für den vorliegenden Bericht angewandt. Weil die Items 3-7 von QUADAS bei strikter Auslegung immer verletzt sind, ergibt sich eine hohe Subjektivität bei der Beurteilung, ob es sich um eine für den Zweck einer Studie „relevante“ oder „nicht relevante“ Verletzung handelt.

Auch wenn man diese grundsätzliche Problematik ignoriert, bleiben die Gründe für eine Klassifizierung des Verzerrungspotentials von Studien nicht objektiv nachvollziehbar. Dazu nur einige Beispiele: In Tabelle 18 wird das Verzerrungspotential der Studie von Schreus et al. als niedrig und das der Studie von Buchmann et al. als hoch klassifiziert. Bei beiden Studien ist ein Kriterium als nicht erfüllt, bei Buchmann et al. 2 Kriterien als fraglich und bei Schreus et al. 3 als fraglich angesehen. Nach dem Studium der Tabelle müsste die Klassifikation also umgekehrt sein?

Ebenso wenig nachvollziehbar ist die Verzerrungsargumentation der Studie Ott et al. (5.4.3.2): hier wurden „zu Studienbeginn unterschiedliche Krankheitsstadien“ bemängelt, sowie dass „potenziell relevante Faktoren, wie Alter und Behandlung, in der multifaktoriellen Analyse nicht berücksichtigt wurden“; unter 6.1.3. (S. 152) wird argumentiert, dass Ott et al. ein sehr homogenes Patienten- und Therapiespektrum untersuchten. Außerdem wird bei Chung et al. kritisiert „Die Anzahl der in der multifaktoriellen Analyse eingeschlossenen Faktoren (7) scheint hinsichtlich der Gesamtzahl der eingeschlossenen Patienten (100) zu hoch zu sein.“ (5.3.3.2). Eine einheitliche Linie ist hier nicht zu erkennen.

Ebenso problematisch ist die Beurteilung der Übertragbarkeit der Untersuchungsergebnisse auf die Patientenversorgung in Deutschland. So wird z. B. für die Studie von Higuchi die Übertragbarkeit als gegeben angesehen (Tabelle 37). Dies verwundert sehr, da die Studie an asiatischen Patienten durchgeführt wurde (Unterschiede in der Tumorbilogie von Ösophaguskarzinomen bei asiatischen und europäischen Patienten sind gut dokumentiert, siehe z. B. Rice et al. 2009), als Referenztest in der Studie eine in Europa nicht eingesetzte Klassifikation des histologischen Ansprechens verwendet wurde und die Untersuchungen mit einem PET-Gerät durchgeführt wurden, das in Europa nie verfügbar war. In dem Paper werden keine Angaben gemacht, wie viele Untersucher die Auswertung durchgeführt haben und wie diese qualifiziert waren. Es fehlen auch Angaben zur

Verblindung der Untersucher. Nach unserer Auffassung sollten somit alle Kriterien nach Anhang D4 gegeben sein, dass die Studie *nicht* auf den deutschen Versorgungsalltag übertragbar ist.

Selbst wenn man die Beurteilung des Verzerrungspotentials und die Beurteilung der Übertragbarkeit der Ergebnisse als korrekt akzeptiert, ergibt sich ein unlösbarer Widerspruch zwischen den beiden Beurteilungskriterien. Während zur Verhinderung eines Bias verlangt wird, verblindete Studien durchzuführen, führt diese Verblindung zu einer negativen Beurteilung der Studie im Hinblick auf die Übertragbarkeit auf den Versorgungsalltag. Es wird dadurch unklar, welches Studiendesign vom IQWiG als qualitativ hochwertig akzeptiert würde.

Schlussfolgerung

Der Vorbericht weist schwere inhaltliche und methodische Mängel auf und ist unserer Meinung nach als Grundlage für die Beratung und eine Beschlussfassung durch den G-BA ungeeignet. Die Grundprobleme des Berichtes sind erstens die nicht nachvollziehbare Auswahl der Primärliteratur und der Evidenzsynthesen, zweitens die wenig sorgfältige Exzerpierung und Analyse der ausgewählten Literatur sowie drittens die Beurteilung der Literatur anhand von formalen Kriterien ohne ausreichenden klinischen Sachverstand. Der Bericht kommt deswegen zu Schlussfolgerungen, die konträr zum europäischen, australischen und amerikanischen Standard sind.

Die geforderten randomisierten Studien zum Nachweis eines patientenrelevanten Nutzens sind angesichts der vorliegenden Daten zum Tumorstaging ineffizient und würden zu einer unnötigen Verteuerung der PET- bzw. PET/CT-Diagnostik führen. Angesichts der vorliegenden internationalen Empfehlungen und der klinischen Praxis in der EU, Australien und den USA erscheint es auch fraglich, ob derartige randomisierte Studien ethisch gerechtfertigt sind.

