

**Biomarkerbasierte Tests zur
Entscheidung für oder gegen
eine adjuvante systemische
Chemotherapie beim
primären Mammakarzinom**

**Dokumentation und Würdigung der
Anhörung zum Vorbericht**

Auftrag: D14-01
Version: 1.0
Stand: 27.10.2016

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

22.04.2014

Interne Auftragsnummer:

D14-01

Anschrift des Herausgebers:

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

Tel.: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

E-Mail: berichte@iqwig.de

Internet: 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)	3
4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll.....	5
4.1 Teilnehmerliste der wissenschaftlichen Erörterung	5
4.2 Tagesordnung der wissenschaftlichen Erörterung	6
4.3 Protokoll der wissenschaftlichen Erörterung.....	6
4.3.1 Begrüßung und Einleitung.....	6
4.3.2 Tagesordnungspunkt 1: Grenzwert für die Vollständigkeit der Daten.....	7
4.3.3 Tagesordnungspunkt 2: Mindestdauer der Nachbeobachtung	22
4.3.4 Tagesordnungspunkt 3: Versorgungsstandard	27
4.3.5 Tagesordnungspunkt 4: Verschiedenes	31
Anhang A – Dokumentation der Stellungnahmen	46

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AGO	Arbeitsgemeinschaft gynäkologischer Onkologen
EBCTCG	Early Breast Cancer Trialists' Collaborative Group
ER	Östrogenrezeptor
ESMO	European Society for Medical Oncology
G-BA	Gemeinsamer Bundesausschuss
HER2	Rezeptor des humanen epidermalen Wachstumsfaktors 2
HTA	Health Technology Assessment
NICE	The National Institute for Health and Care Excellence
IHC	immunhistochemisch
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
PAI 1	Plasminogen-Aktivator-Inhibitor 1
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
uPA	Urokinase-Typ Plasminogen-Aktivator

1 Dokumentation der Anhörung

Am 10.11.2015 wurde der Vorbericht in der Version Nr. 1.0 vom 02.11.2015 veröffentlicht und zur Anhörung gestellt. Bis zum 08.12.2015 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 8 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 18.01.2016 im IQWiG diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

Eine Würdigung der in der Anhörung vorgebrachten Aspekte befindet sich im Kapitel „Kommentar“ 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 „Kommentar“ des Abschlussberichts gewürdigt.

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“ (Version 11/2013). 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
Berufsverband der Frauenärzte e. V.	Marion Kiechle	nein	nein	ja	ja	nein	ja
Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V. & Deutsche Gesellschaft für Senologie e. V.	Marcus Schmidt	nein	ja	ja	ja	ja	nein
	Peter Sinn	nein	ja	ja	ja	ja	nein
Deutsche Gesellschaft für Pathologie e. V. & Bundesverband Deutscher Pathologen e. V.	Manfred Dietel	nein	ja	ja	ja	nein	ja
Genomic Health Deutschland GmbH	Matthias Langenfeld	ja	nein	nein	nein	nein	nein
	Erwin Morawski	ja	nein	nein	nein	nein	ja
	Holger Weidenauer	ja	nein	nein	nein	nein	ja
Mamma Mia! Das Brustkrebsmagazin	Eva Schumacher-Wulf	nein	nein	ja	nein	ja	nein
Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V.	Sandra Janatzek	ja	ja	ja	nein	nein	nein
NanoString Technologies GmbH	Uta Mellert	ja	nein	nein	nein	nein	nein
	Karl Sotlar	nein	ja	ja	ja	ja	nein
Sividon Diagnostics GmbH	Ralf Kronenwett	ja	nein	nein	nein	nein	ja
	Christoph Petry	ja	nein	nein	nein	nein	ja

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

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Kreienberg, Rolf	nein	ja	nein	nein	nein	nein
Rutjes, Anne	nein	ja	nein	nein	ja	nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt (Version 11/2013):

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 pharmazeutischem 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, für die Sie tätig sind, 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 pharmazeutischem 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, für die Sie tätig sind, 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
Breuer, Susanne	IQWiG
Dietel, Manfred	Deutsche Gesellschaft für Pathologie e. V. & Bundesverband Deutscher Pathologen e. V.
Ernst, Stefan	Protokollant Sitzungsprotokollarischer Dienst des Landtags NRW
Hausner, Elke	IQWiG
Janatzek, Sandra	Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e. V.
Kiechle, Marion	Berufsverband der Frauenärzte e. V.
Kreienberg, Rolf	Universitätsklinikum Ulm – Klinik für Frauenheilkunde und Geburtshilfe, Direktor (em.)
Kronenwett, Ralf	Sividon Diagnostics GmbH
Lange, Stefan	IQWiG
Langenfeld, Matthias	Genomic Health Deutschland GmbH
Markes, Martina	IQWiG
Mellert, Uta	NanoString Technologies GmbH
Mensch, Alexander	IQWiG
Morawski, Erwin	Genomic Health Deutschland GmbH
Petry, Christoph	Sividon Diagnostics GmbH
Rutjes, Anne	Institut für Sozial- und Präventivmedizin, Universität Bern
Rüttgers, Daniela	IQWiG
Sauerland, Stefan	IQWiG
Scheibler, Fülöp	IQWiG
Schmidt, Marcus	Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V.
Schumacher-Wulf, Eva	Mamma Mia! Das Brustkrebsmagazin
Sieben, Wiebke	IQWiG
Sinn, Peter	Deutsche Gesellschaft für Senologie e. V.
Sotlar, Karl	NanoString Technologies GmbH
Weidenauer, Holger	Genomic Health Deutschland GmbH
Windeler, Jürgen	IQWiG (Moderation)

4.2 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Grenzwert für die Vollständigkeit der Daten
TOP 2	Mindestdauer der Nachbeobachtung
TOP 3	Versorgungsstandard
TOP 4	Verschiedenes

4.3 Protokoll der wissenschaftlichen Erörterung

Datum: 18.01.2016, 13:00 bis 14:55 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: Meine Damen und Herren! Ich begrüße Sie sehr herzlich zur wissenschaftlichen Erörterung des Vorberichts zu den Biomarkern beim primären Mammakarzinom. Einige von Ihnen sehe ich jedenfalls in diesem Raum zum zweiten Mal; wir haben uns schon beim Berichtsplan ausgetauscht.

Ich muss trotzdem, will trotzdem meine üblichen Vorbemerkungen machen, die damit anfangen, dass die Veranstaltung aufgezeichnet wird, dass wir das Wortprotokoll der Veranstaltung veröffentlichen - wie immer -, dass wir Sie bei der Einladung auf diesen Umstand hingewiesen haben und auch deutlich gemacht haben, dass Sie mit dem Erscheinen heute der Veröffentlichung des Wortprotokolls zustimmen. Damit hängt zusammen - noch einmal als Regieanweisung - dass Sie bitte vor jeder Ihrer Wortmeldungen Ihren Namen sagen, damit das Protokoll, der Stenograf, das Tonband alle Äußerungen richtig zuordnen können. Damit hängt auch zusammen - das ist vielleicht ein bisschen schwierig, aber vielleicht können Sie mal schauen, wie weit das geht -, dass ich Ihre Namensschilder möglichst einigermaßen von hier sehen kann; da stehen ein paar Flaschen im Weg. Dann habe ich nämlich die Chance, Sie auch anzusprechen. Das wären die Dinge, die zur reinen Technik zu sagen wären.

Ich habe auch in den ersten Veranstaltungen bereits gesagt, dass diese Erörterung nicht wie andere Anhörungen, die Sie vielleicht kennen oder schon mitgemacht haben, dem erneuten Verlesen von schriftlichen Stellungnahmen dient. Ich bitte Sie ausdrücklich, darauf zu verzichten. Wir haben alle Stellungnahmen bekommen. Wir haben sie alle sorgfältig studiert. Sie können davon ausgehen, dass die Dinge, die wir heute nicht zur Sprache bringen, jedenfalls

von unserer Seite nicht mehr diskussionsbedürftig sind. Das heißt nicht, dass wir nicht vielleicht unterschiedlicher Auffassung sind, aber es ist uns jedenfalls klar, dass und warum wir unterschiedlicher Auffassung sind. Wir haben da keinen Fragen- und Klärungsbedarf.

Die Punkte, die für uns noch offengeblieben sind und bei denen wir aufgrund Ihrer Stellungnahmen und Ihrer Hinweise eher noch Diskussions- und Klärungsbedarf für uns sehen, haben wir in der Tagesordnung zusammengefasst. Der TOP 4 „Verschiedenes“ bietet wie üblich noch einmal Gelegenheit, weitere Punkte anzusprechen, wenn Sie möchten. Aber es ist jedenfalls nicht vorgesehen, dass hier sozusagen noch einmal Dinge, die wir schriftlich nachlesen konnten, wiederholt werden.

Die Tagesordnung liegt vor. Gibt es jetzt schon Fragen zur Tagesordnung? Wie gesagt: „Verschiedenes“ bietet Möglichkeiten, sich dann noch zu äußern, wenn man möchte. Gibt es Fragen zum Ablauf? - Das sehe ich jetzt nicht. Dann können wir mit der Tagesordnung loslegen, und Herr Mensch führt in TOP 1 ein.

4.3.2 Tagesordnungspunkt 1: Grenzwert für die Vollständigkeit der Daten

Alexander Mensch: Vielen Dank Herr Windeler. - Kommen wir zu TOP 1, zum Grenzwert der Vollständigkeit der Daten. In mehreren Stellungnahmen wird der Grenzwert für die Vollständigkeit der Daten von 70 % als zu hoch angesehen. Unsere Frage wäre: Welcher Grenzwert erscheint Ihnen als angemessen, und wie würden Sie diesen begründen?

Moderator Jürgen Windeler: Zwei Wortmeldungen nach angemessener Bedenkzeit. - Bitte schön.

Marcus Schmidt: Aus meiner Sicht ist die Höhe des Grenzwertes gar nicht so entscheidend, sondern der wesentliche Punkt ist - das ist ja auch in den Arbeiten, auf die Sie sich beziehen, von Simon und anderen herausgekommen -: Die Repräsentativität der untersuchten Tumore muss gegeben sein. Das ist sicherlich wichtiger, als sich jetzt darüber zu streiten, ob man einen Grenzwert von 70 % oder 7 % oder 17 % haben möchte. Es kommt auf die Repräsentativität an. - Danke.

Moderator Jürgen Windeler: Danke. - Sie hatten sich gemeldet?

Christoph Petry: Ich habe dem nichts hinzuzufügen.

(Zuruf: Mikro bitte!)

Moderator Jürgen Windeler: Gut, vielen Dank. - Ja, bitte schön.

Marion Kiechle: Natürlich muss auch berücksichtigt werden, dass für die Analyse so einer Biomarkerstudie eine genügend hohe statistische Power neben der klinischen Repräsentativität vorhanden ist. Das wiegt natürlich viel mehr als ein fixer Anteil von Tumorproben aus einer prospektiv-retrospektiven Studie.

Moderator Jürgen Windeler: Frau Janatzek, dann Herr Scheibler.

Sandra Janatzek: Uns ging es im Wesentlichen auch darum, zu sagen, dass, wenn wir uns dafür entschieden haben, prospektiv-retrospektive Analysen von RCTs bzw. randomisierten Studien einzubeziehen, es dann widersprüchlich erscheint, den üblichen Nichtberücksichtigungsanteil anzulegen - mit einer Einschränkung: Wenn sich dieser Nichtberücksichtigungsanteil anders berechnet, als Sie es sonst üblicherweise tun, wenn Sie also im Nenner zum Beispiel bereits berücksichtigen, dass verlorengegangene Patientinnen, sprich Patientinnen, für die kein archivierte Tumormaterial verfügbar ist, die aber, was die erhaltene Randomisation angeht, kein Problem darstellen ... Wenn diese Patientinnen im Nenner unberücksichtigt bleiben, kann das vielleicht trotzdem ein sinnvolles Vorgehen sein, diese Schwelle zu berücksichtigen. Also will ich eigentlich nur sagen: Aus unserer Sicht ist der entscheidende Punkt, in dieser spezifischen Situation mit den prospektiv-retrospektiven Studien zu entscheiden: Können wir davon ausgehen, dass die ausgewerteten Patienten eine die die Randomisation erhaltene Auswahl darstellen?

Moderator Jürgen Windeler: Bitte schön.

Manfred Dietel: Herr Windeler, die Frage an Sie: Dürfen wir auch Gegenfragen stellen? Denn Sie haben jetzt die Frage gestellt: Wie begründen wir das? Ich würde gern die Frage stellen: Wie begründen Sie die 70 %? Warum nicht 65 %, 75 % oder anders? Das war mir nicht eindeutig erklärt, zumindest für mich nicht plausibel in dem Bericht.

Moderator Jürgen Windeler: Selbstverständlich darf hier auch jeder Gegenfragen stellen. Andererseits: Ich habe gerade noch einmal geschaut und in einer Reihe von Stellungnahmen gesehen: 70 % sind nicht adäquat, was für mich nur bedeuten kann: 70 % ist zu hoch. Ich habe nicht viele Begründungen gefunden bis auf den Hinweis: In der Literatur steht 66 %. Da kann ich jetzt keinen großen Unterschied feststellen.

Wir können natürlich die 70 % begründen, aber ich habe hier im Moment gehört, auf die 70 % käme es gar nicht an. Wir müssen uns dann einmal irgendwie darauf verständigen, worauf es eigentlich ankommt und ob die Zahl überhaupt von Bedeutung ist. Ansonsten könnten und sollten wir das jetzt - wir sind genau dazu hier - thematisieren. - Herr Scheibler, Sie hatten sich gemeldet.

Fülöp Scheibler: Vielleicht in der Verlängerung des Argumentes von Frau Janatzek: In anderen Kontexten, also beispielsweise, wenn wir den Nutzen von Medikamenten bewerten, fordern wir ja auch randomisiert-kontrollierte Studien aus gutem Grunde, weil wir den kausal nachweisbaren Effekt des Medikamentes auf die Endpunkte prüfen wollen. Wir gehen dann eben davon aus, dass, wenn ich eine Attrition von $> 30\%$ habe, dann die ursprüngliche Strukturgleichheit von Interventions- und Kontrollgruppe nicht mehr gewährleistet sein kann, dass möglicherweise oder sehr wahrscheinlich ein Confounding die Ergebnisse verzerrt.

Jetzt frage ich mich: Warum soll das bei diesen prospektiv-retrospektiven Analysen anders sein? Ich gebe Frau Janatzek vollkommen recht, dass, wenn beispielsweise in multizentrischen Studien einzelne Zentren keine Tumorproben konserviert haben und diese Zentren dann für die Auswertung nicht infrage kommen, das möglicherweise nicht verzerrend auf die Ergebnisse wirkt. Wenn aber für den großen Teil dieser ursprünglichen Population nicht klar ist, aus welchen Gründen sie aus der Studie bzw. aus der Auswertung verlorengegangen sind, und wenn nicht mit Sicherheit ausgeschlossen werden kann, dass diese Attrition möglicherweise mit Eigenschaften des Tumors oder mit Eigenschaften auch der Therapie in Zusammenhang gebracht werden kann, dann sehe ich die Gefahr, dass eine Verzerrung vorliegen kann. Und diese Verzerrung ist genauso wie in jeder anderen randomisiert-kontrollierten Studie. Ich sehe da keinen Unterschied zwischen den Bewertungen, die wir sonst machen, und den Grenzwerten, die wir in unseren Methoden festgelegt haben. Übrigens: Die wenigen Stellen, die es in der Literatur gibt, nennen da eher 80 %. Da sind wir vergleichsweise liberal. Diesen Unterschied kann ich hier nicht erkennen.

Moderator Jürgen Windeler: Frau Kiechle.

Marion Kiechle: Ich kann Sie ein Stück weit verstehen. Das kennen wir ja, dass wir, wenn wir einen Teil nehmen aus einer Studie, irgendeinen Bias befürchten. Aber wenn man sich dann mal die Ergebnisse anschaut von jetzt zum Beispiel dem Oncotype-DX-Test oder auch dem EndoPredict-Test durchgeführt in dieser Art eine prospektiv-retrospektive Studie und wirklich schaut, wie die Niedrigrisikogruppe verläuft, finden Sie da keinen Unterschied.

Ich hatte im Vorfeld da noch eine Grafik mit diesen Studien zusammengestellt für Sie. Wenn man sich die Ergebnisse anschaut, kommen alle 6 Studien unabhängig voneinander zu dem gleichen Ergebnis. Das macht doch Verzerrung höchst unwahrscheinlich. Alle 6 Studien kommen zu konsistenten Ergebnissen: mit 2 verschiedenen Genexpressionstests und mit unterschiedlichen Kollektiven.

Moderator Jürgen Windeler: 2 Wortmeldungen nach angemessener Bedenkzeit. - Bitte schön. Herr Schmidt hatte sich jetzt zuerst gemeldet.

Marcus Schmidt: Ich wollte als Gegenrede zu Herrn Mensch noch einmal anmerken, dass ich ihm natürlich recht gebe im Kontext von randomisierten klinischen Studien, dass wir aber - ich meine, das ist das Entscheidende - in dem Kontext von Prognoseabschätzungen überhaupt keine prospektiven bzw. prospektiv-randomisierten Studien brauchen. Letzten Endes geht es einfach darum, einzuschätzen: Welche Patientin hat eine gute Prognose, welche Patientin hat eine schlechte Prognose? Das heißt: Da kommt nichts Randomisiertes herein. Daher teile ich auch überhaupt nicht die Bedenken, dass es beim etwas geringeren Prozentsatz der Patienten zu einer Imbalance und zu einer Verzerrung kommen könnte.

Moderator Jürgen Windeler: Bitte schön.

Peter Sinn: Ich habe mir einmal die im Internet vorliegenden Nutzenbewertungen durchgesehen, die systematischen Reviews und die ökonomischen Nutzenbewertungen. Das sind ein paar Dutzend. In keine dieser systematischen Reviews und Nutzenbewertungen wird dieses Kriterium berücksichtigt oder gilt gar als Knock-out-Kriterium. Ich frage mich, warum ausgerechnet jetzt im Rahmen dieser Nutzenbewertung dieses Kriterium derart in den Vordergrund gestellt wird, während es in anderen - es gibt, wie gesagt, zahlreiche systematische Reviews und Nutzenbewertungen - keine Rolle spielt. Wenn man die Arbeit von Simon hineinschaut, sieht man: Er unterscheidet 4 verschiedene Kategorien von der Güte von Biomarkerstudien, und der Tenor dieser Arbeit von Simon 2009, in der dieses Kriterium genannt wird, ist, dass man kritisch sein muss bei Biomarkerstudien bezüglich der Qualität der Methodik. Dieses Kriterium wird bei den Studien der Gruppe A, der Kategorie A in den Vordergrund gestellt. Es werden werden aber auch andere Kriterien für Studien in geringerer methodischer Qualität genannt. Ich frage mich, ob es wirklich gerechtfertigt ist, a priori diesen Standard anzusetzen und damit praktisch die große Mehrzahl aller publizierten Evidenz auszuschließen. Ich will darauf hinaus: Ist es nicht so, dass dadurch, dass Sie praktisch den fast kompletten Satz der publizierten Evidenz ausschließen, ein höheres Verzerrungspotenzial entsteht, als dadurch, dass Sie dieses Kriterium berücksichtigen?

Moderator Jürgen Windeler: Danke. - Jetzt habe ich Herrn Lange und dann noch einmal Herrn Schmidt.

Stefan Lange: Ich frage jetzt für mich ...

(Zuruf: Das Mikro ist nicht an!)

- Jetzt ist es an, okay.

Bis jetzt haben wir noch keine Antwort gehört, was sozusagen Gegenvorschläge sind. Ich will es mal anders versuchen. Ich könnte mir vorstellen, dass wir hier in der Runde Konsens erzielen, wenn wir sagen würden: Da sind nur 5 % der Proben in die Studie eingeschlossen worden. - Würden Sie auch noch sagen: Nö, macht nix? Finden wir irgendwie ausreichend?

Irgendwo wird man doch sagen: Das kann nicht hinhalten, und einen Grenzwert braucht man doch. Oder nehmen wir mal 1 %? - Das würde mich dann doch noch mal interessieren.

Moderator Jürgen Windeler: Jetzt hatte sich Herr Schmidt gemeldet, dann sind Sie dran.

Marcus Schmidt: Zu diesem wesentlichen Punkt, wie hoch der Grenzwert sein soll, möchte ich oder muss ich etwas redundant sein: die Repräsentativität und das, was Frau Kiechle gesagt hat, natürlich die Power, die erzielt werden kann. Natürlich kann man sich vorstellen, dass man mit nur 1 % der untersuchten Proben Schwierigkeit mit der Repräsentativität und vor allen Dingen mit der statistischen Power bekommt, aber es kommt nicht auf den absoluten Wert an.

Moderator Jürgen Windeler: Jetzt da drüben, dann Frau Rutjes.

Christoph Petry: In der Arbeit von Simon, die jetzt in dieser Runde mehrmals zitiert worden ist, steht explizit drin, dass es keinen allgemeinen Grenzwert geben kann. Und das ist, glaube ich, in der Diskussion hier auch von mehreren Diskussionsrednern sehr deutlich gesagt worden: Es gibt keinen Grenzwert.

Insofern wird Ihnen, Herr Lange, die Frage nach der Zahl hier nicht beantwortet werden. Die Antwort gibt es nicht. Es geht ums Verzerrungspotenzial. Das ist bei allen Antworten herauskommen. Herr Schmidt ist ebenso weit gekommen: 1 % reicht, wenn es unverzerrt ist. Wenn Sie eine Studie mit 100.000 Patienten und 1 % haben, können Sie eine ziemlich gute Stichprobe bekommen.

Ich habe eine Antwort gehört aus der Diskussion: Es gibt keine Zahl als Grenzwert. Sie müssen wirklich in die Statistik reingehen und schauen: Ist die repräsentativ für die Kohorte? Und es ist tatsächlich - das war ein Kommentar eben von anderer Seite - auch nicht relevant, wenn man diese Proben, diese Patienten aus einer randomisierten Studie rausnimmt, dass man die Repräsentanz in dieser Randomisierung wirklich wiedertrifft. Die Biomarker versuchen ja gar nicht, die Frage zu beantworten, ob jetzt eine Patientin mit einem Aromataseinhibitor oder mit Tamoxifen behandelt werden soll. Es geht um Beantwortung der Frage nach der Prognose. Und da wirklich entscheidend, allein entscheidend die Frage: Ist das Kollektiv, das man gewählt hat - 66 %, 80 %, 1 % -, relevant und repräsentativ für die Grundgesamtheit der Patientinnen, die eben mit Tamoxifen, mit Aromataseinhibitor behandelt werden?

Moderator Jürgen Windeler: So, Frau Rutjes.

Anne Rutjes: Diese Prozentzahl ist schwierig zu definieren, und ich stimme ein: Es ist sehr wichtig zu verstehen, warum Patienten oder Tumorproben fehlen. Und da liegt dann auch das Problem. Denn oft können wir nicht einschätzen, wir können einfach nicht einschätzen, ob das Fehlen einen Bias hervorruft. Bei manchen Studien sind die Tumorproben schon in vorherigen Studien benutzt, und dann kann ich doch nicht annehmen, dass das ein Missing-Completely-at-random-Problem ist. Nein, wahrscheinlich hängt das zusammen mit den Endpunkten. Dann habe ich schwere Bedenken. Wenn ich jetzt über 30 % fehlende Daten habe, ist es eigentlich für mich unwichtig, ob das über den prädiktiven oder den prognostischen Wert geht, denn auch beim prognostischen Wert muss es repräsentativ sein. Wenn dann über 30 % fehlen, habe ich schwere Bedenken, dass die Zahlen noch irgendetwas aussagen können.

Moderator Jürgen Windeler: Frau Janatzek, dann habe ich Frau Mellert.