Wir stimmen mit dem IQWiG überein, dass angesichts der begrenzten Ressourcen der Einfluss der PET bzw. PET/CT auf die Ergebnisse der Behandlung des Ösophaguskarzinoms weiter untersucht werden sollte, idealerweise unter Berücksichtigung gesundheitsökonomischer Gesichtspunkte. Dies kann jedoch nicht in klinischen Studien von Universitäten geleistet werden, da hierzu die gesamte Behandlungskette aus primärer Diagnostik und Behandlung, Nachsorge und Therapie bei Rezidiven sowie die sich daraus ergebenden Kosten erfasst werden muss. Derartige Studien scheinen dagegen ideal für die Versorgungsforschung im Rahmen der neuen Erprobungsregelung nach §137e SGB V im Sinne einer „coverage with evidence development“ geeignet. In diesem Rahmen kann nicht nur der Einfluss des Tumorstagings auf die Qualität und Wirtschaftlichkeit der Behandlung von Patienten mit Ösophaguskarzinom untersucht werden, sondern auch die Individualisierung der Behandlung durch eine frühzeitige Beurteilung des Tumoransprechens.

Mit freundlichen Grüßen



Prof. Dr. med. Jörg Kotzerke
Präsident der Deutschen Gesellschaft
für Nuklearmedizin (DGN)



Prof. Dr. med. Dirk Hellwig
Vorsitzender des DGN-Ausschusses
Positronen-Emissions-Tomographie (PET)



Prof. Dr. med. Wolfgang Weber
Abteilung Nuklearmedizin,
Universitätsklinikum Freiburg



Prof. Dr. med. Matthias Schmidt
Klinik und Poliklinik für Nuklearmedizin,
Universitätsklinikum Köln



PD Dr. med. Ken Herrmann
Department of Molecular and Medical Pharmacology,
Ahmanson Translational Imaging Division,
David Geffen School of Medicine at the University of California Los Angeles

Literatur:

1. Barber TW, Duong CP, Leong T, Bressel M, Drummond EG, Hicks RJ. ^{18}F -FDG PET/CT Has a High Impact on Patient Management and Provides Powerful Prognostic Stratification in the Primary Staging of Esophageal Cancer: A Prospective Study with Mature Survival Data. *J Nucl Med.* 2012 Jun;53(6):864-71. Epub 2012 May 11.
2. Bruzzi JF, Swisher SG, Truong MT, Munden RF, Hofstetter WL, Macapinlac HA, Correa AM, Mawlawi O, Ajani JA, Komaki RR, Fukami N, Erasmus JJ. Detection of Interval Distant Metastases: Clinical Utility of Integrated CT-PET Imaging in Patients With Esophageal Carcinoma After Neoadjuvant Therapy. *Cancer.* 2007 Jan 1;109(1):125-34.
3. Choi JY, Lee KH, Shim YM, Lee KS, Kim JJ, Kim SE, Kim BT. Improved Detection of Individual Nodal Involvement in Squamous Cell Carcinoma of the Esophagus by FDG PET. *J Nucl Med.* 2000 May;41(5):808-815.
4. Esophageal and esophagogastric junction. In: Edge SB, Byrd DR, Compton CC, et al., eds.: *AJCC Cancer Staging Manual.* 7th ed. New York, NY: Springer, 2010, pp 103-15. [Diese Quelle ist aus Gründen des Copyrights nicht als Anlage beigefügt.]
5. Fletcher JW, Djulbegovic B, Soares HP, Siegel BA, Lowe VJ, Lyman GH, Coleman RE, Wahl R, Paschold JC, Avril N, Einhorn LH, Suh WW, Samson D, Delbeke D, Gorman M, Shields AF. Recommendations on the Use of ^{18}F -FDG PET in Oncology. *J Nucl Med.* 2008 Mar;49(3):480-508.
6. Gillies RS, Middleton MR, Han C, Marshall REK, Maynard ND, Bradley KM, Gleeson FV. Role of positron emission tomography-computed tomography in predicting survival after neoadjuvant chemotherapy and surgery for oesophageal adenocarcinoma. *Br J Surg.* 2012 Feb;99(2):239-45. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22329010>
7. Gillies RS, Middleton MR, Blesing C, Patel K, Warner N, Marshall REK, Maynard ND, Bradley KM, Gleeson FV. Metabolic response at repeat PET/CT predicts pathological response to neoadjuvant chemotherapy in oesophageal cancer. *Eur Radiol.* 2012 May 5.
8. Javeri H, Xiao L, Rohren E, Lee JH, Liao Z, Hofstetter W, Maru D, Bhutani MS, Swisher SG, Macapinlac H, Wang X, Ajani JA. The Higher the Decrease in the Standardized Uptake Value of Positron Emission Tomography After Chemoradiation, the Better the Survival of Patients With Gastroesophageal Adenocarcinoma. *Cancer.* 2009 Nov 15;115(22):5184-92.
9. Jayachandran P, Pai RK, Quon A, Graves E, Krakow TE, La T, Loo BW Jr, Koong AC, Chang DT. Postchemoradiotherapy Positron Emission Tomography Predicts Pathologic Response and Survival in Patients with Esophageal Cancer. *Int J Radiat Oncol Biol Phys.* 2012 Feb 28.
10. Kato H, Kimura H, Nakajima M, Sakai M, Sano A, Tanaka N, Inose T, Faried A, Saito K, Ieta K, Sohda M, Fukai Y, Miyazaki T, Masuda N, Fukuchi M, Ojima H, Tsukada K, Oriuchi N, Endo K, Kuwano H. The additional value of integrated PET/CT over PET in initial lymph node staging of esophageal cancer. *Oncol Rep.* 2008;20(4):857-862.
11. Krause BJ, Herrmann K, Wieder H, zum Büschenfelde CM. ^{18}F -FDG PET and ^{18}F -FDG PET/CT for Assessing Response to Therapy in Esophageal Cancer. *J Nucl Med.* 2009 May;50 Suppl 1:89S-96S
12. Liberale G, Van Laethem JL, Gay F, Goldman S, Nagy N, Coppens E, Gelin M, Nakadi IE. The role of PET scan in the preoperative management of oesophageal cancer. *Eur J Surg Oncol.* 2004;30(9):942-947.
13. Lordick F, Ott K, Krause BJ, Weber WA, Becker K, Stein HJ, Lorenzen S, Schuster T, Wieder H, Herrmann K, Bredenkaamp R, Höfler H, Fink U, Peschel C, Schwaiger M, Siewert JR. PET to assess early metabolic response and to guide treatment of adenocarcinoma of the oesophagogastric junction: the MUNICON phase II trial. *Lancet Oncol.* 2007 Sep;8(9):797-805
14. Lorenzen S, von Gall C, Stange A, Haag GM, Weitz J, Haberkorn U, Lordick F, Weichert W, Abel U, Debus J, Jäger D, Mütter MW. Sequential FDG-PET and induction chemotherapy in locally advanced adenocarcinoma of the Oesophago-gastric junction (AEG): The Heidelberg Imaging program in Cancer of the oesophago-gastric junction during Neoadjuvant treatment: HICON trial. *BMC Cancer.* 2011 Jun 24;11:266.
15. Medical Services Advisory Committee. Positron emission tomography for oesophageal and gastric cancer: Assessment Report; MSAC reference 35b(i). MSAC; 2008 June.