Sandra Janatzek: Wir stimmen den Ausführungen von Frau Rutjes voll zu. Wenn in besonders sorgfältiger Weise Bestrebungen unternommen wurden, Informationen zu erhalten, warum Patientinnendaten in den hier eingeschlossenen prospektiv-retrospektiven RCTs fehlen, ist es aus meiner Sicht unproblematisch - das ist die Position unseres Hauses -, dieses 70%-Kriterium wie immer anzulegen.

Unser Anliegen war aber, noch mal deutlich zu machen, dass wir eine besondere Sorgfalt in der Beschaffung solcher Informationen benötigen. Wir hatten bei der Ausführung, wie Sie die Nichtberücksichtigungsanteile konkret berechnet haben, den Eindruck, dass Sie das getan haben, konnten es aber leider nicht bis ins Detail nachvollziehen. Deshalb hatten wir hier einfach eine Rückfrage. Aber in dieser Weise stimmen wir dem Vorgehen zu.

Moderator Jürgen Windeler: Nun noch Frau Mellert, dann ziehe ich eine kurze Zwischenbilanz.

Uta Mellert: Ich habe 2 Punkte anzumerken. Ich möchte noch einmal darauf hinweisen, was Frau Prof. Kiechle gesagt hat und was hier auch schon mehrfach gefallen ist, dass die Qualität jeder einzelnen Studie bewertet werden muss. Das heißt: Es gibt keine Zahl, es steckt allerdings anerkennenderweise hinter Ihrer Arbeit ein enormer Aufwand, und es ist absolut verständlich, dass natürlich Kriterien gefunden werden müssen, nach denen man das beurteilen kann. Ich möchte aber trotzdem noch einmal zu bedenken geben, dass das vielleicht nicht das richtige Kriterium in dem Fall ist, und ich möchte auch dazu betonen: Es geht ja hier nicht um ein Arzneimittel, sondern um einen diagnostischen Test, der in der Konsequenz für den einzelnen Patienten erhebliche Auswirkungen hat. Umso mehr ist darauf zu achten, dass wir hier mit retrospektiv-prospektiven Daten absolut nach jeder einzelnen Studie bewerten und natürlich fragen: Warum fehlen die Tests?

Aber da gibt eben diverse Gründe. Die haben meiner Meinung nach nichts mit der Qualität zu tun, die hier das Kriterium der 70 % unterstützen würde.

Moderator Jürgen Windeler: Ich würde gerne versuchen, ein bisschen Zwischenbilanz zu ziehen, weil ich den Eindruck habe: Wir können diesen Punkt auch beenden. Wir kommen in keiner Weise voran, weil ich auch nicht erkennen kann, dass allen klar ist, worüber wir hier eigentlich sprechen. Wir bewerten die Diagnostika ganz genauso wie Arzneimittel. Die meisten würden wahrscheinlich sagen, man sollte an Arzneimittel strengere Kriterien anlegen. Sie haben gerade ein bisschen das Gegenteil gesagt. Also: Wir bewerten sie grundsätzlich gleich.

Das Zweite ist: Wir bewerten hier randomisierte Studien, keine Prognosestudien. Dieser Tagesordnungspunkt geht um randomisierte Studien. Das sollte eigentlich klar sein. Bei randomisierten Studien - Herr Scheibler hat vorhin darauf hingewiesen -, ist es nun üblicherweise so, dass man für die Auswertung alle Patienten in die Auswertung einbezieht. Das ist der Standard. Das heißt: Was wir eigentlich hier hineinschreiben sollten, ist: 100 %.

Jetzt sagen Sie alle einheitlich: Wenn durch das Fehlen von Patienten keine Verzerrung - Stichwort Repräsentativität - entsteht - das wäre die richtige Argumentation -, macht das nichts. - Ich möchte Sie ersatzweise statt nach der Begründung für die 70 % fragen, wer von Ihnen denn sagen kann - ein Argument habe ich vorhin bekommen -, dass das repräsentativ

ist. Das Problem ist - das ist leider so; auch da beißt die Maus keinen Faden ab -, dass man das nicht abschließen bewerten kann.

Deswegen ist unsere Position, dass wir sagen: Eigentlich brauchen wir 100 %. Wir haben es hier mit der Auswertung randomisierter Studien zu tun. Wir können möglicherweise in der Situation mit einem gewissen Attrition leben. Die Argumentation dafür und auch dafür, warum wir das in anderen Bewertungen auch machen, kann man übrigens im Methodenpapier nachlesen. Und deswegen können wir uns hier mit 80 % oder sogar hier mit 70 % Vollständigkeit auf der Basis, dass wir uns einigermaßen versucht haben, davon zu überzeugen, soweit das irgendwie möglich ist, ob die Studien dann in eine Verzerrung gehen, begnügen.

Aber die Aussage: „Das muss repräsentativ sein“, hilft uns überhaupt nicht. Denn das kann man in diesen Studien nicht bewerten. Ich habe den Eindruck, dass wir uns von den 70 % nicht wegbewegen, weil uns die Rückmeldung nicht hilft, Repräsentativität zu bewerten. Die Frage der Power hilft uns auch nicht, denn mit dem Powerargument würden wir sagen: so viele Patienten wie möglich und nicht weniger. Im Moment kann ich also noch nicht so richtig greifen, wie wir uns von diesen 70 % - und Sie merken an der Frage, dass wir interessiert daran sind, das durchaus mitzubekommen - eventuell wegbewegen könnten.

Jetzt habe ich Wortmeldungen. Sie waren die Erste, dann gehe ich so herum und dann Frau Mellert noch einmal.

Marion Kiechle: Herr Windeler, wie gesagt: Ich muss noch einmal zusammenfassen. Ich kann Sie irgendwo verstehen. Meine Empfehlung zu Ihrer ersten Frage wäre: Die 70 % würde ich in dem Fall nicht anwenden. Und Sie fragen mich: Warum? Oder: Was sollen wir denn dann hernehmen? Dann möchte ich auf diese 6 Studien zu sprechen kommen, die ja unabhängig voneinander mit 2 unterschiedlichen Biomarkertests zu den gleichen Ergebnissen kommen. Sie sortieren uns nämlich die Niedrigrisikogruppe heraus: die Frauen, die innerhalb von 10 Jahren gut wie keine Fernmetastasen entwickeln.

Ich habe mir auch die Mühe gemacht, zu schauen: Wie ist da die Berücksichtigungsquote? Das variiert von 14 % in der NSABP-B-14-Studie mit dem Oncotype durchgeführt bis zu 46 % in der TransATAC-Studie. Da sehen Sie: Das sind ganz unterschiedliche Margen, aber die Studien kommen unabhängig voneinander zum gleichen Ergebnis.

Um Ihre Frage zu beantworten: „Worauf sollen wir denn schauen“, sage ich: Schauen Sie doch auf die Ergebnisse dieser Studie. Es macht eine Verzerrung eigentlich für mich höchst unwahrscheinlich, wenn 6 voneinander unabhängige Studien zum gleichen Ergebnis kommen.

Moderator Jürgen Windeler: Okay, danke. - Bitte schön.

Christoph Petry: 2 Dinge. Erstens: randomisierte Studien und Prognosestudien. Sie haben im Ihrem Vorbericht auf der Basis der 70 % auch die prognostischen Studien ausgeschlossen. Und tatsächlich sind ja auch viele der Studien, die ursprünglich mal randomisierte Studien

waren, um eine spezifisch pharmazeutische Fragestellung zu beantworten zum Beispiel Aromataseinhibitor versus Tamoxifen, für die Biomarkerstudien prognostisch ausgewertet worden. Da ging es wirklich nur um rezidivfreies Überleben - 5 Jahre, 10 Jahre primär - unter endokriner Therapie, also keine randomisierte Studie, reine Prognosestudie. Und auch hier haben Sie die 70 % ausgeschlossen. Wir haben jetzt sehr viel das Argument diskutiert: 66 %, 70 %. Simon und Hayes: Steht es jetzt drin, oder steht es nicht drin? Es steht nicht drin mit den 66 %. Das ist eine Richtschnur, dass oberhalb dieser Größe die Verzerrung wohl vermieden werden kann. Aber es steht nicht drin, dass man das als harten Grenzwert annehmen soll. Im Gegenteil steht darin, dass es keinen harten Grenzwert gibt.

Stattdessen - das ist hier, glaube ich, jetzt noch nicht diskutiert worden, aber es geht in die Richtung des Arguments von Frau Kiechle - verlangen ja Simon und Hayes, wenn ein Biomarker auf der Basis dieser prospektiv-retrospektiven Studien verwendet wird, dass mindestens 2 Studien verwendet werden. Das heißt, zu Ihrer Frage, wie vermeide ich das Verzerrungspotenzial: Wie gesagt, durch die statistische Analyse ist eine Repräsentanz gegeben - ich denke, dass Sie die durchaus durchführen können: Was ist die Repräsentanz? Oder: Wie ist eine Kohorte zusammengesetzt in Deutschland, bei der die Gynäkologen die Frage nach Chemotherapie oder Nicht-Chemotherapie stellen? Aber auch wenn Ihnen diese Analyse zu trickreich erscheint, haben Sie die Möglichkeit, sich 2 Studien anschauen. Dann sind Sie auf jeden Fall deutlich konsistenter mit Simon und Hayes, als wenn Sie die 66 % fordern, die in der Publikation explizit nicht drinstehen.

Moderator Jürgen Windeler: Frau Mellert.

Uta Mellert: Ich möchte darauf hinweisen, dass die retrospektiv-randomisierten Studien mit einer anderen Fragestellung auf den Weg gebracht worden sind, nämlich nach einer Arzneimittelstudie, die ganz andere Zielführungen hat.

Um Ihre Frage zu beantworten, möchte ich auf den Vorschlag von Frau Prof. Kiechle zurückkommen und sagen: Man müsste nach jeder einzelnen Studie das Verhältnis anschauen nach den Zielführungen genau dieser Fragestellung dieser Biomarkertests und das im einzelnen berücksichtigen, ob es einen Bias gibt oder nicht und wie viele von diesen Tests durchgeführt worden sind und warum welche gefehlt haben.

Moderator Jürgen Windeler: Jetzt hatten Sie sich gemeldet.

Matthias Langenfeld: Ich stimme im Wesentlichen damit überein, dass die Repräsentativität der Studien in dem Simon-Paper von 2009 gefordert wird. Es gibt auch in der B-20-Studie mit dem Recurrence Score einen Versuch, das zu erörtern, indem einfach die Hazard Ratios und die P-Werte der verschiedenen klinischen Faktoren genannt werden, die sowohl in der untersuchten wie auch in der nicht untersuchten Population verglichen werden. Damit wird dargestellt, dass die untersuchte, biomarkeruntersuchte Population und die Studienpopulation der Elternstudie gleichzusetzen sind, sich zumindest da nicht unterscheiden.

Systematisch ist anzumerken, dass natürlich in einer prospektiven durchgeführten Studie die Untersuchungszahl bei 100 % bestehen bleibt. Das wäre ein eleganter Weg, um war aus den 70 % herauszukommen.

In diesem Zusammenhang sei vielleicht noch einmal darauf hingewiesen, dass eine solche prospektive Studie in Form der TAILORx vorliegt.

Moderator Jürgen Windeler: Stefan Lange.

Stefan Lange: Das stimmt, was ja auch ein Beispiel dafür ist, dass das geht. Manchmal wird gesagt, das gehe alles gar nicht. Aber es geht ja.

Vielleicht doch noch einmal: Also der Wert von 66 % steht in der Publikation. Man muss sich fragen: Warum schreibt er ihn dann rein, wenn er gar keine Bedeutung hat?

Vielleicht noch einmal, um klarzumachen, worum es geht: Eigentlich wollen wir prospektive Studien, die randomisiert sind. Jetzt haben wir keine prospektiven Studien, die randomisiert sind. Wir wollen eigentlich die Interaktion, also den prädiktiven Wert der Biomarker in einer solchen prospektiven und randomisierten Studie bestimmen. Jetzt haben wir keine prospektiv-randomisierten Studien. Daher weichen wir auf retrospektiv-prospektiv-randomisierte Studien aus. Das schränkt schon mal die Aussagefähigkeit deutlich ein. Solche Studie sind schon mal sehr viel schwächer. Aber wir gehen diesen Weg.

Aber eigentlich wollen wir 100 %. Dann sagen wir: 100 % kann man nicht immer haben. Bei 99 % würden wir auch sagen: Das 1 % macht wohl nicht so furchtbar viel. - Jetzt gehen wir sogar auf 70 %. „70 %“ bedeutet: Wir berücksichtigen die, auch wenn wir überhaupt nichts über die anderen 30 % wissen. - Natürlich ist das völlig richtig, wenn wir genau wissen, was der Grund ist, warum die Proben ausfallen. Herr Scheibler hatte ein Beispiel gebracht; das könnte auch sein, dass irgendein Umweltereignis irgendein Labor zerstört hat: Gerne, von mir aus, okay, das ist sicher ein für die Forschungsfrage zufälliges Ereignis und hat nichts mit den Ergebnissen zu tun. Aber Sie wissen ganz genau, dass wir genau eben nicht aus den Studien diese Gründe so exakt herausfinden können, warum die Proben fehlen. Es nützt einem auch die Untersuchung mit den mehr oder weniger bekannten prognostischen Faktoren nichts, zu schauen, ob die vergleichbar sind oder nicht. Das ist das Riesenproblem. Wenn das so einfach wäre, bräuchte man den ganzen Aufwand nicht. Also noch einmal die Frage: Über das Allgemeine hinaus, man müsse die Repräsentativität ... Warum müssen wir überhaupt die Repräsentativität prüfen? Ich meine: Das muss doch in den Studien vernünftig dargestellt werden. Das ist doch nicht unser Job. Ich schaue mir gerne Originaldaten an, aber dafür haben wir leider keine Zeit. Das ist doch völlig ausgeschlossen. Also noch einmal: Warum sollen wir das prüfen? Es muss doch in den Studien nachgewiesen werden. Und wenn die das nicht machen, haben Sie ja vielleicht die Daten, dann sagen Sie uns das doch und weisen es uns nach, wie repräsentativ die Daten sind.

Und jetzt gehen wir sogar noch von den prospektiv-randomisierten oder retrospektiv-randomisierten Studien weg und sagen: Okay, wir können uns da ja vorstellen, dass wir auf alleinige Prognosestudien zurückgehen, wenn denn das Ergebnis so irrsinnig eindeutig ist, weil wir dann indirekt einen prädiktiven Wert ableiten. Und selbst daran sollen wir keine Anforderungen stellen können, um irgendwie zu sagen, dass uns diese Daten irgendetwas aussagen? Ich bin schon sehr erstaunt.

16 % bei allem, was recht ist. Diese Prognose würde ich nicht für irgendetwas heranziehen wollen. Und zum Argument, die Studien kämen alle auf die gleichen Daten, sage ich: Wenn alle Studien die gleichen Fehler machen, kommt man immer auf den gleichen Fehler. Das ist ein etwas waghalsiges Argument.

Moderator Jürgen Windeler: Jetzt habe ich Herrn Sotlar, wenn ich das richtig lese,

(Karl Sotlar: Richtig!)

und dann Herrn Sinn.

Karl Sotlar: Herr Lange, glauben Sie, dass Sie heutzutage bei einer aufgeklärten Patientin oder in einem aufgeklärten Patientenkollektiv eine prospektive Studie durchführen könnten, wo Sie in der Gruppe der Low-Risk-Patientinnen - egal, welchen Test wir durchführen - noch eine Randomisierung „Chemotherapie/keine Chemotherapie“ durchführen könnten? Ich glaube nicht, dass ein solches Kollektiv zustandekommen würde in einer ausreichend großen Zahl.

Moderator Jürgen Windeler: Das lenkt jetzt ein bisschen von der Frage ab, dass wir eigentlich bei den Studien, die wir haben, die wir nun leider nur haben, 100 % an Daten haben wollen und fragen, warum 30 % fehlen dürfen.

Karl Sotlar: Aber es war doch die Aufforderung von Herrn Lange, eine prospektive Studie zu dem Thema durchzuführen. Und ich stelle ganz massiv infrage, dass das heutzutage überhaupt möglich wäre.

Moderator Jürgen Windeler: Die Antwort von Herrn Lange war eine Replik auf Hinweis, dass es eine solche Studie gibt, und seiner explizite Äußerung war: Man kann solche Studien offenbar machen, hätte sie also in der Vergangenheit auch schon beginnen können. - Herr Sinn, bitte.

Peter Sinn: Herr Lange hatte die Frage gestellt nach den möglichen Gründen, die für den Nichteinschluss von Biomaterialien in eine Studie ausschlaggebend sein könnten, und angemerkt, dass es aus seiner Sicht nicht auszuschließen ist, dass dadurch ein systematischer Fehler, ein Bias, eine Verzerrung induziert wird. Dazu würde ich anmerken, dass der Hauptgrund, dass Biomaterialien fehlen oder auch nicht eingebracht werden können, einfach der ist, dass das Biomaterial endlich ist. Es geht hier um prognostisch relativ günstige

Tumoren, kleine Tumoren, und der Pathologe - auch wir - sind nicht in der Lage, jede Studie zu bedienen.

Der Pathologe ist aber andererseits geblindet gegenüber dem Outcome, ist auch geblindet gegenüber den übrigen Studienkriterien einschließlich der Therapie, sodass ich überhaupt nicht erkennen kann, warum die Entscheidung oder die Unmöglichkeit, hier Biomaterial einzubringen, weil es möglicherweise aufgebraucht ist, in irgendeiner Weise hier zu einer Verzerrung führt.

Moderator Jürgen Windeler: Vielleicht nur mal der Hinweis: Wenn das Probenmaterial, was Sie haben, wie ich es gerade verstanden habe - vielleicht habe ich es falsch verstanden -, mit der Größe des Tumors irgendwie korreliert, dass also mit dem, was Sie sagen, offensichtlich ist, dass sehr kleine Tumoren in diesen 70 % nicht enthalten sind

(Anne Rutjes: Genau!)

oder systematisch mehr nicht enthalten sind und dass es eine nicht so fernliegende, ja förmlich naheliegende Idee ist, dass da eine Verzerrung entsteht. Vielleicht habe ich Sie aber missverstanden mit dem, was Sie beschrieben haben.

Peter Sinn: Das ließe sich ja prüfen, aber ich denke, dass das nicht zutrifft, weil das Sampling immer eine etwa gleichbleibende Probenmenge auch bei größeren Tumoren und bei kleinen Tumoren hervorbringt.

Moderator Jürgen Windeler: Frau Kiechle.

Marion Kiechle: Ich wollte gern noch etwas zu Ihnen sagen, Herr Lange, zur Repräsentativität. Wie gesagt, ich kann Sie ja voll verstehen: Wenn man so etwas bewertet, muss wirklich schauen, wenn man auf diese 70 % nicht kommt: Stimmen denn dann die anderen Gütekriterien? Aber um Sie vielleicht da ein bisschen zu beruhigen, sage ich: Wenn Sie einmal in die internationale Landschaft der wissenschaftlichen Gremien schauen - schauen Sie sich das NICE an oder auf europäischer Ebene das ESMO oder auch uns, die AGO, die Arbeitsgemeinschaft Gynäkologische Onkologie unter anderem -, haben sich das ja auch anschaut. Sie kommen zu den gleichen Erkenntnissen: Okay, wir haben nicht diese 66 % - wie bewerten wir das? Stimmen denn die anderen Gütekriterien? Und die kommen ja zu diesem positiven Ergebnis, haben also die Repräsentativität oder die Kriterien nach Simon und Hayes auch geprüft - einfach, um Ihnen da noch ein Argument zu geben.

Moderator Jürgen Windeler: Frau Mellert, bitte.

Uta Mellert: Ich möchte auch noch mal kurz auf Sie zurückkommen, Herr Lange,

(Stefan Lange: Schön!)

und auch auf die Forderung, prospektiv-randomisierte Studien in dieser Fragestellung durchzuführen. Ich denke, im Sinne des Patientenschutzes und dessen, was der Herr Prof. Sotlar angemerkt hat, ist es nicht vertretbar, dass wir solche Studien extra auflegen, sondern dass es natürlich nahe liegend ist, dass man Studien nimmt, die prospektiv designt sind mit den Medikamenten und mit den Gegebenheiten in dieser Fragestellung, und dass man auf diese Studien zurückgreift und die Tests zunächst da durchführt, um auch langzeitprognostische Werte darstellen zu können - so weit zu den bisherigen Studien.

Moderator Jürgen Windeler: Herr Lange und dann Herr Schmidt.

Stefan Lange: Herr Windeler hat es eigentlich schon gesagt, aber ich sage es jetzt noch mal, weil Sie wieder schon darauf zurückkommen. Ich habe gar keine Forderungen aufgestellt. Ich habe nur gesagt, wo wir eigentlich gestartet sind und dass wir also schon in den Anforderungen eben viel weiter runtergegangen sind, als es eigentlich üblich ist, sodass wir uns schon auf einem Evidenzniveau bewegen, das sowieso schon schwächer ist, als das, was wir sowieso schon haben, also wir eigentlich noch viel mehr auf potenzielle Verzerrungsmechanismen achten müssen. Das ist das einzige, was ich gesagt habe. Ich habe gar nicht gefordert, dass man noch zig randomisierte Studien macht. Also: Durch die Wiederholung wird es nicht richtiger und auch nicht besser. Das zu sagen, ist mir noch mal ein Anliegen.

Moderator Jürgen Windeler: Herr Schmidt.

Marcus Schmidt: Zu den randomisierten Studien möchte ich dann doch noch sagen, dass man schon klar sagen muss - das war vorhin schon angeklungen: Es galt für das Low-Risk-Stratum. Von Genomic Health gibt es ja jetzt die Ergebnisse, ehrlich prospektiv ausgewertet, nur für das Low-Risk-Stratum bis jetzt. Und jetzt ist gerade 2015. Das, was im Jahr 2004, also 11 Jahre zurück prospektiv-retrospektiv vorhergesagt worden ist, nämlich dass das Low-Risk-Stratum eine sehr gute Prognose hat, ist jetzt mit viel Aufwand, 11 Jahre danach bestätigt worden - das nur noch einmal zu dem doch immer wieder hochgehaltenen Mantra, was prospektiv-randomisierte Evidenz angeht.

Ich muss dann auch noch einmal zu dem zurückkommen, was ich eingangs schon einmal erwähnt hatte, dass zur Frage der Prognose prospektiv-randomisierte Studien in sich betrachtet gar keine echte Rolle spielen. Es ist wichtig, dass das Material prospektiv und sauber gesammelt wird. Wenn ich auf Ihren Berichtsplan zurückkommen darf, sieht man: Von knapp 10.000 Literaturstellen, die mit allem Fleiß vom IQWiG gesammelt worden sind, bleibt eine einzige Studie von Martin übrig: prospektiv-randomisiert, knapp über 70 %, offensichtlich Gewebe eingebracht und ausgewertet.

Dann muss ich allerdings sagen: Erstens ist irgendwie erschütternd, dass man nach 10 Jahren Prognosefaktorforschung mit Genexpressionsanalysen lediglich 1 auswertbare Studie hat und dann - das ist uns dann doch etwas irritierend vorgekommen - wird die noch falsch

interpretiert. Denn die Interpretation dieser Studie, wie sie in Ihrem Bericht dargelegt worden ist, ging ja nicht dahin, dass der betreffende Test - das war der EndoPredict - in diesem Fall den Unterschied zwischen den beiden Armen, weswegen diese Studie ursprünglich randomisiert worden war, festlegen soll, sondern einfach ganz schlichtweg - und damit bin ich auch schon am Ende -: Wie ist die Prognose? Low Risk versus High Risk. - Vielen Dank.

Moderator Jürgen Windeler: Herr Scheibler.

Fülöp Scheibler: Herr Schmidt, das muss ich jetzt ein bisschen richtigstellen. Generell habe ich den Eindruck, dass im Moment die Diskussion sehr durcheinander läuft. Ich finde, es wäre hilfreich, wenn wir zwischen randomisierten Studien mit dem Ziel, eine Wirksamkeit zu testen, und Prognosestudien unterscheiden. Es wurde schon erwähnt: Es gibt Auswertungen aus randomisiert kontrollierten Studien, die dann prognostisch sind. Wir haben die dann als Prognosestudien betrachtet.

Es wäre sinnvoll, glaube ich, für unsere Diskussion, wenn wir sehr präzise unterscheiden würden zwischen prognostische Fragestellungen und prädiktiven Fragestellungen. Für prädiktive Fragestellungen - ich glaube, da sind wir uns alle einig - bedarf es randomisiert kontrollierter Studiendesigns, in denen beide Patientengruppen den Marker erhalten haben. Dann wird in einer Patientengruppe eine Intervention, beispielsweise eine Chemotherapie durchgeführt. Dann wird die Interaktion zwischen dem Wert des Markers und der Tatsache, ob eine Chemotherapie gegeben wurde oder nicht, auf die Outcomes gemessen.

Von dieser Art von Studien - damit wir jetzt mal bei den Fakten dieses Berichtes bleiben - haben wir nur 2 gefunden. Frau Kiechle, Sie haben 6 erwähnt. Da müssen wir uns im Nachgang noch einmal unterhalten. Wir haben nur 2 gefunden. Alle anderen Studien verletzen andere Einschlusskriterien dieses Berichts. Von den beiden Studien, die wir gefunden haben, ist ja eine Studie ausgewertet worden, und die andere Studie ist eingeschlossen. Wir konnten nur die Daten nicht verwerten. Das ist ebendiese Studie, also die Paik-Publikation.

Da muss man jetzt noch mal deutlich sagen: Die in der Analyse verbleibenden Tumorproben betragen etwa 30 % der ursprünglichen Stichprobe. Also: Es geht nicht um 60 % versus 58 % oder 70 % versus 65 %. Es geht darum, dass wir noch 30 % der ursprünglich eingeschlossenen Patienten zur Verfügung haben. Hier hinten wurde es erwähnt: Wenn man sagt, man braucht wenigstens eine Replikation so eines Ergebnisses - für den Marker in diesem Kollektiv haben wir keine Replikation dieses Ergebnisses vorliegen. Alle anderen vergleichenden Interventionsstudien oder randomisiert-kontrollierten Studien haben andere Einschlusskriterien dieses Berichts verletzt.

Zu der prognostischen Fragestellung könnte man jetzt auch viel sagen, aber ich möchte erst einmal bei dieser Nutzenfragestellungen, bei dieser Interaktionsfragestellung stehen bleiben und sagen: Haben wir über die beiden Studien hinaus vergessen, Studien einzuschließen oder sie nicht berücksichtigt? Gibt es über diese beiden Studien hinaus Studien, bei denen die 70-

%-Regel dazu führte, dass wir die Ergebnisse nicht werten konnten? Ja, das trifft für die eine einzige Studie zu. Für alle anderen Fragestellungen gibt es schlicht und einfach keine Evidenz. - So viel zu dem Thema.

Wenn Sie mich fragen würden: „Darf man in diesem Feld noch randomisiert-kontrollierte Studien machen?“, würde ich antworten: Solange eine Unklarheit bzw. Ungewissheit vorherrscht - ja.

Moderator Jürgen Windeler: Ich habe den Eindruck, dass wir dieses Thema jetzt sicherlich nicht abschließend diskutiert haben, aber dass wir hier nicht weiterkommen. Und ich glaube: Alle Argumente, die wir brauchen, auch um unsere Meinungsbildung weiterzuführen, sind jetzt ausgetauscht worden. Vielleicht darf ich mir nur den Hinweis erlauben, dass es natürlich - das liegt in der Natur der Sache - Studien gibt, die nach 10 Jahren Ergebnisse oder Annahmen, die man vor den 10 Jahren getroffen hat, bestätigen, und dass es eben Studien gibt, die nach 10 Jahren Ergebnisse oder Annahmen, die man getroffen hat, in drastischer Weise widerlegen. Das ist der Sinn von randomisierten Studien, dass man genau diese Entscheidung trifft.

Gibt es jetzt noch Wortmeldungen zu Argumenten allgemein, die jetzt noch nicht hier gefallen sind? - Ich glaube, wir haben viele Punkte gehört, die wir notieren, berücksichtigen und bei denen wir dann überlegen, wie wir weiter vorgehen, aber ich habe jetzt nicht den Eindruck, dass uns die Diskussion sehr viel weiterhilft. - Bitte schön.

Christoph Petry: Noch ein Kommentar zu Herrn Lange: Er fragte, wer müsse den Nachweis führen, dass keine Verzerrung in den Studien vorliege: das IQWiG oder derjenige, der das Paper geschrieben hat? Die Antwort es natürlich: derjenige, der das Paper geschrieben hat. Es ist aber allerdings so, wenn man sich diese Paper anschaut: In allen, die mir jetzt einfallen - das ist eine ganze Menge -, finden Sie an irgendeiner Stelle, normalerweise im Supplement, eine Tabelle, wo die Kriterien, also die deskriptive Statistik der Gesamtkohorte mit der deskriptiven Statistik der Stichprobe, also der Biomarkerkohorte verglichen wird. Das hat sich zum Beispiel bei den Leitlinien die AGO angesehen und ist zu dem Schluss gekommen: Da liegt keine Verzerrung vor.

Wenn Sie sagen, die Autoren müssten nachlegen, entgegne ich: Bei allen Studien, die mir jetzt einfallen, sind die Daten vorhanden und können also auch nachgelesen werden.

Moderator Jürgen Windeler: Stefan Sauerland.

Stefan Sauerland: Dieses Argument, dass die Charakteristika der ausgeschlossenen Patientinnen denen entsprechen, die am Ende noch in der Studie drin sind, trifft nicht ganz zu, weil es im Endeffekt darum geht, ob diese Patientinnen das Zielereignis haben. Das heißt: Diese Baselinecharakteristika führen da gar nicht richtig weiter. Und ich glaube: Wir unterliegen gerade auch dem Trugschluss, dass man mit den Informationen aus den Studien, selbst wenn

die ganzen Gründe haarklein erklärt sind in der Publikation, irgendwie eine Entscheidung so viel einfacher machen kann.

Denn wir haben auch gehört, dass das Probenmaterial dann oft verbraucht ist. Da ist relativ klar, dass es da einen Zusammenhang zum Risiko und auch zu dem Outcome der Patientinnen gibt. Aber es gibt jede Menge anderer Gründe, bei denen wir uns sicherlich tagelang streiten könnten, ob denn diese Gründe irgendwie mit der Therapie oder mit dem Therapieerfolg zusammenhängen. Wenn eine Frau die Studienteilnahme verweigert oder wenn eine Frau nach Australien auswandert, würden wir erst einmal denken: Das hat eigentlich nichts damit zu tun, welche Therapie erfolgte oder ob ein Rezidiv eingetreten ist oder nicht oder ob ein Rezidivrisiko besteht. Aber auch da kann man sich Mechanismen vorstellen, die dazu führen, dass eben hier auch solche Follow-up-Verluste einer Verzerrung der Auswertung herbeiführen.

Moderator Jürgen Windeler: Gut. - Bitte.

Eva Schumacher-Wulf: Da jetzt doch ganz offensichtlich wieder prospektive Studien gefordert werden wie schon bei unserer letzten Sitzung hier, würde ich doch gerne noch einmal die Frage stellen, wie so ein Studiendesign aussehen könnte. Ich habe im Anschluss an unsere letzte Sitzung schriftlich angefragt beim IQWiG, ob ich mal einen Vorschlag für ein ethisch vertretbares Studiendesign bekommen könnte. Das habe ich nicht bekommen. Es waren nur ausweichende Antworten. Daher werde ich jetzt noch einmal die Frage stellen, wie Ihrer Meinung nach diese prospektive Studie designt sein könnte.

Moderator Jürgen Windeler: Herr Lange hat zweimal gesagt, dass er das nicht gefordert hat.

(Eva Schumacher-Wulf: Ja!)

Herr Scheibler hat gesagt, dass er persönlich sich so eine Studie vorstellen kann.

(Eva Schumacher-Wulf: Nach welchem Design?)

- Die Studie, die man sich da vorstellen kann, hängt sehr stark an dem konkreten Design. Wir werden nicht anfangen, das Design hier zu diskutieren. Die Diskussion, die vorhin mehrmals geführt worden ist, ist die, dass wir sagen - das dürfen wir, glaube ich, mit allem Fug und Recht sagen, insbesondere da wir mindestens so erschüttert sind wie Sie, was die Evidenzlage angeht, dass wir gerne solche Studien gehabt hätten. Wenn Sie jetzt sagen, solche Studien könne man jetzt nicht mehr machen, ist, glaube ich, der Zeitpunkt, wo man sie hätte anfangen können, sodass die Ergebnis jetzt vorliegen und wir entspannter diskutieren könnten, ist leider lange vorbei, bedauerlicherweise lange vorbei. Das hätte man längst machen können.

(Eva Schumacher-Wulf: Das bringt ja jetzt nichts!)

Ich würde vorschlagen, dass wir den Punkt abschließen und zum Tagesordnungspunkt 2 übergehen. Das mache ich jetzt auch und gebe Herrn Mensch wieder das Wort zur Mindestdauer.

4.3.3 Tagesordnungspunkt 2: Mindestdauer der Nachbeobachtung

Alexander Mensch: Für Prognosestudien wird im Bericht eine mediane Mindestdauer der Nachbeobachtung von 10 Jahren gefordert. Diese wird in mehreren Stellungnahmen für nicht notwendig erachtet. Die mittlere Nachbeobachtungszeit sei nach Auffassung dieser Stellungnehmenden für eine Kaplan-Meier-Analyse ohne Relevanz. Es sei geradezu ein Charakteristikum dieser Methode, dass sie zuverlässige Aussagen für Zeitintervalle deutlich jenseits der mittleren Nachbeobachtungszeit zulasse, insofern für diese Zeitintervalle noch Datenpunkte in ausreichender Anzahl vorlägen.

Eine mediane Nachbeobachtung von 10 Jahren bedeutet, dass 50 % aller eingeschlossenen Patienten 10 Jahre oder länger nachbeobachtet wurden. Wie hoch sollte Ihrer Ansicht nach der Anteil der in der Auswertung verbliebenen Patientinnen nach 10 Jahren mindestens sein, um zuverlässige Aussagen für das ursprünglich einzuschließende Patientenkollektiv treffen zu können, oder - anders gesagt - um einen Selektionsbias auszuschließen?

Moderator Jürgen Windeler: Also wieder die Frage nach einer Zahl, diesmal nach einer etwas handhabbareren Zahl, habe ich den Eindruck, als bei den 70 %. - Sie hatten sich gemeldet, bitte schön, Herr Langenfeld.

Matthias Langenfeld: Der ursprüngliche Auftrag des G-BA bezog sich auf die Anwendung von biomarkerbasierten Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie. Auch wenn im Verlauf jetzt schon angedeutet wurde, dass es Möglichkeiten gibt, das rein prognostisch zu betrachten, so ist doch in der ursprünglichen Fragestellung dies eine prädiktive Fragestellung zum Nutzen der Chemotherapie. Dieser ist nach Vorlage der Evidenz beschränkt auf die ersten 5 Jahre nach Anwendung der Chemotherapie.

Wir haben unterschiedliche Risikozeiträume. Die Frage, die eben gestellt wurde, bezieht sich ja auf die 10 Jahre. Im ursprünglichen Antrag wird auf die Wirkung bzw. den Entscheid für oder gegen eine Chemotherapie abgehoben. Chemotherapien sind zum Beispiel nach einer gut publizierten und auch vorliegenden Metanalyse vom EBCTCG nach der Gruppe im Zeitraum zwischen 5 und 10 Jahren ohne Effekt - weder auf das rezidivfreie noch auf das gesamte Überleben.

Unter diesen Umständen würde ich die Frage stellen, warum unter der grundsätzlichen Frage eines prädiktiven Testes die 10 Jahre gesetzt sind und man nicht auf den Zeitraum schaut, der durch Chemotherapie beeinflussbar ist, nämlich 5 Jahre

Moderator Jürgen Windeler: Gut, also Sie würden, um sozusagen die Anknüpfung zur konkreten Frage zu machen, den Vorschlag machen, die mediane Nachbeobachtungszeit nicht

auf 10 Jahre, sondern auf 5 Jahre zu legen, oder würden Sie auch noch den Median niedriger legen, sodass wir irgendwie nur bei 25 % bei 5 Jahren legen?

Matthias Langenfeld: Nein, ich denke, es geht wirklich um die 5 Jahre, und am Median würde ich nichts ruckeln.

Moderator Jürgen Windeler: Okay, „Median“ würde also bedeuten: Die Hälfte der Patientinnen dürfte kürzer und die Hälfte der Patientinnen sollte länger als die 5 Jahre beobachtet worden sein? Nur, damit wir alle klar darüber sind, worüber wir konkret reden. Das ist ein, glaube ich, verständlicher und nachvollziehbarer Alternativvorschlag zu den 10 Jahren, den wir jetzt verfolgt haben. - Frau Kiechle.

Marion Kiechle: Ich wage mich jetzt einmal vor, auch wenn ich kein Biostatistiker von Haus aus bin, und spreche nochmals für den Berufsverband. Es sind doch meines Erachtens die Studiengröße und das Konfidenzintervall viel wichtiger als die mittlere Nachbeobachtungszeit. Darauf schauen wir doch immer. Das ist doch letztendlich unabhängig von der mittleren Nachbeobachtungszeit. Ich möchte nur hervorheben, dass man doch darauf schauen sollte.

Moderator Jürgen Windeler: Ich habe 2 Wortmeldungen: Herrn Dietel und dann da hinten.

Manfred Dietel: Ich glaube, man muss hier auch ein bisschen die Tumorbiologie mit einbeziehen in den Gedankengang. Es geht ja nicht darum, dass alle Tumoren gleich sind. Und es geht bei diesem Test, wie Sie wissen, eher um die Low-Grade-Tumoren, deren Rezidiv in aller Regel überhaupt erst nach 5 Jahren und gehäuft in der zweiten Phase zwischen 5 und 10 Jahren auftaucht. Ob wir nun über Median oder so sprechen, darüber können wir diskutieren.

Es macht also eine kürzere Zeit meines Erachtens wenig Sinn, wenn man sich diesen Tumortyp anschaut. Es kann beim Non-Small Cell Lung Cancer völlig anders sein - darüber müsste man dann getrennt sprechen -, aber gerade die Low-Risk-östrogenrezeptorpositiven/HER-2-negativen Mammakarzinome haben ja primär schon eine relativ gute Prognose. Und deshalb sollte man bei der längeren - egal, wie das jetzt genau definieren - Überwachungszeit meines Erachtens unbedingt bleiben.

Moderator Jürgen Windeler: Wenn ich versuche, das zusammenzufassen, haben Sie nicht ausdrücklich unserem Vorgehen, was den Median nach 10 Jahren angeht, zugestimmt. Aber die 10 Jahre würden Sie jedenfalls bestehen lassen, und Sie würden möglicherweise im Median auf - ich weiß nicht was - 30 %, 25 %, 20 % heruntergehen?

Manfred Dietel: Eher in diese Richtung, ja.

Moderator Jürgen Windeler: Okay, gut.

Christoph Petry: Zu Ihrer Frage, Herr Prof. Windeler, ganz konkret: Die 10 Jahre würde ich auch stehenlassen, aber nicht den Median, weil er für die Kaplan-Meier-Analyse nicht wirk-

lich relevant ist, sondern das untere Ende des Konfidenzintervalls zum Beobachtungszeitpunkt, also bei 10 Jahren, sollte über dem Grenzwert des geforderten Überlebens liegen, also sprich: Bei den Genexpressionstests sagt man normalerweise 90 %. Das IQWiG hat sich dem internationalen Standard ja nicht angeschlossen. Man sagt normalerweise 90 % metastasenfrees Überleben nach 10 Jahren. In der Kaplan-Meier-Analyse sollte die untere Grenze des Konfidenzintervalls oberhalb dieses Wertes liegen. Ob die mediane Nachbeobachtungszeit dann 10 Jahre, 5 Jahre oder 3 Jahre ist, ist egal.

Moderator Jürgen Windeler: Okay, danke.

(Stefan Lange: Da habe ich eine Nachfrage!)

- Ja, bitte, Herr Lange.

Stefan Lange: Das habe ich nicht verstanden. Die untere Grenze muss oberhalb des Wertes liegen?

Moderator Jürgen Windeler: Wenn du die Überlebenszeit nimmst.

Stefan Lange: Für die Überlebenszeit. Aber dann muss doch die untere Grenze oberhalb der 90%-Freiheit liegen. So meinten sie das?

(Marion Kiechle: Genau!)

Moderator Jürgen Windeler: Ja, 10-Jahres-Überlebensrate: 90 %, und das Konfidenzintervall soll zwischen 90 und 100 liegen - auf Deutsch. Wir haben jetzt 2 unterschiedlich konkretisierte, aber ähnlich lautende ... - Ah, sorry, bitte.

Ralf Kronenwett: Zum Begriff „medianes Überleben“ sehe ich noch ein Problem, dass es keinen Standard für die Definition des medianen Überlebens gibt. Also: Nimmt man zensierte Patienten mit hinein? Was ist mit Patienten, die vorher verstorben sind? Da hat man überhaupt nicht die Chance, die weiter zu beobachten. Ist die tatsächliche Beobachtungszeit damit gemeint? Das ist oftmals auch nicht auch nicht definiert in Publikationen. Daher kann die mediane Überlebenszeit über die Definition letztendlich beeinflusst werden.

Moderator Jürgen Windeler: Herr Lange.

Stefan Lange: „Mediane Überlebenszeit“ ist relativ klar definiert. Was in der Tat im Augenblick jetzt wieder etwas schwimmt, ist die Frage: Welche Ereignisse gehen denn in die Schätzung ein? Ich bin jetzt etwas erstaunt, dass Sie verstorbene Patienten dort irgendwie nicht berücksichtigen wollen. Andererseits war vorher die Rede, dass wir uns nur die metastasenfrees Zeit anschauen wollen.

Wie ich das verstanden habe, geht es doch um die Frage des Rezidivs: ob nun lokal oder fern. Beides spielt ja eine Rolle, würde ich jetzt mal annehmen. Das heißt: Nicht nur Metastasen,

sondern auch die Lokalrezidive werden bedeutsam. Die aktuelle Auswertung der Genomics-Studie hat, glaube ich, auch noch die Zweitkarzinome oder die kontralateralen Karzinome mit in die Definition hineingenommen. Aber die Todesfälle gehören da, denke ich, schon irgendwie mit rein oder nicht? Das finde ich jetzt bemerkenswert, dass es jetzt so ein bisschen schwimmt.

Ralf Kronenwett: Natürlich gebe ich Ihnen vollkommen recht: Es geht um die Definition der medianen Überlebenszeit. Was ist, wenn ich Patientinnen nicht mehr nachbeobachten kann, weil sie vorher verstorben sind? Dazu gibt auch Literatur, die auch die Situation beschreibt, dass der Begriff „mediane Überlebenszeit“ nicht klar definiert ist. Daher - denke ich - sollten Sie, wenn Sie das schon fordern, dann auch die klare Definition der medianen Überlebenszeit mit hineinnehmen.

Das heißt: Wenn Sie zum Beispiel eine Kohorte 5 Jahre beobachtet haben, kann es sein, dass Sie im nächsten Jahr in Abhängigkeit davon, wie viele Patientinnen ein Rezidiv haben oder versterben oder aus dem Follow-up gehen, völlig unterschiedliche mediane Nachbeobachtungszeiten bekommen - abhängig davon, wie Sie den Begriff definieren. Das ist das, was ich meine.

(Stefan Lange: Wir reden über die mediane ... - Zuruf: Mikro!)

Stefan Lange: Entschuldigung. - Noch einmal: Sie fragen: Wie ist die mediane Nachbeobachtungszeit definiert?

(Ralf Kronenwett: Ja!)

Nicht die mediane Überlebenszeit, sondern die mediane Nachbeobachtungszeit.

(Ralf Kronenwett: Genau, ja!)

- Okay.

Moderator Jürgen Windeler: Jetzt ist erst Frau Mellert dran, dann kommen Sie an die Reihe.

Uta Mellert: Ich möchte darauf hinweisen, dass es mir um das Risiko des Relapses für Fernrezidive geht und dass die Ivana Sestak zum Beispiel publiziert hat, dass dieses Risiko, wie Herr Prof. Dietel schon angemerkt hat, über 5 Jahre hinaus steigt und deswegen eine Nachbeobachtungszeit über 10 Jahre sinnvoll ist. - Danke.

Moderator Jürgen Windeler: Danke. - Frau Kiechle.

Marion Kiechle: Ich möchte auch noch mal mehr oder weniger das Gleiche sagen und vielleicht noch etwas ergänzen. Es geht hier von der Definition her wirklich ums metastasenfremde

Überleben, denn die Fernmetastasen sind ja quoad vitam bedrohlich bei der Brustkrebsdiagnose, nicht das Lokalrezidiv.

(Manfred Dietel: Ja!)

Moderator Jürgen Windeler: Gut. - Ich habe auch hier jetzt weniger wegen bestehen bleibender unterschiedlicher Argumente, sondern wegen ziemlich kongruenter Argumente den Eindruck, dass wir den Tagesordnungspunkt abschließen können.

Ich fasse noch einmal zusammen, dass Sie sich im Wesentlichen dafür aussprechen, die 10 Jahre zu behalten als Orientierung, was die Zeitlinie angeht, aber den geforderten Anteil entweder auf irgendwie unter 50 % festzulegen oder in der Weise über das Konfidenzintervall zu definieren, was wahrscheinlich auf ein ähnliches Ergebnis hinausläuft. - Ja, bitte schön, Herr Langenfeld.

Matthias Langenfeld: Ich hatte, glaube ich, die erste Bemerkung dazu gemacht. Ich möchte nicht unbedingt die letzte dazu haben. Aber ich denke, dass die Fragestellung der Prädiktivität in der ursprünglichen Anfrage des G-BA einfach auf die Wirkung von Chemotherapie zielt und dass es in dem Patientenkollektiv, in dem wir uns hier befinden - ER positiv, HER-2/neu-negativ, maximal 3 befallene Lymphknoten -, keinen Hinweis dafür gibt, dass eine Chemotherapie den Verlauf über 5 Jahre hinaus beeinflussen kann.

Das heißt, wenn das IQWiG die Fragestellungen nach dem G-BA prüft, wäre es sicherlich zu bevorzugen, die 5 Jahre zu nutzen - mehr als die 10 Jahre. Denn die 10 Jahre stellen im Allgemeinen ein anderes Risiko dar als das, was mit Chemotherapie angegangen werden kann.

Moderator Jürgen Windeler: Gut, danke. - Wobei ich jetzt gerade von den Pathologen meinte, gelernt zu haben, dass das insofern nicht so viel Sinn macht, weil die Ereignisse, für die man sich interessiert, erst nach den 5 Jahren im Wesentlichen auftreten. Insofern macht es keinen Sinn, sozusagen die 5 Jahre anzuschauen und festzustellen, dass nichts - Gott sei Dank muss man für die Frauen sagen - passiert ist.

Matthias Langenfeld: Wenn ich dazu noch kurz Stellung nehmen könnte: Ja, es geht nicht um die Ereignisse. Es geht um die Frage der Beeinflussbarkeit der Ereignisse durch Chemotherapie, durch adjuvante Chemotherapie im frühen Stadium.

Moderator Jürgen Windeler: Noch einmal: Wenn da keine Ereignisse auftreten in 5 Jahren, können wir die Frage auch nicht beantworten. - Herr Dietel, bitte noch einmal zur Klarheit.