16. Omloo JMT, van Heijl M, Hoekstra OS, van Berge Henegouwen MI, van Lanschot JJ, Sloof GW. FDG-PET Parameters as Prognostic Factor in Esophageal Cancer Patients: A review. *Ann Surg Oncol*. 2011 Nov;18(12):3338-52.
17. Ott K, Herrmann K, Krause BJ, Lordick F. The Value of PET Imaging in Patients with Localized Gastroesophageal Cancer. *Gastrointest Cancer Res*. 2008 Nov;2(6):287-94.
18. Rice TW, Rusch VW, Apperson-Hansen C, Allen MS, Chen LQ, Hunter JG, Kesler KA, Law S, Lerut TE, Reed CE, Salo JA, Scott WJ, Swisher SG, Watson TJ, Blackstone EH. Worldwide esophageal cancer collaboration. *Dis Esophagus*. 2009;22(1):1-8.
19. Schreurs LMA, Janssens ACJW, Groen H, Fockens P, van Dullemen HM, van Berge Henegouwen MI, Sloof GW, Pruijm J, van Lanschot JJB, Steyerberg EW, Plukker JT. Value of EUS in Determining Curative Resectability in Reference to CT and FDG-PET: The Optimal Sequence in Preoperative Staging of Esophageal Cancer? *Ann Surg Oncol*. 2011 May 6. [Epub ahead of print] PubMed PMID: 21547703.
20. Vach W, Høilund-Carlsen PF, Gerke O, Weber WA. Generating Evidence for Clinical Benefit of PET/CT in Diagnosing Cancer Patients. *J Nucl Med* 2011 52:775-85S.
21. van Hagen P, Hulshof MCCM, van Lanschot JJB, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BPL, Richel DJ, Nieuwenhuijzen GAP, Hospers GAP, Bonenkamp JJ, Cuesta MA, Blaisse RJB, Busch ORC, ten Kate FJW, Creemers GJ, Punt CJA, Plukker JTM, Verheul HM, Spillenaar Bilgen EJ, van Dekken H, van der Slangen MJC, Rozema T, Biermann K, Beukema JC, Piet AHM, van Rij CM, Reinders JG, Tilanus HW, van der Gaast A; CROSS Group. Preoperative Chemoradiotherapy for Esophageal or Junctional Cancer. *N Engl J Med*. 2012 May 31;366(22):2074-84.
22. Vyas S, Markar SR, Iordanidou L, Read S, Stoker D, Hashemi M, Mitchell I, Winslet M, Bomanji J. The Role of Integrated F-18-FDG-PET Scanning in the Detection of M1 Disease in Oesophageal Adenocarcinoma and Impact on Clinical Management. *J Gastrointest Surg*. 2011 Dec;15(12):2127-35.
23. Weber WA. Is There Evidence for Evidence-Based Medical Imaging? *J Nucl Med* 2011 52:745-76S.
24. Yanagawa M, Tatsumi M, Miyata H, Morii E, Tomiyama N, Watabe T, Isohashi K, Kato H, Shimosegawa E, Yamasaki M, Mori M, Doki Y, Hatazawa J. Evaluation of Response to Neoadjuvant Chemotherapy for Esophageal Cancer: PET Response Criteria in Solid Tumors Versus Response Evaluation Criteria in Solid Tumors. *J Nucl Med*. 2012 Jun;53(6):872-80. Epub 2012 May 11.
25. Zum Büschenfelde CM, Herrmann K, Schuster T, Geinitz H, Langer R, Becker K, Ott K, Ebert M, Zimmermann F, Friess H, Schwaiger M, Peschel C, Lordick F, Krause BJ. ¹⁸F-FDG PET-Guided Salvage Neoadjuvant Radiochemotherapy of Adenocarcinoma of the Esophagogastric Junction: the MUNICON II Trial. *J Nucl Med*. 2011 Aug;52(8):1189-96.

Anlagen:

- Literaturverzeichnis
- Zitierte Literatur (eingereicht in elektronischer Form mit vier E-Mails vom 21. Juni 2012)
- Formblatt zur schriftlichen Stellungnahme zum Vorbericht
- Formblatt zur Offenlegung potenzieller Interessenkonflikte: Prof. Dr. med. J. Kotzerke
- Formblatt zur Offenlegung potenzieller Interessenkonflikte: Prof. Dr. med. D. Hellwig
- Formblatt zur Offenlegung potenzieller Interessenkonflikte: Prof. Dr. med. W. Weber
- Formblatt zur Offenlegung potenzieller Interessenkonflikte: Prof. Dr. med. M. Schmidt
- Formblatt zur Offenlegung potenzieller Interessenkonflikte: PD Dr. med. K. Herrmann

A 1.2 Deutsche Krankenhausgesellschaft e. V.

Autoren:

Acikgöz, Özlem

Dippmann, Anja

Rüther, Klaus

Schlottmann, Nicole



DEUTSCHE
KRANKENHAUS
GESELLSCHAFT

Institut für Qualität und
Wirtschaftlichkeit im Gesundheitswesen
- Stellungnahme zum Vorbericht <D06-01H>
Prof. Dr. med. Jürgen Windeler
Im Mediapark 8
50670 Köln

Bundesverband
der Krankenhausträger
in der Bundesrepublik
Deutschland

EINGEGANGEN

22. Juni 2012

-vorab per E-Mail -

Unser Zeichen
Dez. V/AC

Telefon
Durchwahl +49 (0) 30 3 98 01- [REDACTED]

Telefax
+49 (0) 30 3 98 01- [REDACTED]

Datum
21.06.2012

Stellungnahme der Deutschen Krankenhausgesellschaft zum Vorbericht D06-01H „Positronenemissionstomographie (PET) und PET/CT bei Ösophaguskarzinom“

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

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

Der Gemeinsame 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- und PET/CT-Diagnostik bei Ösophaguskarzinomen 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 Vorbericht des Instituts wie folgt Stellung:

Postfach 120555 · D-10595 Berlin
Wegelystraße 3 · D-10623 Berlin
Telefon +49 (0) 30 3 98 01-0
Telefax +49 (0) 30 3 98 01-3000
dkg@mail@dkgev.de · www.dkgev.de

Rue d'Arlon 50
B-1000 Brüssel
Telefon +32 (0) 2 282- [REDACTED]
Telefax +32 (0) 2 282- [REDACTED]
europa@dkgev.de

Commerzbank AG
Konto-Nummer [REDACTED] (BLZ [REDACTED])
BIC (SWIFT-Code): [REDACTED]
IBAN: [REDACTED]
USt-IdNr. DE [REDACTED]

1. Die letzte Suche nach Primärliteratur in bibliografischen Datenbanken fand am 16.02.2010 statt. Die Veröffentlichung des Vorberichts erfolgt mehr als 2 Jahre später. Auch wenn es sich um ein aufwändiges Auswertungsverfahren handelt, ist dieser lange Zeitraum unverständlich. Die hier vorliegende Stellungnahme erfolgt nunmehr auf der Grundlage einer Evidenzlage, die nicht mehr aktuell ist. Für die Stellungnehmenden eröffnet sich vor diesem Hintergrund die Problematik, dass sie erneut weitere Erkenntnisse ermitteln müssen, um die aktuelle Evidenzlage korrekt darzustellen. Wir bitten Sie daher, diese u.E. lange Verzögerung zu erläutern.
2. Wie Sie richtig einleitend feststellen, finden sich im Bereich des Ösophagus sowohl Plattenepithelkarzinome als auch Adenokarzinome. Allerdings befinden sich die Plattenepithelkarzinome überwiegend im mittleren und oberen Drittel; nicht im mittleren und unteren Drittel (s. S. 1).
3. Der vorliegende Bericht erhebt den Anspruch, auch die ösophagealen Adenokarzinome zum Gegenstand zu haben. Histologisch betrachtet handelt es sich hierbei meist um Karzinome, die Schleimhaut- und Drüsenstrukturen aufweisen und infolge einer Barrett-Metaplasie (nicht Barrett-Ulkus, s. S. 2) auf dem Boden einer Refluxkrankheit entstehen. Die mit zunehmender Häufigkeit auftretenden, zumeist distal gelegenen Adenokarzinome werden seit mehreren Jahren als eigene Entität, nämlich dem "Adenokarzinom des ösophagogastralen Übergangs" (AEG) zusammengefasst und anhand der AEG-Klassifikation nach Siewert eingeteilt. So wird z. B. das Adenokarzinom, das sein Tumorzentrum und seine Tumormasse im distalen Ösophagus aufweist, als Barrett-Karzinom bezeichnet und als AEG I klassifiziert. Der Bericht erwähnt im Hintergrundkapitel das AEG nicht, obgleich es Teil seines Berichtsgegenstands ausmacht. Wir empfehlen daher, das entsprechende Kapitel mit klinisch-fachkundiger Expertise zu überarbeiten.
4. Darüber hinaus fällt auf, dass der Begriff des „Karzinoms des gastroösophagealen Übergangs“ keinen Eingang in Ihre Literatursuchstrategie gefunden hat. Damit kann nicht sicher ausgeschlossen werden, dass relevante Literatur nicht in Betracht gezogen worden ist. Aus diesem Grund erscheint eine Anpassung der Suchstrategie unumgänglich. In diesem Zusammenhang sei darauf verwiesen, dass es sinnvoll wäre, die Literatursuchstrategie bereits im Berichtsplan darzulegen und damit Stellungnehmenden bereits in dieser Phase die Möglichkeit zu geben, diese zu kommentieren.
5. Im Kapitel „Hintergrund“ nutzt der Bericht größtenteils Quellen, die nicht den aktuellen Stand widerspiegeln. So ist z. B. die Berichterstattung des Robert-Koch-Instituts und der Gesellschaft der epidemiologischen Krebsregister in Deutschland bereits in der 8. Ausgabe verfügbar (1). Der Vorbericht bezieht sich auf die 6. Ausgabe und stellt Daten aus 2003/2004 dar. Demzufolge differieren die von Ihnen genannten Zahlen und Häufigkeiten mit den aktuell verfügbaren Daten. Bei den Klassifikationen zitieren Sie zwar den aktuelleren Band 7 (Ihre Referenz Nr. 13), nutzen in der Darstellung aber die veraltete TNM- und UICC-Stadieneinteilungen gemäß Band 6 aus 2003.
6. Zum Thema der derzeit üblichen therapeutischen Verfahren findet sich in der Cochrane Database eine Metaanalyse zur präoperativen Chemotherapie aus 2010 (2), die in Ihrem

Bericht keine Erwähnung findet. Wir bitten Sie, das Kapitel zu überarbeiten und die Aktualität zu berücksichtigen.