Manfred Dietel: Noch einmal zur Klärung: Es geht hier nicht um die Wirkung von Medikamenten. Das ist nicht Ziel dieser Zusammenkunft. Darüber kann man auch gerne noch einmal streiten, aber hier geht darum - ein Prognosemarker -, ob in einem bestimmten Zeitraum Rezidive und bei Fernmetastasen in aller Regel ... Es gibt nur wenige lebensbedrohliche Lokalrezidive. Diese Fernmetastasen, die das Entscheidende sind für die Prognose der

Patientin ... Das ist auch nicht das Lokalrezidiv, sondern die Fernmetastase. Daran stirbt man.

Diese treten gehäuft, deutlich gehäuft erst nach 5 Jahren auf. Deswegen das Argument noch einmal, den längeren Zeitraum beizubehalten.

Moderator Jürgen Windeler: Gut, dann nehmen wir dies als abschließende Bemerkung für den Tagesordnungspunkt und kommen jetzt zum Tagesordnungspunkt 3. - Stefan Lange.

4.3.4 Tagesordnungspunkt 3: Versorgungsstandard

Stefan Lange: Da würden wir gern die Firma Genomic Health ansprechen, die uns freundlicherweise eine Stellungnahme geschickt hat und darin auch auf den jetzigen Versorgungsstandard zu sprechen gekommen ist mit der nicht ganz richtigen Bemerkung, das sei auch Aufgabe des Berichtes, den zu evaluieren. Das stimmt nicht. Das steht auch im Auftrag nicht so drin, sondern er soll berücksichtigt werden. Das ist ein kleiner, aber doch wichtiger und feiner Unterschied. Da haben Sie völlig recht, dass wir das berücksichtigen sollten. Wir haben das bisher vielleicht auch noch nicht in der gebotenen Weise getan.

Mich würde an der Stelle interessieren, ob Sie eigentlich Daten dazu haben - das gilt natürlich auch für andere Tests, aber Sie hatten das angesprochen -, was der Oncotype gegenüber den standardmäßigen Prognosemarkern, die wir haben - angefangen von der Tumorgröße, Lymphknotenstatus, Grading, die üblichen Rezeptoren, also HER-2/neu und die Hormonrezeptoren, aber auch zum Beispiel Ki-67 -, hinaus noch für Informationen liefert. Mich würde interessieren, ob es dazu eigentlich Daten gibt.

Matthias Langenfeld: Danke für die Anfrage. - Ja, dazu gibt es Daten. Die wichtigsten sind sicherlich die vor Kurzem vorgestellten, zunächst einmal auf 3 Jahre beschränkten Daten der planB aus Deutschland, die zum Beispiel Ki-67 haben als Marker oder Grading und die einfach zeigen, dass in einer Patientenpopulation mit unterschiedlich hohem Ki-67 und einem niedrigen Recurrence Score das Ergebnis, also das Outcome der Patienten gut ist. Das heißt: In der Publikation wird so um die 2 % Risiko innerhalb von 3 Jahren ohne Chemotherapie betragen sein.

Es gibt weiterhin Untersuchungen aus retrospektiv-prospektiven durchgeführten Studien, in denen in multivariater Analyse die Bedeutung der klassischen Faktoren wie Tumorgröße, Grading, Alter der Patientin und der Recurrence Score durch die, wie gesagt, multivariate Analyse untersucht werden. Sie zeigen, dass der Recurrence Score jeweils unabhängig zu der Vorhersage des Outcomes beiträgt.

(Stefan Lange: Nachfragen?)

Moderator Jürgen Windeler: Bitte schön.

Stefan Lange: Die erste Antwort habe ich nicht ganz verstanden. Ich fange mit dem zweiten Teil an. Diese Analysen, die beinhalten aber nicht den Ki-67-Marker? Das ist dann so.

Ich meine: Wir sind uns ja einig. Zum Beispiel wissen wir ja, dass Sie so etwas Ähnliches wie HER-2 nur eben auf eine andere Weise bestimmen und Hormonrezeptor - das ist ja keine zusätzliche Information im Grunde genommen, sondern das Gleiche. Das wissen wir alle, also im Wesentlichen. Jetzt fangen wir nicht an, hier über Kleinigkeiten zu reden. Aber im Grunde genommen ist es das Gleiche, was die prognostische Wertigkeit angeht. Noch einmal: Gibt es für das, was jetzt oder auch in der Vergangenheit Gynäkologinnen und Gynäkologen, Onkologinnen und Onkologen ohne die genbasierten Tests an Methoden zur Entscheidungsfindung verwendet haben als Prognosemarker, dort einen Vergleich, was eben noch einmal der Oncotype darüber hinaus an Information, an prognostisch relevanter Information liefert, bitte inklusive Ki-67? Vielleicht können Sie das mit planB noch einmal erläutern. Das habe ich, wie gesagt, nicht verstanden.

Aber die anderen Analysen - zweiter Teil Ihrer Antwort - haben das Ki-67 als wichtigen Proliferationsmarker meines Wissens nicht berücksichtigt.

Moderator Jürgen Windeler: Herr Langenfeld noch einmal.

Matthias Langenfeld: Ich möchte die Frage oder die Ergebnisse insofern konkretisieren, als die Ki-67-Daten tatsächlich vor allem für die planB in Deutschland vorliegen. Der Marker ist nicht in allen Ländern so gebräuchlich wie in Deutschland. Nicht jede Analyse, in der zum Beispiel B14- oder B20-Studie hat ihn berücksichtigt.

Das heißt: Ich beziehe mich in meiner Antwort auf das Ki-67 vor allen Dingen auf die planB-Studie, in der systematisch erhoben worden ist - einmal lokal, einmal zentral, genau wie das Grading. Publierte Analysen liegen dazu vor, wie in der Niedrigrisikogruppe das Ki-67 gestreut ist. Da wir das Outcome nach 3 Jahren kennen, was, wie gesagt, in Kürze publiziert werden wird, zeigt es, dass zumindest hohes Ki-67 nicht immer mit einem schlechten Outcome vergesellschaftet sein muss - genauso wenig wie ein hohes Grading. Das liegt auch für die TAILORx vor.

Systematische Analysen im Rahmen von multivariaten Analysen gibt es zu vielen Studien, eigentlich in allen veröffentlichten Studien, in prospektiv-kontrollierten Studien, in der die traditionellen Parameter erhoben worden sind. Und die zeigen jedes Mal konklusiv, dass der Recurrence Score in dieser multivariaten Analyse statistisch signifikant mit einer Hazard Ratio versehen ist, die klinisch relevant ist, also knapp 2.

Moderator Jürgen Windeler: Ja, bitte.

Christoph Petry: Herr Lange, ich darf die Frage für den EndoPredict beantworten. Es gibt - das ist publiziert worden 2011 von Filipits und 2013 von Dubsky - eine Analyse, in der sowohl für den EndoPredict als auch für die anderen prognostischen Faktoren, und zwar sowohl

einzelnen als auch in Kombination, der C-Index berechnet worden ist. Sie kennen das. Das ist ein statistischer Deskriptor für die diagnostische Leistungsfähigkeit einer Methode. Und in dieser Analyse - das gilt sowohl für die ABCSG-6- als auch für die ABCSG-8-Studie, einzeln und in Kombination - fügte der EndoPredict, festgemacht über den C-Index, den anderen Methoden statistisch signifikante Informationen zu. Das war im Jahr 2011 tatsächlich das erste Mal, dass das für so einen Genexpressionstest gezeigt werden konnte.

Aber zu Ihrer Frage: „Sind die Daten da?“ - Ja, die Daten sind da.

Moderator Jürgen Windeler: Herr Dietel, bitte.

Manfred Dietel: Vielleicht noch eine kurze Ergänzung. Warum haben sich überhaupt Pathologen mit so etwas beschäftigt? Da ist unter anderem der Punkt, dass eben bei dem zweiten Test die klassischen morphologischen Parameter, Prognoseparameter wie Tumorgröße und Lymphknotenstatus, mit einbezogen werden in den Algorithmus. Wir haben uns immer überlegt, warum man nur auf Gene geht und nicht die Diskussion auch dahin führt - denn „4 Lymphknotenmetastasen“ war früher der eigentliche Goldstandard für die Prognosebewertung -, dass diese beiden sehr wichtigen Kriterien Tumorgröße und Lymphknotenstatus mit einbezogen werden.

Das war eigentlich die Verbindung zwischen Pathologie und den molekularen Assays, damit das wirklich zusammengeführt wird, was ich für wichtig halte.

Moderator Jürgen Windeler: Herr Lange.

Stefan Lange: Ich habe jetzt nicht verstanden, was Sie mit dem zweiten Test meinten. Meinten Sie den EndoPredict?

Manfred Dietel: Ja.

Stefan Lange: Ach so, okay.

Moderator Jürgen Windeler: 2 Wortmeldungen noch. Herr Sinn? Erst Herr Morawski und dann Herr Sinn.

Erwin Morawski: Herr Lange, vielen Dank. - Ja, ich glaube, wir haben zum großen Teil ähnliche Ansätze gehabt. Versorgungsstandard war für uns tatsächlich ... Sie haben es gerade konkretisiert - danke dafür -, dass es mit den bisherigen Standardparametern eben keine klare Aussage gibt, und zwar nicht allein in der Low-Risk-Gruppe, sondern in der definierten Gruppe, die ja intermediäre Patienten mit Low-Risk-, manchmal auch mit High-Risk-Anteil umfasst. Bis zu 3 befallene Lymphknoten haben wir, glaube ich, alle als Fokus.

Dazu gibt es, wie Herr Langenfeld ausgeführt hat, aus den prospektiven Studien Ergebnisse, die zeigen, was die klassischen Faktoren im Verhältnis zum Oncotype aussagen können oder was eben nicht. Und dazu gibt es auch - ich denke, das ist vielleicht ganz relevant, weil Sie

gerade auch gesagt haben, Sie betrachten die Diagnostik so ähnlich wie ein solches Produkt, also ähnlich wie ein Pharmazeutikum - die Forderung, dass man auch hier breite Versorgungsforschung vielleicht mit berücksichtigen kann. Da gibt es jetzt 2 ganz neue Studien, die veröffentlicht wurden: einmal aus dem israelischen Umfeld mit knapp über 2 000 Patienten, die 5 Jahre nachverfolgt wurden, nachdem sie zuerst mit dem Oncotype klassifiziert wurden, und jetzt in San Antonio publiziert 44 600 Patienten vom National Cancer Institute der USA, in der 40 000 Patienten node-negative und 4 600 Patienten node-positive nachverfolgt wurden - auch über 5 Jahre. Da geht auch zusätzlich noch eine Analyse auf das Grading ein. Also: Hier sieht man auch, dass Grading-1, 2, 3, dass Alter und dass Tumorgröße keine so klare Aussage treffen können, wie es der Oncotype kann - das noch, wenn man sich die Ergebnisse anschaut, 5 Jahre nachbeobachtet mit einem sehr niedrigen Rezidiv- respektive Mortalitätsrisiko.

Also ich denke: Versorgungsstandard a) retrospektiv-prospektiv, b) prospektiv, was ja auch bei Simon und Hayes als die nächsthöhere und wichtigere Stufe betrachtet wird, und dann noch Praxisrealität, Versorgungsforschung im besten Sinne aus 2 relativ großen Kohorten, insgesamt, wie gerade ausgeführt, über 46 000 Patienten.

Moderator Jürgen Windeler: Herr Sinn, bitte.

Peter Sinn: Die Frage wurde gestellt, ob man nicht mithilfe von Ki-67 Ähnliches erreichen kann wie mit den molekularen Prognosetests. Das ist nicht der Fall. Hier geht es um Versorgungsstandards, und Ki-67 gehört zum Versorgungsstandard. Die Definition eines Mammakarzinoms mit intermediärem Risiko bezieht ja zumindest in unserer Praxis Ki-67 bereits ein, sodass die Tumoren mit einem intermediären Ki-67 - das ist die Mehrzahl dieser Tumoren - dann tatsächlich die Kandidaten sind für die Indikation dieses Tests. Da hilft eben Ki-67 nicht weiter.

Moderator Jürgen Windeler: Herr Lange, dann Frau Mellert.

Stefan Lange: Welches sind jetzt die Kandidaten? Was für einen Ki-67-Wert haben die?

(Fülöp Scheibler: Intermediär!)

Peter Sinn: Das ist intermediär. Das heißt: Unsere Onkologen würden ein Mammakarzinom mit einem hohen Ki-67 von - sagen wir einmal - 50 % und diesen übrigen genannten tumorbiologischen Faktoren nicht einem Genexpressionstest unterwerfen, genauso wenig wie ein Mammakarzinom mit - sagen wir - einem Ki-67 von 5 %, aber wohl durchaus schon mit einem intermediären Ki-67 von - sagen wir - 20 % oder 25 %.

Marion Kiechle: Herr Sinn, es tut mir leid. Aber ich muss jetzt hier mal hineingrätschen.

(Peter Sinn: Bitte!)

Wenn man sich die Publikation anschaut von Filipits, sich die Überlebenskurven von Frauen mit einem niedrigen Ki-67 anschaut und dann den EndoPredict draufsetzt, sieht man, dass man da sehr wohl mit dem EndoPredict Frauen herausfischen kann, die trotz des niedrigen Ki-67 ein schlechtes Überleben haben und umgekehrt auch.

Das bedeutet nämlich andersherum gesagt: Das trifft nach den Daten von Oncotype und den Daten von EndoPredict, die uns zur Verfügung stehen, zu, dass diese Tests den herkömmlichen prognostischen Kriterien überlegen und unabhängig vom Ki-67 sind.

Zugegeben: Es gibt eine große Überschneidung der Mengen, aber in der multivariaten Analyse sind die besser.

Moderator Jürgen Windeler: Ist die Frage ausreichend beantwortet?

(Zustimmung von Stefan Lange)

- Gut. - Dann Frau Mellert.

Uta Mellert: Ich möchte noch einmal auf die anfängliche Frage zurückkommen, was der Added Value von den Gensignaturtests ist und dann auf Prosigna zu sprechen kommen. Es ist Ihnen sicherlich nicht entgangen, dass Prosigna in dem Patientenreport die Subtypen aufzeigt: Luminal A, Luminal B, basal-like und HER-2-enriched.

Das ist insofern ein Added Value, als der Art zusätzliche Informationen darüber bekommt, wie sich die Tumorbilogie zusammensetzt und innerhalb einer immer noch bestehenden Intermediärgruppe Fragen beantworten kann, was das denn für den Patienten für Konsequenzen hat. Aus der Erfahrung mit diesem Test kann ich zusätzlich sagen, dass wir auch unter Low-Risk-Patienten interessante Erkenntnisse gewinnen, wenn wir diese Tumortypen erfahren können.

Zum anderen ist publiziert, dass es - es gab eine Untersuchung gegenüber den Immunhistochemieparametern - also definitiv einen Added Value hat gegenüber diesen Parametern, dass auch hier der Lymphknotenstatus und die Tumorgöße einfließen und dass wir zusätzlich eben hier besser entscheiden können, welcher Risikostatus für die Patientin besteht.

Moderator Jürgen Windeler: Weitere Wortmeldungen sehe ich jetzt zu diesem Punkt nicht. Dann können wir zum Punkt „Verschiedenes“ übergehen.

4.3.5 Tagesordnungspunkt 4: Verschiedenes

Moderator Jürgen Windeler: Wer hat den Wunsch, dazu etwas zu sagen? - Herr Dietel, Sie machten schon entsprechende Bewegungen, bitte schön.

Manfred Dietel: Danke schön. - Ich würde gerne einen allgemeinen Punkt etwas aus der Sicht der Pathologie ansprechen, und zwar dahingehend, dass wir er uns wirklich mit der

Versorgungsrealität - nicht nur wir, sondern auch Sie - beschäftigen sollten. Ich glaube, es ist wichtig, dass Kriterien wie Verlässlichkeit, Lebensqualität etc. mit in den ganzen Prozess zumindest eingedacht werden.

Da kann ich Ihnen nur, um den Hintergrund meiner Ausführungen klarzumachen, sagen, dass wir im Jahre 2014 in meinem Institut 120 Assays gemacht haben und im Jahre 2015 520. Das heißt: Der klinische Bedarf, die Fragestellung der Kliniker sowohl in den eigentlichen Kliniken, aber ganz deutlich auch im niedergelassenen Bereich ist massiv. Es ist wirklich eine sehr verantwortungsvolle Aufgabe, die Sie übernommen haben, hierzu eine Stellungnahme abzugeben, die sich nicht nur meines Erachtens 5 % rauf, 5 % runter, Konfidenzintervall - was weiß ich: so oder so - ..., sondern die die gesamte Situation, die Versorgungssituation, die Lebensqualität und die Realität, in der wir uns befinden in der täglichen Routine, mitbetrachten sollte. Wir haben jahrelang gemeinsam - Kliniker und Pathologen - versucht, vernünftige und verlässliche Prognosekriterien zusammenzustellen. Wir sind jahrelang nicht wirklich weitergekommen. Der einzige Schritt jetzt wirklich nach vorne - das ist unabhängig von den Assays; ich möchte mich da auch betont ganz allgemein ausdrücken - sind diese molekularen Tests, die wir ja übrigens auch für viele andere Tumoren mittlerweile routinemäßig durchführen.

Ich wollte dies nochmal sozusagen als Rahmen des Ganzen - deswegen habe ich es auch zum Schluss gesagt - anmerken, weil es mir wirklich ein Anliegen ist, dass hier sorgfältig überlegt werden muss, ob nicht Patientinnen eine Therapie zugemutet wird, wenn die Beurteilung so bleibt, wie sie im Augenblick ist, die sie nicht benötigen. Das ist ein Vorgang, der sicherlich abgewogen werden muss, der aber doch in der Gesamtsituation - finde ich; von der Lebensqualität her hatte ich das schon gesagt - mitberücksichtigt werden sollte.

Moderator Jürgen Windeler: Vielen Dank. - Ich habe 2 Wortmeldungen, Frau Kiechle und Herrn Sauerland. Ich würde aber gerne an dieses Statement, dem ich mich im Grundsatz natürlich - nicht nur bei diesem Thema, bei jedem anderen Thema auch - anschließen kann, was immer das dann für Konsequenzen hat, eine Frage anschließen, denn das passt wunderbar hierzu. Ich habe vor ein paar Tagen - leider jetzt erst - eine Arbeit gelesen, die noch nicht so alt ist - ich glaube, sie wurde 2012/2013 publiziert -, die aus deutschen Zentren - ich glaube, Heidelberg war dabei, aber ich glaube, Heidelberg war nicht alleine - beschrieben hat, was denn nun aus den Patientinnen wird, die diese Tests angeboten bekommen haben und die auch angewendet worden sind. Die Arbeit hat mich etwas Schweiß gekostet, denn ich fand die Darstellung nicht so unmittelbar eindrücklich. Aber was ich meinte, verstanden zu haben aus der Arbeit, ist, dass die Therapieentscheidung für eine Chemotherapie nach einem niedrigen Risikoscore in einem Drittel der Fälle geändert worden ist.

(Manfred Dietel: In welche Richtung?)

- In Richtung, auf die Chemotherapie zu verzichten.

(Manfred Dietel: Ja!)

Das bedeutet umgekehrt in dieser Erhebung aus Deutschland, dass bei zwei Dritteln der Frauen die Chemotherapie trotz eines niedrigen Risikoscores weiter empfohlen worden ist - wohlgemerkt; da waren die Frauen noch nicht beteiligt - und von den Frauen, soweit ich das der Arbeit weiter entnehmen kann, auch weitestgehend dann so umgesetzt worden ist.

Jetzt würde ich natürlich die Frage erst stellen wollen, ob das auch Ihrer Erfahrung entspricht. In dieser Arbeit ist auf eine ganze Reihe anderer Studien verwiesen worden - auch aus dem internationalen Bereich -, die sich alle im Bereich zwischen 20 % und 30 % Änderungen der Therapieentscheidung in Richtung auf Verzicht auf eine Chemotherapie bewegt haben.

Wenn ich dies jetzt als Ergebnis aus der Praxis und Versorgungsforschung betrachte, muss ich natürlich sagen: Dann scheinen mir die Appelle, Frauen eine nicht benötigte Chemotherapie zu ersparen, auch noch in ganz andere Richtungen angebracht zu sein als an unsere Bewertung.

Frau Kiechle hatte sich gemeldet und dann ...

Marion Kiechle: Also dann vielleicht einmal aus Sicht einer Ärztin, die vor der Patientin sitzt und sie beraten soll, wie es weitergeht nach der Brustkrebsoperation oder sogar schon vorher. Ganz ehrlich: Ich bin sehr dankbar, dass wir diese Tests haben. Natürlich - ich komme von der Universitätsklinik - haben wir das auch wissenschaftlich begleitet und werden demnächst unsere eigenen Daten publizieren; die sind jetzt zusammengeschrieben. Wir verfügen jetzt über knapp 400 Frauen mit primärem Mammakarzinom, bei denen wir mit den herkömmlichen prognostischen Kriterien nicht weitergekommen sind. Klar davon ausgeschlossen sind die HER2-Positiven. Da wissen wir: Die brauchen eine Chemo. Oder auch die Tripelnegativen. Wenn so eine Patientin vor mir sitzt, brauche ich nicht lange nachzudenken. Andersherum genauso: Wenn da jetzt eine Frau, 85, G1 ... Darüber brauche ich auch nicht nachzudenken. Die kriegt natürlich keine Chemotherapieempfehlung.

Aber die große Gruppe dazwischen fragt mich natürlich: Was sollen wir machen? Hormonsensibles Mammakarzinom. Ich bin dankbar, dass wir die Tests haben. Es ist so: Unsere eigenen Daten zeigen uns, dass die Anwendung dieses Tests bei 50 % der Patienten tatsächlich greift. Ohne diese Tests hätten wir gut zwei Dritteln dieser Frauen eine Chemotherapie - aus dieser Gruppe jetzt, wohl gemerkt - empfohlen. Mit Anwendung der Tests liegt der Wert bei 32 %.

Ähnliche Daten gibt es international aus Spanien, wobei die Diskrepanz nicht ganz so hoch war. Die Spanier waren etwas - sagen wir einmal - salopper. Wir handeln natürlich im Sinne „in dubio pro reo“. Wenn wir es nicht wissen, empfehlen wir - wir in Deutschland haben eher so ein Sicherheitsdenken - naturgemäß eher den Patientinnen eine Chemotherapie, nach dem Motto: Mach mal besser eine, vielleicht nützt das ja doch etwas. - Andere Population sind

etwas - sagen wir mal - großzügiger. Aber trotzdem: Es gibt die Daten auch international, dass diese Tests die klinische Entscheidung deutlich beeinflussen.

Moderator Jürgen Windeler: Danke. - Herr Morawski.

Erwin Morawski: Herr Windeler, zu Ihrer Frage: Wie ist tatsächlich die „Therapietreue“? 2 Studien dazu: zum einen die deutsche planB, also die prospektive Studie im High-Risk-Kollektiv. Da wurde der Low-Risk-Gruppe, 88 %, keine Chemotherapie verabreicht, nur 12 %. Zu der damaligen Zeit war es untypisch, in diesem Patientenkollektiv die Frage zu stellen, also sprich: Selbst da, in einem potenziellen High-Risk-Kollektiv, haben sich nur 12 % für die Chemo entschieden.

Wenn man jetzt noch einmal auf die gerade veröffentlichten Versorgungsforschungen geht, also auf die beiden großen Studien, die ich gerade angesprochen habe, sieht man: Gerade bei Clalit wird darauf sehr deutlich eingegangen. Die Low-Risk-Gruppe hat 1 % Chemotherapie tatsächlich bekommen, also: Die Therapietreue mit dem Oncotype - darum ist uns das auch so wichtig, die Versorgungsforschung und die prospektiven Studien nach vorne zu heben - gibt die Sicherheit für die Patientin, dass sie da auch vertrauen kann und dass dem auch gefolgt wird.

Moderator Jürgen Windeler: Herr Sauerland.

Stefan Sauerland: Es sind ein paar Parameter in den Raum geworfen worden, die uns natürlich immer mal wieder begegnen. Das eine war die Versorgungssituation. Da ging es primär - so habe ich es verstanden - um die Anwendernachfrage. Das kann natürlich erst einmal kein Kriterium sein.