7. Die Darstellung der verfügbaren Leitlinien und Empfehlungen der wissenschaftlichen Fachgesellschaften weist große Mängel auf. Sie beziehen sich größtenteils auf Quellen, die nicht mehr aktuell oder verfügbar sind und vernachlässigen relevante Leitlinien.
8. So findet die deutsche S3-Leitlinie, die – wenn vorhanden - in jedem IQWiG-Bericht Erwähnung finden sollte, keinen Eingang in den Bericht. Darüber hinaus sind mehrere Leitlinien verfügbar, die die Karzinome des Ösophagus und des gastroösophagealen Übergangs zum Gegenstand haben, in Ihrem Bericht aber keine Erwähnung finden.
9. Laut deutscher S3-Leitlinie kann die PET für Patienten mit Adenokarzinomen des gastroösophagealen Übergangs und kurativer Therapieoption nach konventionellem Staging mit CT/Endo-Ultraschall für das Staging von loko-regionär und nicht loko-regionär gelegenen Lymphknoten und zum Nachweis / Ausschluss anderer Fernmetastasen in Betracht gezogen werden (3). Die von Ihnen genannte SIGN-Leitlinie aus dem Jahre 2006 wird zurzeit einem Review unterzogen (4). Darin findet sich der Hinweis, dass die Empfehlung, PET nicht routinemäßig beim Staging einzusetzen, aufgrund der höheren Sensitivität bei der Detektion von Fernmetastasen möglicherweise revidiert werden wird. Die „European Organisation of Research and Treatment of Cancer“ (EORTC) sieht die FDG-PET als hilfreich für die Bewertung des Therapieansprechens bei Karzinomen des gastroösophagealen Übergangs (AEG I und II) an (5). Die kanadische Leitlinie „Sask Cancer Agency Esophageal Cancer Guidelines“ von 2011 empfiehlt die PET als mögliche Option bei T- und M-Staging (6).
10. Die Empfehlungen des Freiburger Tumorzentrums, auf die Sie an mehreren Stellen eingehen (Ihre Referenz Nr. 1) sind über zehn Jahre alt (von 2001) und auf der von Ihnen genannten Website nicht mehr zugänglich. Die von Ihnen genannten ESMO-Guidelines (Ihre Referenz Nr. 14) existieren bereits in einer aktuelleren, überarbeiteten Version (7).
11. Da sich im Bericht keine Information findet, ob eine Aktualisierung der Leitlinienkapitel vorgesehen ist, bitten wir Sie, Ihre Quellenangaben zu überprüfen und insbesondere die genannten aktuellen Leitlinien in den Bericht miteinzubeziehen.
12. Die mit den Ein- und Ausschlusskriterien erfolgte Limitierung der einzubeziehenden Literatur auf Studien der Evidenzstufen I und II erachten wir, wie bereits mehrfach von uns in Stellungnahmen und in der Kritik Ihrer Methodenpapiere angemerkt, als nicht konform mit den Vorgaben der Verfahrensordnung des G-BA. Die von Ihnen so nur eingeschränkt vorgenommene Berücksichtigung der zum Zeitpunkt der Literaturrecherche vorhandenen medizinisch-wissenschaftlichen Evidenz führt dazu, dass das IQWiG im Vorbericht keine einzige Studie zur Frage des patientenrelevanten Nutzens der PET bei Ösophaguskarzinom berücksichtigt hat. Laut Verfahrensordnung des G-BA gilt für diagnostische Methoden, dass Erkenntnissicherheit insbesondere durch Unterlagen der Evidenzstufe I entsprechend der Verfahrensordnung gewonnen werden sollte. Sollte es begründet unangemessen oder unmöglich sein, Studien der Evidenzstufe I zu fordern oder durchzuführen, so können auch qualitativ angemessene Unterlagen niedrigerer Evidenzstufen her-

angezogen werden. Die aktuelle Verfahrensordnung des G-BA (in Kraft seit 01.02.2012) besagt auch, dass es u. a. bei seltenen Erkrankungen unmöglich oder unangemessen sein kann, Studien dieser Evidenzstufe durchzuführen oder zu fordern. Da das Ösophaguskarzinom nach EU-Definition, die auch in der Verfahrensordnung des G-BA verankert ist (3. Kapitel, § 5), mit einer standardisierten Erkrankungsrate von unter 1 Fall pro 10.000 Einwohner (in Deutschland) eine seltene Erkrankung ist, bitten wir zu erläutern, warum trotz dieser G-BA-Vorgabe der Bericht Studien der Evidenzklasse I und II fordert. Wir weisen erneut darauf hin, dass der G-BA für seine Entscheidungen die „Darstellung und Bewertung des aktuellen medizinischen Wissensstandes“, wie im §139a Absatz 3 Ziffer 1 SGB V ausdrücklich benannt, benötigt. Vor diesem Hintergrund wird der G-BA ansonsten erneut weitere Erkenntnisse ermitteln müssen.