Das zweite Kriterium haben Sie jetzt genannt: die Anzahl der Patientinnen, die am Ende einen Therapiewechsel bekommen. In der Tat geht es natürlich, ohne dass das einen Effekt auf die Therapie hat, nicht, dass diese Testung einen Nutzen entwickelt. Aber umgekehrt ist es eben kein hinreichendes Kriterium, um darüber Nutzen abzuleiten.

Das dritte Kriterium, das jetzt von Herrn Dietel genannt wurde, war die Lebensqualität. Da würde ich gern nachfragen. Natürlich ist Lebensqualität sehr wichtig. Aber Sie haben nicht gesagt, zu welchem Zeitpunkt die erhoben werden sollte. Denn es ist natürlich klar: Wenn ich der Frau sagen kann: „Da ist nichts. Das Risiko ist minimal. Sie brauchen keine Chemotherapie.“, steigt spätestens nach der Chemotherapie, aber wahrscheinlich schon bei der Mitteilung der guten Nachricht die Lebensqualität. Aber das ist nicht der Sinn der Sache. Ich möchte Sie nur noch einmal bitten, zu klären, dass wir doch eigentlich mit der Lebensqualität alleine nichts anfangen können. Wir brauchen schon eine onkologische Untersuchung, die auf Langzeitdaten fußt, oder?

Manfred Dietel: Dagegen habe ich ja nichts. Aber ich wollte diesen zweiten Gesichtspunkt der Lebensqualität und der Versorgungsrealität schon noch einmal betonen. Es geht ja nicht

darum, dass wir die Patientinnen und Patienten, wobei ich persönlich das gar nicht tue - nicht, weil ich Pathologe bin, aber ich kenne das von unseren Onkologen -, zu irgendetwas überzeugen wollen, sondern es geht um eine sachliche Diskussion zwischen Patientin und Arzt. Wenn die Patientin sagt: Ich will auf gar keinen Fall eine Chemotherapie haben, auch wenn sie High Risk ist von den üblichen Parametern, und man jetzt den Assay hat - egal welchen, wie gesagt - und er eindeutig anzeigt, dass es notwendig ist - nach diesem molekularen Test, inkludiert man die morphologischen Kriterien -, ist die Diskussion zwischen Ärztin oder Arzt und Patientin natürlich eine völlig andere und auch die Überzeugungskraft, die Patienten doch zu einer in diesem Falle notwendigen Chemotherapie zu überreden, ist deutlich höher, und vice versa. Wenn eine Patientin sagt: „Ich möchte aber sicher sein, 100 % und Pipapo“ und man dann sagt: „Passen Sie auf, der Test zeigt eindeutig: Es ist ein Low Risk, vielleicht sogar ein sehr niedriges Risiko“, kann man ganz anders mit der Patientin diskutieren. Das kann vielleicht noch einmal ein Kliniker oder eine Klinikerin kommentieren. So ist es mir zumindest von unseren Kollegen immer wieder berichtet worden.

Eva Schumacher-Wulf: Ich habe noch eine Frage zu der eben zitierten Studie, an der Heidelberg beteiligt war. Könnten Sie nochmals sagen, wo Sie die Studie gelesen haben - zum Nachlesen? Habe ich es richtig verstanden, dass Patienten getestet wurden und dass innerhalb der Low-Risk-Gruppe trotzdem noch zwei Dritteln eine Chemotherapie empfohlen wurde?

Moderator Jürgen Windeler: Das war die Tabelle, die ich da gelesen habe. Können Sie die Quelle sagen?

Stefan Sauerland: Das ist die Arbeit von Eiermann et al., Annals of Oncology, 2013. Das ist, glaube ich, bekannt.

Moderator Jürgen Windeler: Herr Schmidt kommentiert die Arbeit oder meine Interpretation.

Marcus Schmidt: Kurz dazu: Klar, die Arbeiten interpretieren wir nicht genau, weil man nicht genau weiß, um welche Arbeiten es sich handelt. Aber ich dachte mir schon, dass es um Eiermann geht. Es gibt auch eine vergleichbare Arbeit - ich weiß das, weil ich als Koautor Man-in-the-Middle war - aus der Charité, die Ähnliches gezeigt hat.

Letzten Endes ist es der Transfer oder das Umwidmen von etwa einem Drittel, 35 % bis 38 %, von der einen Risikogruppe - jetzt klinisch-pathologisch - Low Risk zu High Risk und vice versa. Das ist der Punkt. Aber in diesen Arbeiten war meiner Erinnerung nach nicht ausgeführt, ob die Patienten diese Therapie dann auch wahrgenommen haben. Das wäre natürlich der nächste relevante Schritt. Da bin ich natürlich ganz auf der Seite von dem, was die Vorredner schon genannt haben. Jetzt eher von meiner klinischen Perspektive: Wenn ich mich mit der Patientin im Rahmen der partizipativen Entscheidungsfindung unterhalte, die wir immer durchführen, wenn wir uns über Chemotherapie - ja oder nein - unterhalten, benutze ich natürlich am liebsten die Parameter, die die beste derzeit erhältliche Aussagekraft

haben. Daher ist es bei vielen Patientinnen real hilfreich, solche Genexpressionstests aufgrund der überlegenen Aussagekraft einsetzen zu können. - Das sage ich jetzt wirklich mit der klinischen Brille, danke.

Moderator Jürgen Windeler: Gut, also die Quelle haben wir geklärt. Die Interpretation war tatsächlich ein Drittel von Chemotherapie initial zu nicht mehr Chemotherapie danach. - Die Arbeit lässt sich über die Adhärenz, sagt man modernerweise, der Frauen zu dieser Therapieempfehlung etwas ungenau aus. Es sind insgesamt nur sehr, sehr wenige Frauen, die der ursprünglichen Empfehlung nicht gefolgt sind. Dazu gibt es leider keine Tabelle. Nach dem, was ich da entnehmen kann, ist es so, dass die meisten Frauen insbesondere dieser Empfehlung, keine Chemotherapie trotz des Low-Risk-Scores zu benutzen, nicht widersprochen und es genauso gemacht haben. So ist es in der Arbeit angegeben. Aber wie gesagt, wer das nachlesen will: Eiermann et al., 2013. - So, Frau Kiechle.

Marion Kiechle: Sorry, da muss ich ... Entweder habe ich da etwas falsch verstanden, oder ... In der klinischen Versorgung ist es genau umgekehrt. In dieser besagten Gruppe, hormonsensitives Mammakarzinom, HER-2/neu-negativ, kommen die Frauen auf uns zu bzw. da ist die Frage: Chemotherapie - ja oder nein? Die Tests dienen in dieser Situation in erster Linie dazu, festzustellen: Wie hoch ist es Risiko, dass die Frau Fernmetastasen entwickelt und daran stirbt? Wenn sie ein niedriges Risiko hat - das ist untersucht mit einer rein endokrinen Therapie -, können wir anhand der jetzigen Studienlage eine Prognose ableiten. Wenn die Prognose sehr gut ist, das Risiko innerhalb von 10 Jahren Metastasen zu entwickeln, nicht höher ist als 5 %, können sagen: Liebe Frau Meier/Müller/XY, machen Sie eine rein endokrin adjuvante Therapie. Sie können auf eine Chemotherapie verzichten.

Ganz ehrlich: Keine einzige Frau hat sich bislang gegen dieses Urteil gewehrt, denn die meisten Frauen hoffen auf so ein Ergebnis. Sie sind ja froh, wenn sie keine Chemotherapie machen müssen. Das ist die hauptsächliche klinische Fragestellung.

Umgekehrt: Wenn eine Frau kommt und von vornherein in dieser Situation sagt: „Frau Professor, ich mache auf gar keinen Fall eine Chemotherapie. Da können Sie mir kommen, womit Sie wollen“, mache ich den Test auch nicht, denn - das muss ich dann ganz ehrlich sagen - , das ist dann meine Konsequenz: Warum soll ich dann die Krankenkasse damit belasten oder belästigen, wenn die Frau von vornherein dagegen ist.

Okay, ich sage ihr: „Frau Meier, es gibt diese Möglichkeit“ - „Nein, ich mache es nicht.“ - Gut, dann kann ich darauf verzichten. Das ist der klinische Alltag.

Moderator Jürgen Windeler: Vielen Dank. - Wobei Sie natürlich verstehen werden, dass uns solche niedergelegten Erfahrungen aus einer Reihe von deutschen Zentren dann auch durchaus interessieren und auch durchaus beeindrucken. - Dann kommen Sie bitte noch einmal dran, dann Sandra Janatzek.

Eva Schumacher-Wulf: Der klinische Alltag wurde dargestellt. Der Alltag in Patientenkreisen sieht so aus, dass Patientinnen seit Jahren wissen, dass es da eine Möglichkeit gibt zu testen, dass sie auch wissen, dass in vielen anderen Ländern seit Langem die Kostenübernahme geregelt ist. Es gibt überhaupt keine Diskussion. Jetzt auch in San Antonio - egal, bei welchen internationalen Kongressen - wird nicht diskutiert, ob das ein Nutzen ist oder nicht, sondern da werden vielleicht Tests verglichen oder wie auch immer. Aber dass ein Nutzen da ist, wird überhaupt nicht mehr diskutiert.

Dass das Ganze jetzt in einer Pressemeldung so dargestellt wurde vonseiten des IQWiG, als wären da offensichtlich einige Experten einer PR-Kampagne aufgesessen, fand ich eine krasse Äußerung, um es einmal so zu sagen. Das heißt ja nicht nur, dass die Experten in Deutschland einer Kampagne aufsitzen, sondern weltweit. Das sorgt für großen Frust bei Patientinnen.

Moderator Jürgen Windeler: Frau Sandra Janatzek.

Sandra Janatzek: Ich habe jenseits dieser Diskussion hier noch einen Nachfrage. Ist dafür jetzt der Zeitpunkt?

Moderator Jürgen Windeler: Ja.

Sandra Janatzek: Wir haben unter TOP 2 über die Nachbeobachtungszeiten für die Prognosestudien gesprochen. Meine Frage ist: Wie wird eigentlich begründet, dass für die randomisierten Studien eine Nachbeobachtungszeit von 1 Jahr ausreichend ist?

Moderator Jürgen Windeler: Frau Rutjes.

Anne Rutjes: Wir würden nicht sagen: Es ist ausreichend. - Wir haben erlaubt, dass es mindestens 1 Jahr sein muss, um etwas über das Schadenprofil sagen zu können. Aber für die Effektivität ist es absolut nicht ausreichend.

Moderator Jürgen Windeler: Frau Kiechle.

Marion Kiechle: 1 Jahr auch bei einer prospektiven Studie reicht definitiv nicht aus. Da möchte ich noch einmal auf das zurückkommen, was Herr Dietel auch gesagt hat, da ja beim Mammakarzinom gerade in dieser Gruppe die Rezidive bzw. Fernmetastasen jenseits der 5 Jahre auftreten. Die Ereignisse letztendlich, auf die wir schauen ... Eine Nachbeobachtungszeit auch bei prospektiven Studien von 1 Jahr ist zu wenig.

Vielleicht mag es bei anderen Karzinomen wie Lunge oder Pankreas vielleicht ausreichend sein, aber bei Brustkrebs nicht in dieser Situation.

Moderator Jürgen Windeler: Herr Scheibler.

Fülöp Scheibler: Dürfte ich dann bei der Gelegenheit nachfragen: Was wäre denn dann aus Ihrer Sicht für die randomisierten, prospektiv-geplanten Studien eine ausreichende Mindest-nachbeobachtungszeit?

Marion Kiechle: Eine Nachbeobachtungszeit über 10 Jahre wäre wünschenswert, wobei Sie dann schon auch auf die Studiengröße und die Konfidenzintervalle schauen müssen.

Moderator Jürgen Windeler: Herr Langenfeld.

Matthias Langenfeld: In den Studien - Herr Windeler, Sie hatten jetzt auch darauf hingewiesen, dass es durchaus Gleichheit gibt zwischen den pharmazeutischen Produkten und den Biomarkern - bzw. in den Berichten zu den Interventionsstudien reicht ein Zeitraum von 5 Jahren im Allgemeinen aus. Das würde ich gerne anmerken.

Moderator Jürgen Windeler: Um an der Stelle die Kongruenz herzustellen: Wenn wir uns für sehr seltene und sich spät entwickelnde Ereignisse interessieren - ich erinnere einmal an eine Studie, die in Deutschland für das frühe Prostatakarzinom gemacht wird -, wird man lange schauen müssen. Das liegt in der Natur der Sache. Wenn man sich für seltene Ereignisse interessiert, sind in der Regel die Effekte klar. Da wird man noch länger schauen müssen. An diesen Stellen etwas herauszubekommen, ist leider relativ schwierig und insbesondere langwierig.

(Zustimmung von Manfred Dietel, Matthias Langenfeld und Erwin Morawski)

Herr Langenfeld.

Matthias Langenfeld: Danke noch einmal. - Ich weise deshalb darauf hin, weil in einem der Verfahren - ich glaube, es war Pertuzumab, und es geht zumindest um die Indikation Brustkrebs und um die Indikation Chemotherapie plus - auch der Zeitraum von 5 Jahren gewählt worden ist. Wir befinden uns zumindest in der gleichen grundsätzlichen Indikation.

Moderator Jürgen Windeler: Jedenfalls wenn wir über das gleiche Stadium und die gleiche Prognose des Brustkrebses sprechen, wobei ich nicht sicher bin, dass das so ist.

Marion Kiechle: Entschuldigen Sie, wenn ich mich da einmische, aber wenn eine Patientin Pertuzumab bekommt, ist sie HER-2/neu-positiv. Über diese Entität diskutieren wir hier nicht.

Moderator Jürgen Windeler: Ich wollte es nicht so deutlich sagen, aber so ähnlich hätte ich es auch verstanden, ja.

(Teilweise Heiterkeit)

Gut, gibt es weitere Wortmeldungen? - Bitte schön, Frau Mellert.

Uta Mellert: Vielleicht möchte ich noch einen Ausblick geben zum Tagesordnungspunkt 3. Wenn wir den Wert dieser Tests bemessen wollen und in die Zukunft schauen und sehen, ob wir ... Jetzt reden wir von einer bestimmten Patientengruppe, und nachher reden wir von tripelnegativen Patientinnen. Wir reden immer mehr von Subgruppen und immer mehr von individualisierten Therapien. Und wir werden immer mehr Hilfe brauchen, um wirklich Patientinnen zielgerichtet therapieren zu können. Dann sind diese Tests in ihrer Reife noch weiterentwicklungsfähig. Das ist ja das, was jetzt auch passiert. Dann werden wir immer mehr die Diskussion bekommen und immer mehr dankbar sein, dass wir solche Tests haben, weil wir immer besser Subgruppen bestimmen können anhand der molekularen Daten, die wir von den Gentests bekommen. - Danke schön, nur um die Wichtigkeit beizumessen, dass wir hier am Anfang sind.

Moderator Jürgen Windeler: Was man aus unserer Perspektive interpretieren kann, ist, dass man vieles nicht weiß. - Herr Lange.

Stefan Lange: Ich würde gern die Gelegenheit nutzen, noch einmal eine Frage spezifisch an die Klinikerinnen und Kliniker im Raum zu richten. Ich habe letztes einen Vortrag gehört - es ging auch um das Thema -, in dem die Vortragende sehr eindringlich davor gewarnt hat, 2 dieser neuen Tests zu machen, weil das die Patientinnen verunsichern könnte. Jetzt würde mich interessieren - ich muss gestehen, ich bin da etwas unbeleckt, was die Leitlinien angeht; ich bin ja auch nicht in dem Projekt; deswegen spreche ich jetzt nicht fürs IQWiG, sondern für mich, was Sie denn jetzt eigentlich empfehlen. Welchen denn jetzt eigentlich? EndoPredict? Oncotype? Prosigna? Dann gibt es noch den uPA / PAI-1. Was empfehlen Sie denn jetzt? Und was sagen Sie einer Frau, die beispielsweise EndoPredict gemacht hat mit einem vielleicht nicht so wünschenswerten Ergebnis und sagt: Ich möchte jetzt aber doch noch mal den Oncotype machen? Oder umgekehrt. Was sagen Sie denen?

Moderator Jürgen Windeler: Herr Schmidt hat eine Antwort.

Marcus Schmidt: Das ist natürlich eine klinisch absolut relevante Frage. Zumindest das Statement der AGO ist ganz klar: Man sollte letztendlich nur einen dieser molekularbiologischen Tests machen. Welchen Test macht man? Ich denke, da verrate ich kein Geheimnis, wen auch immer Sie hier fragen: Es gibt Argumente für den EndoPredict. Es gibt Argumente für den Oncotype. Es gibt Argumente für Prosigna. - Das ist keine Frage. Aber um die Patienten nicht über die Maßen zu verwirren, weil eben diese Tests nicht hundertprozentig kongruent sind, ist die ganz klare therapeutische Empfehlung oder diagnostische Empfehlung, kein Testhopping zu machen. - Danke.

Moderator Jürgen Windeler: Sandra.

Sandra Janatzek: Aus unserer Sicht löst sich das Problem nur auf, wenn wir einmal einen validen Vergleich zur bisherigen Therapieentscheidung haben. Danach können sich ja dann

für diesen Biomarker auch Vergleiche zwischen verschiedenen biomarkerbasierten Entscheidungsstrategien anschließen.

Moderator Jürgen Windeler: Frau Kiechle.

Marion Kiechle: Um vielleicht etwas Salomonisches dazu zu sagen: Ich habe ja schon von diesen 6 Studien gesprochen, die ich da immer aufführe, wo die Tests ziemlich deckungsgleich sind in der in der Sortierung der Low-Risk-Gruppe, die keine Chemo braucht. Da sind die ziemlich kongruent. Das ist eigentlich auch die wichtigste Entscheidung, die wir erst einmal haben wollen für unsere Patientinnen in der Anwendung dieser Tests.

Es gibt dann schon feine Unterschiede. Es gab jetzt auch Publikationen dazu in San Antonio. Aber das bringt jetzt hier für die Diskussion oder für die für Ihre Entscheidung nicht sehr viel. Der eine Test ist bei Nodal-positiven besser, der andere ist da nicht so gut. Schauen Sie sich die Sachen an.

(Manfred Dietel: 5 Jahre, 10 Jahre!)

Aber bei der Einsortierung in die Low-Risk-Gruppe sind die ziemlich gleich. Das finde ich jetzt erst einmal aus klinischer Sicht das Wichtigste.

Moderator Jürgen Windeler: Ich habe jetzt Herrn Langenfeld, dann Herrn Lange und Frau Mellert.

Matthias Langenfeld: Es gibt kleine Serien an gut kontrollierten Ergebnissen, wo die gleichen Proben der gleichen Patienten zum Beispiel mit dem Recurrence Score und dann auch mit dem EndoPredict verglichen werden. Die Korrelation zwischen den Tests ist relativ nachvollziehbar und nie über 0,7/ 0,8. Der Irrtum, den man mit dem einen Test versus den anderen begehen kann, liegt in der Größenordnung von 20 %, was aus unserer Sicht durchaus eine Differenzierung notwendig macht.

Zum Zweiten ist die Gruppengröße, wenn man sich Studien anschaut, die 2 oder mehr Tests miteinander vergleichen, häufig unterschiedlich, also nicht nur qualitativ, sondern auch quantitativ werden Patienten grundsätzlich ähnlich, aber nicht übereinstimmend in jedem Falle einsortiert.

Moderator Jürgen Windeler: Stefan Lange.

Stefan Lange: Also erstens würde ich Frau Langenfeld gerne bitten, dass sie uns diese 6 Studien explizit benennt, damit wir auf ihren Schatz sozusagen auch zugreifen können.

Moderator Jürgen Windeler: Frau Kiechle war jetzt angesprochen.

Stefan Lange: Bitte?

Moderator Jürgen Windeler: Frau Kiechle.

Stefan Lange: Frau Kiechle. - Entschuldigung! Sorry, ich war gerade in Gedanken.

Dann würde ich gern Frau Kiechle antworten, dass Sie sagen, der eine ist da besser, der andere ist dort besser. - Ich weiß, ehrlich gesagt, gar nicht, was Sie jetzt mit „besser“ meinen. Denn das ist genau unser Problem. „Besser“ würde bedeuten: Wir wüssten, dass es auch richtig ist, was die Tests dort prognostizieren, nämlich dass betroffene Frauen im Niedrigrisikobereich keine Chemotherapie benötigen und umgekehrt.

Dann tatsächlich jetzt auch zu Herrn Langenfeld - sorry -: Ich bin jetzt etwas erschüttert, wenn Sie vom Korrelationskoeffizienten sprechen. Sie haben gesagt: Korrelationskoeffizient 0,7 bis 0,8. Das Quadrat des Korrelationskoeffizienten nennt man Bestimmtheitsmaß. Das ist der Anteil erklärter Varianz. Der liegt gerade einmal, wenn wir bei 0,7 sind, bei 50 %. Das ist keine besonders hohe Korrelation und sagt uns eigentlich, dass hier offensichtlich völlig unterschiedliche Dinge vorhergesagt werden. Das finde ich etwas beunruhigend.

Moderator Jürgen Windeler: Frau Mellert.

Uta Mellert: Ich hatte vorhin schon mal erwähnt: Es gibt auch eine Studie, wo der Prosigna-Test versus der IHC-Parameter verglichen wird, der einen deutlichen Benefit versus den Standardparametern zeigt, wenn es darum geht.

Es gibt auch eine Studie, die den Vergleich zwischen Oncotype DX und Prosigna gezeigt hat, in der die Unterschiede nicht so signifikant waren und sich primär um die Intermediärgruppe drehen. Das mag auch Begründungen haben, die wir im Einzelnen ... Aber wir können Ihnen die Studie natürlich zustellen.

Moderator Jürgen Windeler: Herr Dietel.

Manfred Dietel: Ich würde Sie gern fragen: Ist das ein offizieller Teil der Anhörung, den wir jetzt gerade begehen, oder ist es ein persönliches Gespräch von Herrn Lange? Ich finde problematisch, was im Augenblick passiert. Er hat eine persönliche Frage gestellt. Dann wird hin und her diskutiert. Ich frage mich: Ist das ein offizieller Teil? Das ist vorher nicht besprochen. Das kann man ja machen, aber ich halte es für problematisch, in diesem Zusammenhang persönliche Meinungen zu äußern, um mal zu hören, was die Kliniker so sagen oder was die Pathologen sagen oder wer auch immer. Das ist ja interessant, das kann man auch draußen machen, aber das gehört eigentlich nicht hierher. Das ist meine persönliche Meinung.

Moderator Jürgen Windeler: Gut, dann nehmen wir das als eine solche. - Ja, bitte.

Erwin Morawski: Ganz kurz zu Frau Janatzek noch einmal: Wir stimmen Ihnen in dem Punkt, dass es natürlich auf die Qualität und auf die Ergebnisqualität ankommt, zu. Ich denke: Das ist auch die Historie, die wir auch schon versucht haben, aufzumachen. Erst einmal:

Validierungen, retrospektiv-prospektiv belegen, dass die Daten tatsächlich die Qualität bringen.

Dann prospektive Studien: Ich glaube, Herr Langenfeld hat das vorhin in einem Satz anklingen lassen. Die TAILORx-Daten liegen Ihnen vor. Die planB-Daten werden wir Ihnen in sehr baldiger Zukunft zur Verfügung stellen können. Dann stehen also 2 prospektive Studien zur Verfügung.

Und aus unserer Sicht auch wichtig: die Validierung auch noch mal im klinischen Alltag durch die großen Kohorten, wo die Praxis belegt: Wie es das Ergebnis für die Patientin nach 5 Jahren? Sprich: erster Test, 5 Jahre Nachbeobachtung und dann das Ergebnis.