13. Auf S. 17 des Vorberichts weisen Sie darauf hin, dass „eine Änderung im Patientenmanagement alleine (ohne Verknüpfung zu den oben genannten Zielgrößen) ... keine patientenrelevante Zielgröße im Sinne dieses Berichts dar[stellte] und ... daher nur ergänzend betrachtet [wurde]“. In diesem Zusammenhang muss darauf hingewiesen werden, dass eine in einer Studie ggf. dargestellte Änderung des Patientenmanagements dazu führen muss, dass nach Studien gesucht wird, die einen patientenrelevanten Nutzen der beschriebenen Änderung des Managements zum Gegenstand hatten, nicht jedoch die Anwendung der PET. Sollte nicht so vorgegangen worden sein, wäre auch diese Information wichtig und daher darzustellen, da in diesem Fall der G-BA selbst solche Studien identifizieren und auswerten müsste.
14. In Kapitel 4.1.2.1. sollte ergänzt werden, dass eine Therapie, die zwischen Prüftechnologie und Vergleichstechnologie einerseits und dem Referenztest andererseits bei allen Patienten durchgeführt wurde, nicht als Verzerrungspotenzial angesehen wurde, da eine differenzielle Wirkung dieser Therapie auf die der Prüf- bzw. Vergleichstechnologie unterzogenen Patientengruppen nicht angenommen werden kann. Sollte diese Aussage nicht zutreffen, müssten ggf. die aus diesem Grund ausgeschlossenen Diagnose- und Prognose-Studien neu bewertet und in der Regel eingeschlossen werden.
15. In Kapitel 4.5.2.2 sollte ergänzt werden, dass im Falle einer Überlappung der Konfidenzregionen andere statistische Testverfahren herangezogen wurden (z.B. McNemar-Test), die der jeweiligen Publikation entweder direkt oder indirekt entnommen werden können.
16. Übertragbarkeit auf den deutschen Versorgungskontext (z.B. Kap. 5.4.2.3): Es werden u.a. eine verblindete Bewertung oder eine Bildinterpretation durch 2 Bewerter als Anhaltspunkte für die Nichtübertragbarkeit auf den deutschen Versorgungskontext angesehen. Diese Sichtweise sollte auf ihre Sinnhaftigkeit überprüft werden; bei konsequenter Anwendung würde man z.B. zu dem Schluss kommen müssen, dass kein RCT auf den deutschen Versorgungskontext übertragbar ist.
17. Schon in früheren Berichten des Instituts wurde deutlich, dass insbesondere im Zusammenhang mit der diagnostischen Methode PET bzw. PET/CT die Forderung nach randomisierten, simultanen Vergleichen zwischen einem Vorgehen mit und ohne PET in vielen Konstellationen nicht sachgerecht ist und von verantwortlichen Fachärzten nicht durchgeführt werden kann (z. B. prä-OP-PET als Zusatzinformation erst nach Ausschöpfung der

Alternativtechnologien). Da der vorliegende IQWiG-Bericht dies aber einfordert und dementsprechend gemäß seinen Einschlusskriterien 37 Primärstudien mit dem Ausschlussgrund „Studiendesign entspricht nicht den Einschlusskriterien“ ausschließt, bitten wir diesen klinisch und damit auch methodisch relevanten Aspekt unter Hinzuziehung entsprechender Fachexpertise im Bericht zu beleuchten und die Definition der Einschlusskriterien dementsprechend zu erläutern bzw. zu modifizieren.

18. Wie bereits in unserer Stellungnahme zum Berichtsplan angemerkt, erachten wir die Beteiligung von klinischen Sachverständigen in Bezug auf den gesamten Bericht für unerlässlich. Insbesondere die Studienauswahl und Ergebnisinterpretation sollten unter onkologischer, chirurgischer und nuklearmedizinischer Fachbegleitung erfolgen. Aus dem vorliegenden Vorbericht ist leider nicht ersichtlich, ob dies der Fall war. Wir bitten Sie zukünftig bereits im Vorbericht – und nicht erst im Abschlussbericht - Informationen zu beteiligten Sachverständigen darzulegen.
19. Im Bericht findet sich die Information, dass in Hinblick auf patientenrelevante Endpunkte bei Erstellung des Berichtsplans, der für mehrere Entitäten galt, am 04.09.2009 Patientenvertreter konsultiert wurden und dabei kein Patientenvertreter für das Krankheitsbild Ösophaguskarzinom gefunden werden konnte. Leider findet sich kein Hinweis darüber, ob im Zuge der Erstellung des Vorberichts erneut versucht worden ist, themenbezogene Patientenvertreter zu gewinnen. Sollte dies nicht erfolgt sein, bitten wir um eine Erläuterung der Gründe und ggf. die Information, ob Patientenvertreter eine Beteiligung abgelehnt haben.
20. Bei Ihrer Bewertung diagnostischer Tests berufen Sie sich auf die „Mehrzahl der westlichen Gesundheitssysteme“, die „Überlegungen zu den Konsequenzen und dem Nutzen und Schaden diagnostischer Tests für Arzt, Patient und Gesellschaft in Betracht zieht“. Dieser Sichtweise habe sich die Verfahrensordnung des G-BA angeschlossen. Wie bereits mehrfach von uns angemerkt, ist dies nicht der Fall. Der diesbezüglich von Ihnen genannte Paragraph (§ 18) ist in der Verfahrensordnung des G-BA nicht existent. Eine Limitierung der Bewertung auf die von Ihnen genannten Grundlage findet sich in der Verfahrensordnung des G-BA nicht. Darüber hinaus bleibt nach wie vor unklar, was das IQWiG mit dem Ausdruck „westliches Gesundheitssystem“ meint.
21. Wir gehen davon aus, dass sich, auch gemäß der *Methoden 4.0*, im Abschlussbericht eine geschlechtergerechte Sprache wiederfinden wird.

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

Bitte haben Sie Verständnis dafür, dass wir mit Blick auf das geltende Urheberrecht und einer diesbezüglich noch ausstehenden endgültigen rechtlichen Prüfung in unserem Hause zunächst von der Übermittlung der von uns zitierten Volltexte absehen.

Mit freundlichen Grüßen
Der Hauptgeschäftsführer
Im Auftrag



Dr. med. N. Schlottmann
Geschäftsführerin
Dezernat Medizin



Ö. Acikgöz, Ärztin
Referentin
Dezernat Medizin



Prof. Dr. med. K. Rüther
Referent
Dezernat Medizin



Dr. med. A. K. Dippmann
Referentin
Dezernat Medizin

1. Robert-Koch-Institut und Gesellschaft der epidemiologischen Register in Deutschland, Krebs in Deutschland 2007-2008; 8. Ausgabe, 2012; frei verfügbarer online-Zugriff über http://www.gekid.de/Doc/krebs_in_deutschland_2012.pdf
2. Kelly Vogt, Deborah Fenlon, Sarah Rhodes, Richard Malthaner. Preoperative chemotherapy for resectable thoracic esophageal cancer, Published Online: 14 APR 2010; Cochrane Database, frei verfügbarer online-Zugriff über <http://online.library.wiley.com/doi/10.1002/14651858.CD001556.pub2/abstract>
3. S3-Leitlinie „Magenkarzinom“ – Diagnostik und Therapie der Adenokarzinome des Magens und ösophagogastralen Übergangs (AWMF-Regist.-Nr. 032-009-OL, frei verfügbarer online-Zugriff über http://www.dgvs.de/media/Leitlinie_Magenkarzinom_2011.pdf
4. Management of esophageal and gastric cancer; Proposed Review of SIGN guideline consultation summary, 2012; frei verfügbarer online-Zugriff über http://www.sign.ac.uk/pdf/SIGN87_gastric_cancer_review.pdf
5. Lordick F, Ruers T, Aust DE, Collette L, Downey RJ, El Hajjam M, Flamen P, Haustermans K, Ilson D, Julié C, Krause BJ, Newiger H, Ott K, Roth A, Van Cutsem E, Weber WA, Lutz MP; European Organisation of Research and Treatment of Cancer (EORTC) Gastrointestinal Group: Workshop on the role of metabolic imaging in the neoadjuvant treatment of gastrointestinal cancer. Eur J Cancer. 2008 Sep;44(13):1807-19. Epub 2008 Jul 17
6. Provincial Esophageal Cancer and Gastro-esophageal junction Cancer Treatment Guidelines (As per consensus at the Provincial Thoracic Oncology meeting on March 12, 2011) Sask Cancer Agency Esophageal Cancer Guidelines 2011, frei verfügbarer online-Zugriff über <http://www.saskcancer.ca/Esoophageal%20Guidelines>
7. Stahl M, Budach W, Meyer H-J, Cervantes A. Esophageal cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol (2010) 21 (suppl 5): v46-v49, frei verfügbarer online-Zugriff über http://annonc.oxfordjournals.org/content/21/suppl_5/v46.long