Kurz zum Punkt von Herrn Dietel, der natürlich sehr relevant ist: Was heißt es für die Versorgungsrealität? Weltweit wurden über 600 000 Tests bis jetzt mit dem Oncotype durchgeführt und in Deutschland mehrere Tausend Tests mit steigender Zahl, insgesamt über 13 000. Ich glaube: Das zeigt schon auch die Realität hier. Ich stimme Herrn Dietel zu: Das ist ein wachsender Markt, und das ist eine wachsende Bedeutung für die Patientin. Die Bedeutung hat Frau Schumacher-Wulf auch gesagt: Die Patientin braucht eine Entscheidungshilfe. Ich glaube: Das ist ein ganz wichtiger Punkt.

Aber nach wie vor, um noch einmal auf den Punkt zu kommen: Je besser validiert, desto sicherer für die Patientin, desto sicherer für den Anwender.

Moderator Jürgen Windeler: Gut, vielen Dank. - Zu der Frage, wie weit steigende Verkaufszahlen eine Bedeutung haben, hat sich, glaube ich, vorhin schon jemand kommentierend geäußert. Das will ich jetzt nicht noch einmal machen. - Gibt es noch weitere Wortmeldungen zum Punkt „Verschiedenes“? - Ja, bitte schön, Herr Petry.

Christoph Petry: Von mir noch ein kurzer Kommentar zu Ihnen, Herr Lange, zu der mehr oder weniger guten Korrelation der unterschiedlichen Genexpressionstests: Die Tests sind ja dazu optimiert, die Patientinnen zu finden, die Metastasen bekommen, sprich: Die Niedrigrisikogruppe soll nach Möglichkeit gänzlich frei von Metastasen sein. Das heißt, da natürlich die Summe aus Sensitivität und Spezifität immer endlich ist, dass Sie in der Hochrisikogruppe bei all diesen Tests relativ viele Patientinnen, ungefähr 50 %, haben, die tatsächlich, obwohl in der Hochrisikogruppe, auch ohne Chemotherapie keine Metastase bekommen. Da haben Sie natürlich einen gewissen Austausch zwischen den Gruppen, zwischen den Tests. Das heißt: Patientinnen, die im einen Tests in der in der Hochrisikogruppe sind, sind im anderen in der Niedrigrisikogruppe.

Das Wichtige ist, dass die keine Metastase bekommen, und das sehen sie tatsächlich auch. Das ist relativ selten. Insofern: Ja, die Korrelation ist nicht besonders, aber die klinische Anwendbarkeit der Tests wird durch diese mäßige Korrelation in keiner Weise beeinträchtigt. Ich glaube, das ist ganz wichtig zu erklären.

Dann zu ihrer zweiten Frage, wie man dann die Tests vergleicht: Es gibt ja mittlerweile - abgesehen von diesem relativ kleinen analytischen Studien, in den diese mehr oder weniger gute Korrelation festgestellt worden ist - auch recht große Studien, in denen mit ungefähr 1 000 Patienten dann wirklich verglichen ist, was hat der eine Test gesagt hat, was der andere Test gesagt hat und was dann am Ende nach 10 Jahren herausgekommen ist. Diese Vergleiche sind mittlerweile auch publiziert. Ich möchte keine Werbeveranstaltung für den einen oder für den anderen Test machen. Sie können das aber nachlesen. Die Daten sind da.

Moderator Jürgen Windeler: Gut, vielen Dank. - Herr Sinn.

Peter Sinn: Ich hätte noch eine Frage an das IQWiG. Und zwar ist mir aufgefallen, dass es ja etwa 10 bis 20 große internationale Nutzenbewertungen gibt, die etwa auf einer Augenhöhe sind - mit IQWiG und systematischen Reviews - und die praktisch alle zu einer vorsichtig positiven Bewertung kommen und sagen: Eine abschließende Evidenz gibt es nicht. Aber alles deutet darauf hin, dass es einen Nutzen für die Patientin gibt. Wie ist Ihre persönliche Einschätzung: Liegen diese zum Teil sehr ausführlichen Nutzenbewertungen - denken Sie an die NICE-Bewertung - falsch? Sind die alle auf dem Holzweg, oder was unterscheidet diese von Ihrer? Also ich meine: Das mit dem Knock-out-Kriterium hatten wir schon bei Punkt 1. Aber warum kommen die alle zu einem eigentlich komplett anderen Ergebnis?

Moderator Jürgen Windeler: Sie haben nach unserer persönlichen Einschätzung gefragt. Die geben wir jetzt vielleicht nicht mehr ab. Aber Herr Scheibler möchte eine vonseiten des IQWiG geben.

Fülöp Scheibler: Wir haben im Bericht, im hinteren Abschnitt „Einordnung der Ergebnisse“ unser Ergebnis im Lichte, im Gegensatz oder auch teilweise entsprechend den Berichten anderer HTA-Berichte und auch systematischer Übersichten beschrieben. Ich glaube, da ist unsere gemeinsame Meinung ganz gut zusammengefasst.

Moderator Jürgen Windeler: So gibt es weitere Wortmeldungen? - Sandra Janatzek.

Sandra Janatzek: Da ich vorhin vielleicht zu unspezifisch den Vergleich zur bisherigen Therapieentscheidung in den Raum geworfen habe, wollte ich das gern noch einmal konkretisieren. Aus unserer Sicht ist - ich bitte um Verzeihung, dass ich mich wiederhole im Vergleich zur Erörterung zum Berichtsplan - die vergleichende Nutzenbewertung im Vergleich zur herkömmlichen Therapieentscheidung unerlässlich. Aus unserer Sicht müssen wir den Vergleich auch mit führen bei den Prognosestudien - zumindest in der ganz nahe liegenden Weise, dass der G-BA natürlich wird wissen wollen, ob die Low-Risk-Gruppe, die EndoPredict oder Oncotype identifizieren kann und die den in ihrem Bericht dargelegten Kriterien genügt, nicht auch auf herkömmliche Weise identifiziert worden wäre. Also: Diese banale Frage muss sicherlich beantwortet werden.

Darüber hinaus fiel mir beim Schauen auf die 5-Jahres-Daten des Low-Risk-Stratums der TAILORx-Studie auf, dass sich ja schon ein nicht unerheblicher Anteil von Frauen mit G3-

Tumoren in dieser Low-Risk-Gruppe befindet. Das waren 111 Frauen, 7 %. Mit Blick darauf muss man vielleicht auch überlegen ob, man sich insbesondere der Patientinnengruppe mit diskordanten Therapieempfehlungen - Biomarker versus herkömmliches Vorgehen - nochmals spezifisch widmen sollte.

Moderator Jürgen Windeler: Herr Schmidt.

Marcus Schmidt: Mit dem kurzen Versuch, diese, wie sie sagten, banale Frage, die natürlich klinisch absolut relevant ist für die Patientinnen, zu beantworten. Ich denke: Man hat es zumindest näherungsweise durch die Arbeiten Eiermann und Müller beantwortet - das war die Erstautorin von der anderen Arbeit mit dem anderen Test -, die gelaufen sind und die gezeigt haben, dass etwas mehr als ein Drittel diskordant eingestuft werden zwischen klinisch-pathologischen, also dem, was man ansonsten nach alter Sitte hat, und dem Ergebnis eines Genexpressionstests. Die klinische Erfahrung lehrt natürlich schon, dass die Frauen in den meisten Fällen, wenn man mit diesem Testergebnis argumentiert, dann auch die Chemotherapie dementsprechend annehmen oder bzw. mit Kusshand am liebsten natürlich darauf verzichten, wenn sie im Low-Risk-Stratum sind.

Ich denke: Diese Frage - das wird nicht mehr viel Mühe machen - ist im Wesentlichen beantwortet. - Danke.

Moderator Jürgen Windeler: Bitte schön.

Eva Schumacher-Wulf: In Patientenkreisen ist außerdem bekannt, dass viele Frauen, und zwar sehr großer Teil, keinen Nutzen von der Chemotherapie haben. Und trotzdem haben die Frauen im Moment keine Möglichkeit, es sei denn, sie zahlen selbst oder es sind Privatpatientinnen, Gewissheit zu bekommen, ob sie mit Blick auf die Therapieentscheidung einen Nutzen haben oder nicht.

So eine Chemotherapie durchzustehen, wenn man nicht einmal weiß, ob man davon profitiert, ist hart. Die Frauen sind einfach mittlerweile informiert und wollen Gewissheit haben.

Moderator Jürgen Windeler: Herr Sotlar.

Karl Sotlar: Ich möchte jetzt doch noch einmal einzelne Daten aus einer Studie, die wir durchgeführt haben, deren Ergebnisse jetzt zur Publikation anstehen, zum Besten geben - unter Benutzung des Prosigna-Tests. Jetzt auch noch einmal zum Vergleich mit der herkömmlichen Versorgung: Frau Mellert hat ja gesagt, der Prosigna-Test liefere eben auch noch die wichtige Information des sogenannten intrinsischen oder molekularen Subtyps. In dieser Gruppe, von der wir hier sprechen, also der hormonrezeptorpositiven, also luminalen Tumoren, ist es die Unterscheidung luminal A versus luminal B.

In dieser Studie differierten in etwa 37 % diese molekularen Subtypen, was das Ergebnis des Prosigna-Tests und eben die Einschätzung über die immunhistochemische Bestimmung dieses

luminalen Subtyps betraf. Es ist ferner so, dass ungefähr 5 % der Patientinnen aus der molekularen Low-Risk-Gruppe am Ende der Chemotherapie bekommen haben, was ein sehr niedriger Prozentsatz ist. Ich kann Ihnen jetzt im Einzelnen nicht sagen, warum das dann so war, und das 16 von 18 Patientinnen aus der High-Risk-Gruppe, die auf der Basis der konventionellen Therapieempfehlung nach Kenntnis des Prosigna-Tests mit dem Ergebnis High Risk gewichtet haben auf Chemotherapieanwendung.

Also ich denke: Die Patientinnen haben das Ergebnis dieser Testung entsprechend ernst genommen und gewürdigt. Und natürlich ist das eine Gruppe von Patientinnen, die da eingeschlossen wurde, die sich natürlich bereit erklärt hat, das Ergebnis des Tests im Grunde auch zu respektieren, und nicht von vornherein schon gesagt hat: Egal, wie der Test jetzt ausfällt, ich mache das oder ich mache jenes.

Moderator Jürgen Windeler: Vielen Dank. - Dann sehe ich jetzt keine weiteren Wortmeldungen mehr. Ich bedanke mich sehr, dass Sie hier waren, und bedanke mich noch einmal sehr dafür, dass Sie mit uns jetzt, sozusagen in der zweiten Runde noch einmal einige, für die Bewertung sicherlich relevante Fragen ausgetauscht haben.

Wir werden uns jetzt unter Berücksichtigung der Stellungnahmen, auch unter Berücksichtigung der Dinge, die wir heute gehört haben, auf den Weg machen und uns zu einem Abschlussbericht in gebotener Form und in gebotener Zeit durchringen.

Vielen Dank und kommen Sie gut nach Hause.

(Beifall)

Anhang A – Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
A.1 – Stellungnahmen von Organisationen, Institutionen und Firmen	A 2
A.1.1 – Berufsverband der Frauenärzte (BVF) e. V.	A 2
A.1.2 – Deutsche Gesellschaft für Gynäkologie und Geburtshilfe (DGGG) e. V. & Deutsche Gesellschaft für Senologie (DGS) e. V.	A 8
A.1.3 – Deutsche Gesellschaft für Pathologie (DGP) e. V. & Bundesverband Deutscher Pathologen (BDP) e. V.	A 14
A.1.4 – Genomic Health Deutschland GmbH	A 25
A.1.5 – Mamma Mia! Das Brustkrebsmagazin.....	A 45
A.1.6 – Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen (MDS) e. V.	A 48
A.1.7 – NanoString Technologies GmbH	A 54
A.1.8 – Sividon Diagnostics GmbH	A 68

A.1 – Stellungnahmen von Organisationen, Institutionen und Firmen

A.1.1 – Berufsverband der Frauenärzte (BVF) e. V.

Autoren:

- Albring, Christian
- Kiechle, Marion



Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Stellungnahme zum Vorbericht D14: Biomarkerbasierte Tests

Herrn Professor

Dr. med Jürgen Windeler

Im Mediapark 8

50670 Köln

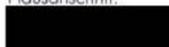
Geschäftsstelle
Postfach 20 03 63
80003 München
Telefon (0 89) 2 444 66 0
Telefax (0 89) 2 444 66 100
E-Mail bvf@bvf.de
Internet <http://www.bvf.de>

Betr.: Schriftliche Stellungnahme zum Vorbericht des IQWiG zu „Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Therapie beim primären Mammakarzinom (Auftrag D14-01 des G-BA)

Sehr geehrter Herr Professor Windeler,

als Ärztinnen und Ärzte in Forschung und Lehre, in der Klinik und der Niederlassung sind wir täglich mit der Beurteilung neuer Untersuchungs- und Behandlungsmethoden und deren Überführung in die klinische Praxis befasst. Daher haben wir, der Vorstand des Berufsverbands der Frauenärzte, Ihren Vorbericht zu biomarkerbasierten Tests beim Mammakarzinom (Auftrag D14-01) geprüft und eingehend diskutiert. Wir sind zu der Auffassung gekommen, dass das Ergebnis Ihres Vorberichts auf einer problematischen gesundheitspolitischen Zielsetzung und im Wesentlichen zwei methodischen Fehlern beruht.

Ausschließliches Kriterium für die Entscheidung für oder gegen Chemotherapie bei Patientinnen mit primärem Mammakarzinom ist derzeit das individuelle Metastaserisiko einer Patientin^{1,2}. Methoden, um festzustellen, welche spezifische Chemotherapie besonders erfolgversprechend ist, sind weiterhin unbekannt. Bei Patientinnen mit HER2-positiven oder triple-negativen Karzinomen wird aus guten Gründen von einem so hohen Risiko ausgegangen, dass eine Chemotherapie normalerweise indiziert ist. Bei Patientinnen mit östrogenrezeptorpositivem und HER2-negativem Tumor versuchen die Gynäkologen und Onkologen weltweit, mit der Hilfe verschiedener Methoden das Metastaserisiko so präzise wie möglich zu fassen, um eine evidenzbasierte Therapieentscheidung zu ermöglichen. Als Grenzwert zwischen hohem und niedrigem Risiko hat sich dabei nach internationalem Konsens ein 10 Jahres-Metastase Risiko von 10% etabliert^{3,4}. In Vorbericht (Seite 149) versuchen Ihre Sachbearbeiter zu begründen, warum dieser Grenzwert bei 5% liegen sollte. Sie versuchen dabei, die Anzahl der durch Chemotherapie herbeigeführten „Hospitalisierungen“ gegen die durch einen Verzicht auf Chemotherapie möglicherweise bedingten Metastasierungen bzw. Todesfälle abzuwägen. Diese Gegenüberstellung ist jedoch medizinisch nicht sinnvoll: Bei der Betrachtung der Argumente, die gegen eine Chemotherapie sprechen, geht es nur in zweiter Linie um Nebenwirkungen, die wir durch „Hospitalisierung“ auffangen können. Ganz wesentlich für die Schaden-Nutzen-Betrachtung einer Chemotherapie sind ihre irreversiblen Nebenwirkungen. Dabei



sind neben akuten Todesfällen vor allem durch Infektionen und Sepsis Schädigungen der sensorischen Nerven der Patientinnen zu nennen; weiterhin Kardiomyopathien, nicht selten mit Todesfolge, und Sekundärtumore, primär des hämatopoetischen Systems und der Reproduktionsorgane. In Summe geht die Literatur davon aus, dass die Chemotherapie bei über 6% der mit ihrer Hilfe behandelten Brustkrebspatientinnen innerhalb von 10 Jahren ursächlich zum Tod oder zu schweren Langzeitschäden führt^{5, 6}. Vor diesem Hintergrund ist es nicht angemessen, bei der Festlegung eines Grenzwerts zur Entscheidung pro oder kontra Chemotherapie nur vermeidbare Hospitalisierungen zu betrachten. Zur Begründung ab welchem Metastaserisiko keine Chemotherapie verabreicht werden sollte, ist deswegen eine andere Überlegung zielführender: Die relative Reduktion des Tumorrisikos innerhalb von 10 Jahren, die durch eine Chemotherapie erreichbar ist, beträgt ca. 30%⁷. Bei einem Metastaserisiko von 10% bewirkt eine Chemotherapie somit eine absolute Risikoreduktion von 3%. Dem steht ein Todesrisiko durch die Chemotherapie von im Mittel ebenfalls etwa 3% entgegen – womit der heute international verwendete Grenzwert von 10% Metastaserisiko bei der Entscheidung für oder gegen Chemotherapie einfach begründet werden kann. Nimmt man, wie Ihre Sachbearbeiter im Vorbericht es tun, eine Reduktion des Rezidivrisikos durch Chemotherapie von relativ nur 20 % an, würde sich der Grenzwert, unterhalb dessen auf eine Chemotherapie verzichtet werden sollte, auf ca. 15 % Metastaserisiko innerhalb von 10 Jahren erhöhen. Dieser Grenzwert wäre gegenüber dem internationalen Konsens zu hoch – wir könnten aber zumindest die Argumentation, die zu diesem Grenzwert führt, nachvollziehen. Der von Ihren Sachbearbeitern vorgeschlagene Grenzwert von 5 % ist dagegen medizinisch nicht begründbar. Seine Anwendung hätte zur Konsequenz, dass wir den Einsatz von Chemotherapie bei Patientinnen mit ER-positivem und HER2-negativem Mammakarzinom gegenüber der heutigen Praxis nahezu verdoppeln müssten. Das ist aber nicht nur gesundheitspolitisch und gesundheitsökonomisch kontraindiziert; es würde vor allem die Behandlung der Brustkrebspatientinnen gegenüber dem Status Quo verschlechtern und die Gesamtmortalität durch Brustkrebs voraussichtlich erhöhen.

Konsistent mit Ihrem Vorschlag, 5% Metastaserisiko innerhalb von 10 Jahren als Grenzwert für die Chemotherapieindikation anzusetzen – und leider medizinisch ebenso schädlich für die Patientinnen - ist die Vorgabe Ihrer Sachbearbeiter, dass die Biomarker bei mindestens 15 % der Patientinnen der Zielkohorte einen Verzicht auf Chemotherapie begründen müssten (Seite 21 Ihres Vorberichts). Bei dieser Vorgabe haben Sie – abgesehen von unserer oben bereits erläuterten Argumentation - außer Acht gelassen, dass wir bereits heute bei ca. 40% der Zielpopulation der Biomarker keine Chemotherapie verabreichen. Entsprechend wäre eine Chemotherapieindikation für 85% der Patientinnen basierend auf einem neuen Biomarker ein herber gesundheitspolitischer Rückschlag für Deutschland und ein Schritt in eine therapeutisch völlig falsche Richtung. Die Konsequenz der Verwendung eines solchen Biomarkers wäre, dass wir bei Patientinnen mit rezeptor-positivem Mammakarzinom ca. doppelt so viele Chemotherapien verordnen müssten wie bisher. Wie Ihre Sachbearbeiter im Ihrem Vorbericht richtig schreiben, benötigen jedoch ca. 80% der Patientinnen der Zielgruppe KEINE Chemotherapie – warum sollten wir auf der Basis eines neuen Biomarkers nun 85% der Patientinnen EINE Chemotherapie verabreichen? Durch den Einsatz zumindest einiger der Biomarker, die Sie in Ihrem Vorbericht auf Grund angeblich nicht vorhandener Evidenz ablehnen, ist es wissenschaftlich gesichert möglich, den Einsatz von Chemotherapie gegenüber heute deutlich zu verringern^{8, 9, 10}. Wir bitten Sie daher Ihren Vorbericht dahin gehend zu überarbeiten, dass Sie den Grenzwert für die Entscheidung für oder gegen Chemotherapie bei 10% Tumorrisiko innerhalb von 10 Jahren ansetzen und nur Biomarker zur Verwendung empfehlen, die keinen gegenüber heute erhöhten Einsatz von Chemotherapie zur Folge hätten.

In Ihrem Vorbericht führen insbesondere zwei Ihrer Forderungen dazu, dass Sie praktisch die gesamte wissenschaftliche Literatur zu den Biomarkern bestehend aus zahlreichen oft hochrangig publizierten Studien an tausenden Patientinnen pauschal und ohne weitere Prüfung als nicht aussagekräftig zurück weisen. Diese Forderungen Ihrer Sachbearbeiter sind:

1. dass bei prospektiv-retrospektiven Prognosestudien mindestens 70 % der ursprünglichen Studienpopulation eingeschlossen sein müssen (Seite 6 des Vorberichts) und
2. dass die mittlere Nachbeobachtungszeit bei prognostischen Studien mindestens 10 Jahre betragen müsse (Seite 19 des Vorberichts, letzte Zeile).

Randomisierte Studien dagegen lassen Sie dagegen zu, selbst wenn die Nachbeobachtungszeit nur ein Jahr beträgt.

Als Begründung für Forderung 1 führen Ihre Sachbearbeiter unter anderem die zentrale Publikation von Simon und Hayes an, die die wesentlichen theoretischen Grundlagen für die bisher publizierten prognostischen Biomarkerstudien liefert¹¹. Allerdings begründen Sie Ihre Forderung 1, indem Sie diese Arbeit unvollständig und vor allem sinnenstehend zitieren: Simon und Hayes schreiben, dass man bei einer Rekrutierung von zwei Drittel der Patientinnen aus der Originalstudie in eine prognostische Substudie davon ausgehen könne, dass Verzerrungseffekte vermieden werden können. Die Autoren machen jedoch deutlich, dass dies nicht – so wie Ihre Sachbearbeiter es fälschlicher Weise getan haben – als eine undifferenzierte Forderung zu verstehen sei. Stattdessen führen die Autoren differenziert aus, wie Verzerrungen unabhängig vom Rekrutierungsanteil vermieden werden können. Auch die übrige in Ihrem Vorbericht zitierte Literatur liefert keine Begründung für die Sinnhaftigkeit der von Ihren Sachbearbeitern erhobene 70%-Forderung und die darauf basierte pauschale Zurückweisung praktisch der gesamten Literatur zu den Biomarkern beim Mammakarzinom. Sowohl Hayden¹² als auch Altmann¹³ empfehlen differenzierte Prüfungen zur Vermeidung von Verzerrungseffekten der Stichprobe gegenüber der Grundgesamtheit. Zahlreiche der von Ihnen zurück gewiesenen Biomarker Studien werden gerade der zentralen Forderung von Altmann nach Verwendung von kontinuierlichen Klassifikatoren statt einfachen prognostischen Kategorisierungen exemplarisch gerecht. Wir weisen Ihre Forderung 1 daher als wissenschaftlich nicht begründbar zurück und bitten Sie, die prognostischen Studien zu den Biomarkern, die bereits Eingang in nationale und internationale Behandlungsleitlinien für Brustkrebs gefunden zu haben, für Ihrem Hauptbericht in angemessener Weise zu prüfen.

Auch Forderung 2 nach einer mittleren Nachbeobachtungszeit von mindestens 10 Jahren für Prognosestudien (siehe Seite 22 Ihres Vorberichts) und Ausschluss der Studien, die diese Anforderung nicht erfüllen, ist ebenfalls wissenschaftlich nicht begründbar. Der Berufsverband der Frauenärzte begrüßt, dass Ihre Sachbearbeiter hier offensichtlich versuchen, den Erfordernissen der Langzeitsicherheit der Patientinnen und der vergleichsweise häufigen Spätmetastasierung¹⁴ bei rezeptor-positivem Brustkrebs Rechnung zu tragen. Die Forderung einer definierten mittleren Nachbeobachtungszeit macht jedoch deutlich, dass Ihre Sachbearbeiter die Eigenschaften der statistische Methode des Kaplan-Meier Schätzers zur Auswertung von klinischen Studien im allgemeinen und Prognosestudien im Besonderen bietet, nicht berücksichtigt haben. Die mittlere Nachbeobachtungszeit ist für eine Kaplan-Meier Analyse ohne Relevanz. Es ist geradezu ein Charakteristikum dieser Methode, dass sie zuverlässige Aussagen für Zeitintervalle deutlich jenseits der mittleren Nachbeobachtungszeit zulässt, sofern für diese Zeitintervalle noch Datenpunkte in ausreichender Anzahl vorliegen. Ist dies nicht mehr der Fall, werden die Konfidenzintervalle für diese

Zeitpunkte breit bzw. unbestimmt. Mit der Forderung, dass die obere bzw. untere Grenze der Konfidenzintervalle der Kaplan-Meier Schätzung zum 10 Jahreszeitpunkt bestimmte Kriterien einhalten muss, ist der Patientinnensicherheit vollumfänglich genüge getan. Auf der Basis einer willkürlich gesetzten Grenze der mittleren Nachbeobachtungszeit ganze Studien bei der Bewertung der Biomarker auszuschließen, ist somit zweifellos methodisch falsch. Der Berufsverband der Frauenärzte bitte Sie, diesen Fehler in Ihrem abschließenden Bericht zu korrigieren.

Nicht akzeptabel im Sinne der Sicherheit der Patientinnen wird der Vorbericht Ihrer Sachbearbeiter durch das Kriterium, dass randomisierte Studien schon ab einer Nachbeobachtungszeit von einem Jahr berücksichtigt werden sollen (siehe Seite 19). Nachbeobachtungszeiten von nur einem Jahr – oder auch von nur 5 Jahren - verletzen elementare Sicherheitsanforderungen für die Patientinnen, da eine so kurze Nachbeobachtungszeit bei rezeptorpositivem Brustkrebs keine ethisch oder wissenschaftlich vertretbare Aussage zur Wirksamkeit einer therapeutischen oder diagnostischen Methode erlaubt. Der Berufsverband der Frauenärzte fordert das IQWiG daher auf, konsistent mit der wissenschaftlichen Literatur und den gültigen Leitlinien bei der Bewertung von Biomarkern nur die Studien – egal ob prognostisch oder randomisiert - zu betrachten, die in der Kaplan-Meier Analyse eine statistisch valide Aussage über ein Behandlungsergebnis nach 10 Jahren erlauben. Neue prospektive Studien, die lediglich eine Aussage über kürzere Nachbeobachtungszeiträume zulassen, können lediglich als zusätzliche Evidenz bei der Einordnung bereits publizierten Studien mit langer Nachbeobachtungszeit dienen.

Wir hoffen unsere kurze Begründung hat gezeigt, dass der von Ihnen veröffentlichte Vorbericht das angestrebte Ziel einer sachlichen und wissenschaftlich fundierten Bewertung der untersuchten Biomarker nicht erreicht hat. Seine Schlussfolgerung, dass es für den Nutzen der Marker keinen Anhaltspunkt gebe, ist entsprechend primär das Ergebnis wissenschaftlich nicht angemessener Beurteilungskriterien und der medizinisch wie gesundheitspolitisch verfehlten Zielsetzung, die Indikation von Chemotherapie zu erweitern.

Der Berufsverband der Frauenärzte bittet daher um eine umgehende Neubewertung der Biomarkertests unter Berücksichtigung international akzeptierter und medizinisch sinnvoller Standards, um eine verbesserte Behandlung unserer Patientinnen zu ermöglichen und ihre Sicherheit auch künftig nicht zu gefährden.

Mit freundlichen Grüßen,



Prof. Dr. Marion Kiechle
Direktorin der Frauenklinik der
Technischen Universität München



Dr. Christian Albring
Präsident Berufsverband der
Frauenärzte

¹ Interdisziplinäre S3 Leitlinie für die Diagnostik, Therapie und Nachsorge des Mammakarzinoms, AWMF-Register Nummer 032 – 045OL; <http://www.senologie.org/fileadmin/downloads/S3-Brustkrebs-v2012-OL-Langversion.pdf>

² http://www.ago-online.de/fileadmin/downloads/leitlinien/mamma/maerz2015/de/2015D_05_Prognostische_und_praediktive_Faktoren_NEU_150216.pdf

³ Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MINDACT trial. *J Clin Oncol.* 2008;26:729-35.

⁴ Harbeck N, Schmitt M, Meisner C, Friedel C, Untch M, Schmidt M, Sweep CG, Lisboa BW, Lux MP, Beck T, Hasmüller S, Kiechle M, Jänicke F, Thomssen C; Chemo-N 0 Study Group. Ten-year analysis of the prospective multicentre Chemo-N0 trial validates American Society of Clinical Oncology (ASCO)-recommended biomarkers uPA and PAI-1 for therapy decision making in node-negative breast cancer patients. *Eur J Cancer.* 2013;49:1825-35.

⁵ Mackey JR, Martin M, Pienkowski T, Rolski J, Guastalla JP, Sami A, Glaspy J, Juhos E, Wardley A, Fornander T, Hainsworth J, Coleman R, Modiano MR, Vinholes J, Pinter T, Rodríguez-Lescure A, Colwell B, Whitlock P, Provencher L, Laing K, Walde D, Price C, Hugh JC, Childs BH, Bassi K, Lindsay MA, Wilson V, Rupin M, Houé V, Vogel C; TRIO/BCIRG 001 investigators. Adjuvant docetaxel, doxorubicin, and cyclophosphamide in node-positive breast cancer: 10-year follow-up of the phase 3 randomised BCIRG 001 trial. *Lancet Oncol.* 2013;14:72-80.

⁶ Karp JE, Blackford A, Visvanathan K, Rugo HS, Moy B, Goldstein LJ, Stockerl-Goldstein K, Neumayer L, Langbaum TS, Hughes ME, Weeks JC, Wolff AC. Myelodysplastic Syndrome and/or Acute Myelogenous Leukemia (MDS and/or AML) after a Breast Cancer Diagnosis: The National Comprehensive Cancer Network (NCCN) Experience. *SABCS 2012* (oral presentation)

⁷ 16. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Peto R, Davies C, Godwin J, Gray R, Pan HC, Clarke M, Cutter D, Darby S, McGale P, Taylor C, Wang YC, Bergh J, Di Leo A, Albain K, Swain S, Piccart M, Pritchard K. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet.* 2012;379:432-44.

⁸ Kiechle M, *Frauenarzt* 2014; 55: 10-11.

⁹ Müller BM1, Keil E, Lehmann A, Winzer KJ, Richter-Ehrenstein C, Prinzler J, Bangemann N, Reles A, Stadie S, Schoenegg W, Eucker J, Schmidt M, Lippek F, Jöhrens K, Pahl S, Sinn BV, Budczies J, Dietel M, Denkert C. The EndoPredict Gene-Expression Assay in Clinical Practice - Performance and Impact on Clinical Decisions. *PLoS One.* 2013 Jun 27;8(6):e68252. Print 2013.

¹⁰ J. Albanell1,2,3,* , A. González4, M. Ruiz-Borrego5, E. Alba6, J. A. García-Saenz7, J. M. Corominas2,3,8, O. Burgues9, V. Furio10, A. Rojo11, J. Palacios12, B. Bermejo13, M. Martínez-García1,2, M. L. Limon5, A. S. Muñoz6, M. Martín14, I. Tusquets1,2, F. Rojo2,15, R. Colomer4, I. Faull16 and A. Lluch13, Prospective transGEICAM study of the impact of the 21-gene Recurrence Score assay and traditional clinicopathological factors on adjuvant clinical decision making in women with estrogen receptor-positive (ER+) node-negative breast cancer. *Ann Oncol* (2012) 23 (3): 625-631. doi: 10.1093/annonc/mdr278.

¹¹ Simon RM1, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009 Nov 4;101(21):1446-52. doi: 10.1093/jnci/djp335. Epub 2009 Oct 8.

¹² Jill A. Hayden, DC, PhD; Danielle A. van der Windt, PhD; Jennifer L. Cartwright, MSc; Pierre Co`te´ , DC, PhD; and Claire Bombardier, MD. Assessing Bias in Studies of Prognostic Factors, *Ann Intern Med.* 2013;158:280-286.

¹³ Douglas G Altman, Systematic reviews of evaluations of prognostic variables. *BMJ* 2001;323:224-8

¹⁴ Yu KD1, Wu J, Shen ZZ, Shao ZM. Hazard of breast cancer-specific mortality among women with estrogen receptor-positive breast cancer after five years from diagnosis: implication for extended endocrine therapy. *J Clin Endocrinol Metab.* 2012 Dec;97(12):E2201-9. doi: 10.1210/jc.2012-2423. Epub 2012 Sep 19.

A.1.2 – Deutsche Gesellschaft für Gynäkologie und Geburtshilfe (DGGG) e. V. & Deutsche Gesellschaft für Senologie (DGS) e. V.

Autoren:

- Beckmann, Matthias W.
- Schmidt, Marcus
- Schulz-Wendtland, Rüdiger



DGGG e.V. • Hausvogteiplatz 12 • 10117 Berlin

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

Stellungnahme zum Vorbericht D14-01

Prof. Dr. med Jürgen Windeler

Im Mediapark 8

50670 Köln

Per Mail: berichte@iqwig.de

cc/Herrn Prof. Diethelm Wallwiener; Frau Fragale; Frau Frohloff

Präsident

Prof. Dr. Diethelm Wallwiener
Ärztlicher Direktor
Universitäts-Frauenklinik Tübingen

Repräsentanz der DGGG und
Fachgesellschaften
Hausvogteiplatz 12
D – 10117 Berlin
Telefon: +49 (0) 30 514883333
Telefax: +49 (0) 30 51488344
info@dggg.de
www.dggg.de

DGGG-Stellungnahensekretariat

Frauenklinik
Universitätsklinikum Erlangen
Universitätsstraße 21-23
91054 Erlangen
Telefon: +49 (0) 9131-85-44063
+49 (0) 9131-85-33507
Telefax: +49 (0) 9131-85-33951
E-Mail: fk-dggg-stellungnahmen@uk-erlangen.de
www.frauenklinik-uk-erlangen.de

01.12.2015

Stellungnahme der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe (DGGG)

zu D14-01 Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mamma-Karzinom

Im Namen der Arbeitsgemeinschaft Gynäkologische Onkologie (AGO) in der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe geben wir zum Vorbericht (vorläufige Nutzenbewertung) „Biomarker-basierter Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mamma-Karzinom“ (Auftragsnummer: D14-01) nach kritischer Diskussion folgende Stellungnahme ab:

Anmerkungen zur projektspezifischen Methodik

Da Patientinnen mit einem geringen absoluten Risiko nur einen geringen absoluten Nutzen durch eine adjuvante Chemotherapie haben [1], begrüßen wir es, dass in dem Vorbericht auch Prognosestudien mit herangezogen worden sind. Wir können allerdings den im Vorbericht vorgeschlagenen Grenzwert (vgl. auch A10) für das 10-Jahres Rezidivrisiko von 5 % (obere Grenze des 95 %-Konfidenz-intervalls) statt der national und international üblichen 10% nicht nachvollziehen. Auch wenn natürlich alle Anstrengungen unternommen werden müssen, Rezidive nach Möglichkeit zu verhindern, würde ein solch niedriger Grenzwert notwendigerweise zu einer noch häufigeren Übertherapie führen.

Laut Vorbericht wurden vom Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) nach systematischer Literaturrecherche 4270 Treffer zu randomisierten kontrollierten Studien (RCT) sowie 5350 Treffer zu Prognosestudien aufgeführt. Nach der bibliografischen Literaturrecherche wurde die überwiegende Mehrzahl der Arbeiten „aufgrund fehlender Relevanz ausgeschlossen“. Somit wurden lediglich jeweils 2 Studien als relevant für die Fragestellung der vorliegenden Nutzenbewertung identifiziert.



Als grundsätzlich problematisch erachten wir die Selektionskriterien für die Auswahl der berücksichtigten Studien, die auch zu offensichtlichen Widersprüchen führen: So wurde zum Beispiel als Prognosestudie eine prospektiv-retrospektive Studie am Kollektiv der Austrian Breast and Colorectal Cancer Study Group Trial 8 (ABCSG-8) Studie [2] berücksichtigt, während mit dem lapidaren Hinweis „Nicht E4 (Studientyp)“ eine zweite prospektiv-retrospektive Studie an eben diesem ABCSG-8 Kollektiv [3] nicht berücksichtigt wurde.

Das sogenannte „prospektiv-retrospektive“ Design zur Evidenzbeurteilung von Biomarkern ist mittlerweile international bei der Untersuchung von Biomarkern anerkannt [4, 5]. Die Validität der Ergebnisse von prospektiv-retrospektiven Studien konnte aktuell prospektiv überzeugend bestätigt werden: Die initial 2004 prospektiv-retrospektiv beschriebene günstige Prognose der low-risk Gruppe von Oncotype DX® [6] entspricht grundsätzlich dem 2015 in der prospektiven TAILORx Studie publizierten low-risk stratum [7]. Somit kann, wie von Simon et al. postuliert, Level I Evidenz nicht nur prospektiv (Level IA) sondern auch prospektiv-retrospektiv (Level IB) erreicht werden [4]. Bezüglich der Länge der Nachbeobachtungszeit liegen sogar potentielle Vorteile bei dem prospektiv-retrospektiven Ansatz. Da beim hormonrezeptorpositiven Mammakarzinom viele Rezidive erst nach mehr als 5 Jahren auftreten und daher Patientinnen von einer erweiterten endokrinen Therapie profitieren [8], können prospektiv-retrospektive Studien mit einer Nachbeobachtungszeit von 10 Jahren aussagekräftiger sein als prospektive Studien mit einer Nachbeobachtungszeit von 5 Jahren, da bei einer Nachbeobachtungszeit von 10 Jahren auch späte Metastasen berücksichtigt werden können. In diesem Zusammenhang sollte beachtet werden, dass, anders als unter A.2.1.2.5 (Studiendauer) festgelegt, nicht nur Studien mit einer „mediane Dauer der Nachbeobachtung ... mindestens 10 Jahre“ berücksichtigt werden sollten. Analysen zum 10-Jahres-Überleben mittels Kaplan-Meier-Schätzer sind von einer kürzeren medianen Nachbeobachtungszeit nicht betroffen, da der Schätzer auch in solchen Fällen unverzerrt bleibt. Entscheidend ist in diesem Kontext, dass die Nachbeobachtungszeit einen Zeitraum von mindestens 10 Jahren umfasst.

Wir sehen es als problematisch an, dass „prospektiv-retrospektive“ Studien bei der Zielpopulation von Östrogerezeptor (ER) positiven und Human Epidermal Growth Factor Receptor -2 (HER2) negativen Patientinnen offensichtlich nur unzureichend berücksichtigt worden waren. Interessanterweise ging im Vorbericht nur eine prospektiv-retrospektive Studie in die vorläufige Nutzenbewertung ein:

Tabelle 1: Studienpool der Nutzenbewertung

Studientyp	Studie	Biomarker	Endpunkte / Zielgrößen		
			Fernrezidive	Rezidive	Gesamtüberleben
RCT	Paik 2006	Oncotype DX	-	-	-
	Martin 2014	EndoPredict	○	-	○
Prognosestudie	Gnant 2014	PAM50 Risk of Recurrence Score	-	-	-
	Sgroi 2013	Breast Cancer Index	-	-	-

-: Für den Bericht auswertbare Daten standen nicht zur Verfügung; ○: für den Bericht extrahierte Daten



Bei dieser Studie [9] wurde im Vorbericht allerdings nur ein möglicher Unterschied zwischen den beiden Behandlungsarmen in Bezug auf das Ergebnis des Genexpressionstests (EndoPredict®) betrachtet. Hier fand sich wie zu erwarten kein signifikanter Unterschied. Nicht berücksichtigt wurde im Vorbericht allerdings die für die Fragestellung des Vorberichts relevante prognostische Aussagekraft der Testergebnisse. Patientinnen in der Niedrigrisikogruppe hatten ein 10-Jahres Metastasenfreies Überleben von 93%, Patientinnen in der Hochrisikogruppe eines von 70% (Hazard Ratio [HR] = 4,8; 95% Konfidenzintervall [CI] = 2,5 – 9,6]; $P < 0,0001$) [9]. Diese prognostische Aussagekraft war unabhängig von den traditionellen prognostischen Faktoren wie Tumorgroße, Anzahl der befallenen Lymphknoten, histologischer Differenzierungsgrad, Alter, Hormonrezeptorstatus oder Ki-67. Dies ist aus klinischer Sicht der eigentliche Grund, Genexpressionstests zu verwenden. Die zur Begutachtung stehenden biomarkerbasierten Tests wurden nicht entwickelt, um einen Unterschied zwischen zwei unterschiedlichen Chemotherapieprogrammen nachzuweisen, sondern um das Risiko der Patientinnen für eine spätere Metastasierung abzuschätzen.

Somit bleibt festzuhalten, dass von 4 im Vorbericht nach bibliographischer Literaturrecherche ausgewählten prospektiv-retrospektiven Studien nur eine in der vorläufigen Nutzenbewertung berücksichtigt wurde. Diese wurde allerdings, wie oben ausgeführt, in Bezug auf den Auftrag des Gemeinsamen Bundesausschusses (G-BA) („Bewertung von biomarkerbasierten Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom“) bedauerlicherweise aus unserer Sicht falsch interpretiert.

Die übrigen prospektiv-retrospektiven Studien wurden laut Vorbericht anhand von Kriterien nicht berücksichtigt, die keine klare evidenzbasierte Grundlage haben („Ergebnisse fließen in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen basieren ...“). Simon und Mitarbeiter adressierten die mögliche Problematik eines Selektionbias [4] und führten an, dass die untersuchten Präparate repräsentativ für die in der Studie untersuchten Patientinnen sein sollten („It is clearly desirable that the available specimens from the archived bank should be representative of the patients who were accrued to the study as a whole...“). Gleichzeitig stellten sie allerdings klar, dass es keine Garantie gibt, dass die rekrutierten Studienpatientinnen repräsentativ für die Gesamtpopulation sind („...although there is no guarantee that the study patients are themselves representative of the general population of patients“). Um diese Problematik zu adressieren, schlugen Simon und Mitarbeiter zwar vor, dass in prospektiv-retrospektiven Untersuchungen wenigstens zwei Drittel der Studienpopulation untersucht werden sollten („...we suggest that the correlative study should include at least two-thirds of the total accrued patients“), gleichzeitig machten sie allerdings klar, dass keine Minimalanforderungen bezüglich der Prozentzahl umfassend geltend gemacht werden können („...there are no minimal requirements that can be universally applicable...“).

In einer weiteren Arbeit wurden die von Simon et al. aufgestellten Kriterien zum Erreichen von Level IB Evidenz mittels prospektiv-retrospektiver Biomarkerstudien erneut aufgegriffen [5]. Hier sprachen die Autoren nur noch davon, dass ausreichend archiviertes Gewebe analysiert werden sollte, um eine ausreichende statistische Power zu haben und um repräsentativ für die untersuchten Patientinnen zu sein („Adequate amounts of archived specimen must be available... to have adequate statistical power and for the patients included in the evaluation to be clearly representative of the patients in the trial.“). Sie verzichteten allerdings aus guten Gründen auf eine genaue Festlegung, wieviel Prozent der Studienpatientinnen untersucht werden sollten. Dies ist auch aus unserer Sicht gerechtfertigt, da es keine plausible Evidenz für eine genaue Prozentzahl des zu untersuchenden Archivmaterials gibt. Bei Prognosestudien



kommt es auf eine Repräsentativität der untersuchten Patientinnen für das Studienkollektiv sowie auf eine ausreichende statistische Power an, um Unterschiede im Überleben nachweisen zu können.

Im vorliegenden Vorbericht wurde explizit formuliert: „Simon et al. [36] fordern beispielsweise eine Vollständigkeit der archivierten Tumorproben von mindestens 66 %.“ Dies ist aus unserer Sicht, wie oben ausgeführt, eine Fehlinterpretation des genauen Wortlauts der Empfehlung von Simon und Mitarbeitern. Tatsächlich krankt der Vorbericht an dieser willkürlich festgelegten Grenze, die alleine bereits zum Ausschluss nahezu aller vorliegenden prospektiv-retrospektiven Prognosestudien führen muss. Aus den oben angegebenen Gründen halten wir diesen Ausschluss von publizierten Prognosestudien mittels des 70% Kriteriums für nicht begründbar und somit auch für nicht gerechtfertigt.

Schlussfolgerung

In praxi muss zur Frage einer adjuvanten Chemotherapie mit jeder Patientin mit einem frühen Mammakarzinom eine partizipative Entscheidungsfindung vorgenommen werden. Bei dieser Entscheidungsfindung spielt das absehbare Rückfallrisiko der jeweiligen Patientin eine wesentliche Rolle. Letztlich ist es aus klinischer Sicht entscheidend, wie biomarkerbasierte Tests im Kontext mit den vorliegenden „etablierten“ Prognosefaktoren wie Tumorgröße, Nodalstatus, histologischer Differenzierungsgrad oder daraus zusammengesetzten Prognosescores (z.B. Adjuvant! Online oder klinisch-pathologische Risikoeinteilung der S3 Leitlinie) mit dem Überleben der Patientinnen assoziiert sind. Hier geben die prospektiv-retrospektiven Studien bei den von der AGO mit Level I bewerteten Genexpressionsanalysen eine eindeutige Antwort. Sowohl EndoPredict® [3] als auch Oncotype DX® [6] oder Prosigna® [2] sind in multivariaten Analysen den „traditionellen“ Prognosefaktoren überlegen und bieten daher eine bessere Abschätzung des Überlebens der Patientinnen. Auch wenn diese Studien nicht perfekt sind, zeigen sie jedoch eindeutig, dass die genannten biomarkerbasierten Tests das Überleben genauer voraussagen können als die bislang verwendeten Prognosefaktoren. Ein Ausschluss der genannten Studien, wie im Vorbericht vorgenommen, ist daher aus Sicht der Kommission Mamma der AGO nicht sinnvoll und auch für die Patientinnen, die zu Recht eine möglichst akkurate Risikoeinschätzung vor der Entscheidung über eine adjuvante Chemotherapie erwarten, nicht hilfreich.

Aus den angeführten Gründen müssen wir daher dem Fazit des Vorberichts („Für einen Nutzen bzw. Schaden einer biomarkerbasierten Entscheidung für oder gegen eine adjuvante Chemotherapie beim primären Mammakarzinom liegt kein Anhaltspunkt vor, da die Daten von 3 der 4 eingeschlossenen Studien aufgrund der hohen Unvollständigkeit nicht für die vorliegende Nutzenbewertung verwendet werden können.“) entschieden widersprechen und bitten um eine adäquate Berücksichtigung der zu diesem wichtigen Thema umfänglich vorliegenden Literatur.



Literatur

1. Peto R, Davies C, Godwin J et al. (2012) Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet (London, England)* 379(9814):432–444. doi: 10.1016/S0140-6736(11)61625-5
2. Gnant M, Filipits M, Greil R et al. (2014) Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 25(2):339–345. doi: 10.1093/annonc/mdt494
3. Filipits M, Rudas M, Jakesz R et al. (2011) A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clinical cancer research : an official journal of the American Association for Cancer Research* 17(18):6012–6020. doi: 10.1158/1078-0432.CCR-11-0926
4. Simon RM, Paik S, Hayes DF (2009) Use of archived specimens in evaluation of prognostic and predictive biomarkers. *Journal of the National Cancer Institute* 101(21):1446–1452. doi: 10.1093/jnci/djp335
5. McShane LM, Hayes DF (2012) Publication of tumor marker research results: the necessity for complete and transparent reporting. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 30(34):4223–4232. doi: 10.1200/JCO.2012.42.6858
6. Paik S, Shak S, Tang G et al. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine* 351(27):2817–2826. doi: 10.1056/NEJMoa041588
7. Sparano JA, Gray RJ, Della Makower F et al. (2015) Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *The New England journal of medicine*. doi: 10.1056/NEJMoa1510764
8. Davies C, Pan H, Godwin J et al. (2013) Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet (London, England)* 381(9869):805–816. doi: 10.1016/S0140-6736(12)61963-1
9. Martin M, Brase JC, Calvo L et al. (2014) Clinical validation of the EndoPredict test in node-positive, chemotherapy-treated ER+/HER2- breast cancer patients: results from the GEICAM 9906 trial. *Breast cancer research : BCR* 16(2):R38. doi: 10.1186/bcr3642

Die Stellungnahme wurde von

Herrn Prof. Dr. Marcus Schmidt, Klinik und Poliklinik für Geburtshilfe und Frauenkrankheiten, Universitätsmedizin der Johannes Gutenberg-Universität Mainz

erstellt.

Prof. Dr. Diethelm Wallwiener
Präsident der DGGG e.V.

Prof. Dr. Matthias W. Beckmann
Leitlinienbeauftragter DGGG

A.1.3 – Deutsche Gesellschaft für Pathologie (DGP) e. V. & Bundesverband Deutscher Pathologen (BDP) e. V.

Autoren:

- Schirmacher, Peter
- Schlake, Werner

DEUTSCHE
GESELLSCHAFT FÜR
PATHOLOGIE E.V.

Seit 1897 – dem Leben verpflichtet

Deutsche Gesellschaft für Pathologie e.V. / Robert-Koch-Platz 9 / 10115 Berlin



Institut für Qualität und Wirtschaftlichkeit im
Gesundheitswesen – IQWiG
Stellungnahme zum Vorbericht [D14-01]
Herrn Prof. Dr. med. Jürgen Windeler
Im Mediapark 8
50670 Köln

Vorab per E-Mail am 7.12.2015
An: berichte@iqwig.de

Ihr Zeichen
Unser Zeichen DGP JM Win
Datum 07.12.2015

**Gemeinsame Stellungnahme der Deutschen Gesellschaft für Pathologie e.V. (DGP) und
des Bundesverbandes Deutscher Pathologen (BDP) zum Vorbericht D14-01:
„Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische
Chemotherapie beim primären Mamma-Karzinom“**

Sehr geehrter Herr Prof. Dr. Windeler,

anbei senden wir Ihnen die gemeinsame Stellungnahme der DGP und des BDP
zum oben genannten Vorbericht (D14-01). Da die Stellungnahme von beiden
Gesellschaften unterzeichnet wurde, erhalten Sie sie in zweifacher Ausführung –
einmal von der DGP mit der Unterschrift des Vorsitzenden **Herrn Prof. Dr. med.
Peter Schirmacher**, und einmal vom BDP mit der Unterschrift des Präsidenten
Herrn Prof. Dr. med. Werner Schlake.

Die Stellungnahme und die dazugehörigen Formulare, jeweils ausgefüllt von
den Unterzeichnern, erhalten Sie **außerdem bis spätestens 13.12.2015 per
Post**.

Mit freundlichen Grüßen

Jörg Maas
Generalsekretär

**Deutsche Gesellschaft
für Pathologie e.V.**

Geschäftsstelle
Robert-Koch-Platz 9
10115 Berlin

geschaeftsstelle@pathologie-dgp.de
www.pathologie-dgp.de
T + 49 30 25760727
F + 49 30 25760729

Vorsitzender
Prof. Dr. med. Peter Schirmacher
Geschäftsführendes
Vorstandsmitglied
Prof. Dr. med. Holger Moch

Bankverbindung

IBAN: [REDACTED]
BIC: [REDACTED]
Registergericht
Amtsgericht Berlin
[REDACTED]

pathologie-dgp.de

DEUTSCHE
GESELLSCHAFT FÜR
PATHOLOGIE E.V.

Seit 1897 – dem Leben verpflichtet



Deutsche Gesellschaft für Pathologie e.V. | Robert-Koch-Platz 9 | 10115 Berlin

Institut für Qualität und Wirtschaftlichkeit im
Gesundheitswesen (IQWiG)
Stellungnahme zum vorläufigen Berichts-
plan/Amendment zum Berichtsplan D14-01
Prof. Dr. med. Jürgen Windeler
Im Mediapark 8
50670 Köln

Vorab per E-mail an: berichte@iqwig.de

Ihr Zeichen
Unser Zeichen DGP Schir, IQWiG
Datum 07.12.2015

Deutsche Gesellschaft
für Pathologie e.V.
Geschäftsstelle
Robert-Koch-Platz 9
10115 Berlin

geschaeftsstelle@pathologie-dgp.de
www.pathologie-dgp.de
T + 49 30 25760727
F + 49 30 25760729

Vorsitzender
Prof. Dr. med. Peter Schimacher
Geschäftsführendes
Vorstandsmitglied
Prof. Dr. med. Holger Moch

Bankverbindung

IBAN: [REDACTED]
BIC: [REDACTED]
Registergericht
Amtsgericht Berlin
[REDACTED]

Betreff:

Schriftliche Stellungnahme zum Vorbericht des IQWiG „Biomarkerbasierter Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mamma-Karzinom“ (Auftragsnummer: D14-01) vom 02.11.2015

Sehr geehrte Damen, sehr geehrte Herren,

die Institute für Pathologie in Deutschland haben seit mehreren Jahren Erfahrungen in der routinemäßigen Durchführung von molekularen Genexpressionsuntersuchungen bei Brustkrebs. Aufgrund unserer klinischen Erfahrungen haben wir große Bedenken gegen den o.g. Vorbericht des IQWiG zu Biomarkertests beim Mammakarzinom. Wir sind – wie auch die internationalen Experten der aktuellen St. Gallen Leitlinie 2015 - der Auffassung, dass beim hormonrezeptor-positiven Mammakarzinom oft eine unnötige und für die Patientinnen schädliche Übertherapie durchgeführt wird. Viele dieser Patientinnen benötigen keine Chemotherapie und können alleine mit einer endokrinen Therapie behandelt werden. Mit großer Sorge sehen wir, dass Genexpressionstests in vielen europäischen und außereuropäischen Ländern bereits ein fester Bestandteil der Routinediagnostik bei Brustkrebs sind, wohingegen in Deutschland hier ein immer stärkerer Rückstand entsteht.

Als Wissenschaftler, Pathologen und klinisch tätige Ärzte, die sich mit der Validierung und klinischen Anwendung von Biomarkern befassen, möchten wir uns hiermit erlauben, folgende Kommentare zum Vorbericht abzugeben:

Zu 4.1 Ergebnisse der Informationsbeschaffung:

Wir möchten darauf hinweisen, dass die Ergebnisse der Informationsbeschaffung insgesamt unvollständig sind. Im IQWiG-Bericht sind lediglich die folgenden vier Studien

pathologie-dgp.de

genannt worden: NSABP-B20-Oncotype (Paik 2006), GEICAM9906-Endopredict (Martin 2014), ABCSG8-PAM50 (Gnant 2014), sowie TransATAC-BCI (Sgroi 2013). Von diesen Studien ist dann nur eine einzige (Martin 2014) ausführlicher analysiert worden.

Es gibt mindestens acht weitere publizierte Studien, die vom IQWiG nicht berücksichtigt wurden. Dabei behandelt sich um folgende Studien:

- NSABP-B14 OncotypeDx (Paik 2004)
- SWOG8814 OncotypeDx (Albain 2010)
- TransATAC-OncotypeDx (Dowsett 2010, Cusick 2011).
- TransATAC-PAM50 (Dowsett 2013)
- TransATAC-EndoPredict (Dowsett SABCS 2015)
- ABCSG-6 (EndoPredict Filipits 2011, Dubsy 2013a, Dubsy 2013b)
- ABCSG-8 EndoPredict (Filipits 2011; Dubsy 2013a, Dubsy 2013b)
- Geicam 9906-PAM50 (Martin 2014, SABCS 2013).

Die Studien sind in der Fachwelt allgemein bekannt und anerkannt und sind den vom IQWiG eingeschlossenen Studien zumindest gleichwertig. **Die Aussagekraft des Berichtes ist somit stark eingeschränkt, da nur 30% der publizierten Studien (4 von 12 Studien) identifiziert wurden und nur eine einzige Studie ausführlicher analysiert wurde.** Somit besteht ein hohes Verzerrungsrisiko.

Zu 4.3 Extraktion der berichtsrelevanten Daten

In diesem Teil des Berichtsplans werden drei der ursprünglich identifizierten vier Studien direkt wieder ausgeschlossen. Ausschlaggebend ist dabei jeweils, dass die Biomarker-Studien weniger als 70% der Biomaterialien der ursprünglichen klinischen Studie umfassen.

Hierzu möchten wir folgendes anmerken:

- Selbstverständlich ist es wichtig, dass die Kohorte einer Biomarker-Studie insgesamt repräsentativ ist. Der vom IQWiG genannte Grenzwert von 70% ist aber vollkommen willkürlich und nicht wissenschaftlich begründet.
- Für den klinischen Nutzen eines Biomarkers ist es entscheidend, dass 1) die verwendeten Tumorproben aus einer prospektiven Studie stammen (und somit standardisiert dokumentiert wurden) und dass 2) die untersuchten Tumorproben repräsentativ für die Gesamtpopulation derjenigen Patientinnen sind, an denen diagnostische Test in der klinischen Realität angewendet wird (hier also für HR-positive HER2-negative Tumoren mit 0 bis 3 befallenen Lymphknoten).
- Die prospektiv-retrospektive Prognosestudie ist grundsätzlich unabhängig von der ursprünglichen klinischen Studie und muss daher als separate Studie bewertet werden. So ist z. B. die TransATAC Studie eine eigenständige Biomarker-Studie; sie ist statistisch unabhängig von der klinischen ATAC Studie. Auch die Fallzahlkalkulationen sind in beiden Fällen unterschiedlich: So hat die klinische Studie das Ziel, zwei verschiedene Therapien (z.B.: Tamoxifen vs. Aromatasehemmer) zu vergleichen und ist für dieses Studienziel gepowert, hier sind typischerweise sehr hohe Patientinnenzahlen notwendig. Die Prognosestudie hat aber ein statistisch

- deutlich einfacheres Ziel (Darstellung der Prognose einer Niedrigrisiko-Gruppe), daher sind hier auch geringere Fallzahlen statistisch ausreichend.
- Es gibt also keinen statistischen Grund, die drei Studien, d.h. NSABP-B20-Oncotype (Paik 2006), GEICAM9906-Endopredict (Martin 2014) und ABCSG8-PAM50 (Gnant 2014), auszuschließen. Auch die oben genannten 8 weiteren Studien sollten ebenfalls berücksichtigt werden.
 - Es muss hier betont werden, dass **eine umfassende Analyse aller vorliegenden prospektiv-retrospektiven Studien auch im Hinblick auf die Patientinnen-Sicherheit unerlässlich ist**, da sich die Sicherheit eines diagnostischen Tests u.a. darüber beurteilen lässt, wie konsistent die Testergebnisse in verschiedenen Studienkohorten sind (wenn hier Studien ausgeschlossen werden, führt dies zu Verzerrungen).
 - Zudem haben die prospektiv-retrospektiven Studien typischerweise sehr lange Nachbeobachtungszeiten und können damit auch das Risiko der Spätmetastasen beurteilen. Dieses Risiko ist gerade bei Patientinnen mit einem hormonrezeptor-positiven Karzinom sehr hoch und muss in die klinische Beurteilung der Sicherheit eines Tests mit einbezogen werden.
 - Selbstverständlich muss bei jeder Studie das Verzerrungsrisiko adäquat geprüft werden. Hierzu ist es aber notwendig, die Studienkohorte der Biomarkerstudie mit der Gesamtpopulation an luminalen Tumoren in der Bevölkerung zu vergleichen. Nur so lässt sich darstellen, dass die Ergebnisse auf die aktuelle klinische Situation übertragbar sind. Diese Vorgehensweise sollte auch vom IQWiG gewählt werden.

Zu 4.5 Ergebnisse aus RCTs

Als einzig übriggebliebene Studie ist die GEICAM9906-EndoPredict Studie in die detailliertere Analyse einbezogen worden. Allerdings möchten wir anmerken, dass das IQWiG hier eine falsche Indikation für den Biomarkertest verwendet. Im Vorbericht wird die Frage untersucht, ob Endopredict prädiktiv ist für das Ansprechen auf eine FEC Chemotherapie vs. FEC+P Chemotherapie. Dies ist nicht der Fall, aber es entspricht auch nicht der Indikation des Tests. Denn Endopredict ist ein prognostischer Test zur Identifizierung einer Niedrigrisiko-Gruppe unter alleiniger endokriner Therapie. Die einzige noch verbliebene Analyse im Vorbericht geht somit hier an der klinischen Prognose-Fragestellung vorbei.

Laufende Studien

Im Vorbericht werden sechs aktuell laufende Studien zu Biomarkertests beim Mammakarzinom genannt. Selbstverständlich werden diese Studien sehr wichtige Ergebnisse zur Bewertung der Biomarkertests liefern, erste vielversprechende Ergebnisse aus der TailorX Studie sind ja bereits publiziert worden. Wir möchten aber darauf hinweisen, dass bei einer genauen Durchsicht der Studienprotokolle schnell deutlich wird, dass die Mehrzahl dieser Studien die im Prüfplan vorgegebenen Kriterien des IQWiG offensichtlich nicht oder nicht vollständig erfüllen werden. Es handelt sich teilweise um

nicht vollständig randomisierte Studien, die nur nodal-negative Patientinnen einschließen und veränderte Cut-offs verwenden.

Die prospektiven Studien ergeben somit nur in Kombination mit den vorhandenen Daten aus prospektiv-retrospektiven Studien ein vollständiges Bild. Es ist im Interesse der Patientinnensicherheit unerlässlich, hier das Gesamtbild unter Berücksichtigung aller vorhandenen Daten zu betrachten.

Wir halten die Strategie des IQWiG, zuerst alle vorhandenen publizierten Studien zu verwerfen, um dann in Zukunft auf die Ergebnisse prospektiver Studien zu warten, für wissenschaftlich und gesundheitspolitisch fragwürdig. Denn durch die Verzögerung, die bei prospektiven Studien notwendigerweise immer wieder entsteht, würde das deutsche Gesundheitssystem auf lange Sicht vom wissenschaftlichen Fortschritt abgekoppelt. Zudem werden die prospektiven Studien in absehbarer Zeit nur 5-Jahre-Überlebensdaten liefern. Da das hormonrezeptorpositive Mammakarzinom häufig erst spät metastasiert, sind zusätzliche 10-Jahres-Daten unerlässlich, die nur aus prospektiv-retrospektiven Studien sicher gewonnen werden können.

Auch ein Vergleich von verschiedenen Tests in der identischen Patientinnenkohorte lässt sich nur in prospektiv-retrospektiven Studien durchführen, derartige Analysen sind zur vergleichenden Beurteilung der Sicherheit der verschiedenen Assays unerlässlich und werden auch in der klinischen Diskussion immer wieder gefordert.

Wir möchten abschließend darauf hinweisen, dass die vom IQWiG ausgeschlossenen Studien sowohl von den Zulassungsbehörden anderer Länder (FDA - USA, NICE – Großbritannien, INCA-Frankreich, Schweiz, etc.) als auch von praktisch allen nationalen und internationalen Expertengruppen (AGO, St. Gallen, ESMO etc.) als Grundlage für die Bewertung der Biomarkertests herangezogen werden. Auf dieser Grundlage sind derartige Test in vielen Ländern jetzt fester Bestandteil der klinischen Realität, um die Anzahl nicht indizierter Chemotherapien bei zahlreichen am ER+/HER2-Mammakarzinom (ca. 60-70% aller Mammakarzinome) erkrankten Patientinnen zu reduzieren. Dass sich dadurch eine finanzielle Entlastung der Krankenkassen ergibt, sei hier nur am Rande erwähnt. Dieser Entwicklung sollte sich das deutsche Gesundheitssystem nicht verschließen.

Als Vertreter der Deutschen Gesellschaft für Pathologie und des Bundesverbandes der Deutschen Pathologen würden wir es sehr begrüßen, wenn unsere hier geäußerten Vorschläge Gehör finden würden.

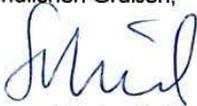


DEUTSCHE
GESELLSCHAFT FÜR
PATHOLOGIE E.V.

Seit 1897 - dem Leben verpflichtet

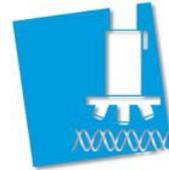
Wir sind selbstverständlich sehr gerne bereit, uns in die weitere Diskussion konstruktiv einzubringen.

Mit freundlichen Grüßen,



Prof. Dr. med. Peter Schirmacher
Vorsitzender der Deutschen
Gesellschaft für Pathologie

Dieses untereinander abgestimmte Schreiben wird parallel in identischer Form als gemeinsame Stellungnahme der Deutschen Gesellschaft für Pathologie und des Bundesverbandes der Deutschen Pathologen übermittelt.



Institut für Qualität und Wirtschaftlichkeit
im Gesundheitswesen
Stellungnahme zum vorläufigen Berichtsplan/
Amendment zum Berichtsplan D14-01
Prof. Dr. med. Jürgen Windeler
Im Mediapark 8
50670 Köln

Vorab per email an: berichte@iqwig.de

29.11.2015

Betr.: Schriftliche Stellungnahme zum Vorbericht des IQWiG „Biomarkerbasierter Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mamma-Karzinom“ (Auftragsnummer: D14-01) vom 02.11.2015

Sehr geehrte Damen, sehr geehrte Herren,

die Institute für Pathologie in Deutschland haben seit mehreren Jahren Erfahrungen in der routinemäßigen Durchführung von molekularen Genexpressionsuntersuchungen bei Brustkrebs. Aufgrund unserer klinischen Erfahrungen haben wir große Bedenken gegen den o.g. Vorbericht des IQWiG zu Biomarkertests beim Mammakarzinom. Wir sind – wie auch die internationalen Experten der aktuellen St. Gallen Leitlinie 2015 - der Auffassung, dass beim hormonrezeptorpositiven Mammakarzinom oft eine unnötige und für die Patientinnen schädliche Übertherapie durchgeführt wird. Viele dieser Patientinnen benötigen keine Chemotherapie und können alleine mit einer endokrinen Therapie behandelt werden. Mit großer Sorge sehen wir, dass Genexpressionstests in vielen europäischen und außereuropäischen Ländern bereits ein fester Bestandteil der Routinediagnostik bei Brustkrebs sind, wohingegen in Deutschland hier ein immer stärkerer Rückstand entsteht.

Als Wissenschaftler, Pathologen und klinisch tätige Ärzte, die sich mit der Validierung und klinischen Anwendung von Biomarkern befassen, möchten wir uns hiermit erlauben, folgende Kommentare zum Vorbericht abzugeben:

Zu 4.1 Ergebnisse der Informationsbeschaffung:

Wir möchten darauf hinweisen, dass die Ergebnisse der Informationsbeschaffung insgesamt unvollständig sind. Im IQWiG-Bericht sind lediglich die folgenden vier Studien genannt worden: NSABP-B20-Oncotype (Paik 2006), GEICAM9906-Endopredict (Martin 2014), ABCSG8-PAM50 (Gnant 2014), sowie TransATAC-BCI (Sgroi 2013). Von diesen Studien ist dann nur eine einzige (Martin 2014) ausführlicher analysiert worden.

Es gibt mindestens acht weitere publizierte Studien, die vom IQWiG nicht berücksichtigt wurden. Dabei behandelt sich um folgende Studien:

- NSABP-B14 OncotypeDx (Paik 2004)
- SWOG8814 OncotypeDx (Albain 2010)
- TransATAC-OncotypeDx (Dowsett 2010, Cusick 2011).
- TransATAC-PAM50 (Dowsett 2013)
- TransATAC-EndoPredict (Dowsett SABCS 2015)
- ABCSG-6 (EndoPredict Filipits 2011, Dubsky 2013a, Dubsky 2013b)
- ABCSG-8 EndoPredict (Filipits 2011; Dubsky 2013a, Dubsky 2013b)
- Geicam 9906-PAM50 (Martin 2014, SABCS 2013).

Die Studien sind in der Fachwelt allgemein bekannt und anerkannt und sind den vom IQWiG eingeschlossenen Studien zumindest gleichwertig. **Die Aussagekraft des Berichtes ist somit stark eingeschränkt, da nur 30% der publizierten Studien (4 von 12 Studien) identifiziert wurden und nur eine einzige Studie ausführlicher analysiert wurde.** Somit besteht ein hohes Verzerrungsrisiko.

Zu 4.3 Extraktion der berichtsrelevanten Daten

In diesem Teil des Berichtsplans werden drei der ursprünglich identifizierten vier Studien direkt wieder ausgeschlossen. Ausschlaggebend ist dabei jeweils, dass die Biomarker-Studien weniger als 70% der Biomaterialien der ursprünglichen klinischen Studie umfassen.

Hierzu möchten wir folgendes anmerken:

- Selbstverständlich ist es wichtig, dass die Kohorte einer Biomarker-Studie insgesamt repräsentativ ist. Der vom IQWiG genannte Grenzwert von 70% ist aber vollkommen willkürlich und nicht wissenschaftlich begründet.
- Für den klinischen Nutzen eines Biomarkers ist es entscheidend, dass 1) die verwendeten Tumorproben aus einer prospektiven Studie stammen (und somit standardisiert dokumentiert wurden) und dass 2) die untersuchten Tumorproben repräsentativ für die Gesamtpopulation derjenigen Patientinnen sind, an denen diagnostische Test in der klinischen Realität angewendet wird (hier also für HR-positive HER2-negative Tumoren mit 0 bis 3 befallenen Lymphknoten).
- Die prospektiv-retrospektive Prognosestudie ist grundsätzlich unabhängig von der ursprünglichen klinischen Studie und muss daher als separate Studie bewertet werden. So ist z. B. die TransATAC Studie eine eigenständige Biomarker-Studie; sie ist statistisch unabhängig von der klinischen ATAC Studie. Auch die Fallzahlkalkulationen sind in beiden Fällen unterschiedlich: So hat die klinische Studie das Ziel, zwei verschiedene Therapien (z.B.: Tamoxifen vs. Aromatasehemmer) zu vergleichen und ist für dieses Studienziel gepowert, hier sind typischerweise sehr hohe Patientinnenzahlen notwendig. Die Prognosestudie hat aber ein statistisch deutlich einfacheres Ziel



(Darstellung der Prognose einer Niedrigrisiko-Gruppe), daher sind hier auch geringere Fallzahlen statistisch ausreichend.

- Es gibt also keinen statistischen Grund, die drei Studien, d.h. NSABP-B20-Oncotype (Paik 2006), GEICAM9906-Endopredict (Martin 2014) und ABCSG8-PAM50 (Gnant 2014), auszuschließen. Auch die oben genannten 8 weiteren Studien sollten ebenfalls berücksichtigt werden.
 - Es muss hier betont werden, dass **eine umfassende Analyse aller vorliegenden prospektiv-retrospektiven Studien auch im Hinblick auf die Patientinnen-Sicherheit unerlässlich ist**, da sich die Sicherheit eines diagnostischen Tests u.a. darüber beurteilen lässt, wie konsistent die Testergebnisse in verschiedenen Studienkohorten sind (wenn hier Studien ausgeschlossen werden, führt dies zu Verzerrungen).
 - Zudem haben die prospektiv-retrospektiven Studien typischerweise sehr lange Nachbeobachtungszeiten und können damit auch das Risiko der Spätmetastasen beurteilen. Dieses Risiko ist gerade bei Patientinnen mit einem hormonrezeptorpositiven Karzinom sehr hoch und muss in die klinische Beurteilung der Sicherheit eines Tests mit einbezogen werden.
- Selbstverständlich muss bei jeder Studie das Verzerrungsrisiko adäquat geprüft werden. Hierzu ist es aber notwendig, die Studienkohorte der Biomarkerstudie mit der Gesamtpopulation an luminalen Tumoren in der Bevölkerung zu vergleichen. Nur so lässt sich darstellen, dass die Ergebnisse auf die aktuelle klinische Situation übertragbar sind. Diese Vorgehensweise sollte auch vom IQWiG gewählt werden.

Zu 4.5 Ergebnisse aus RCTs

Als einzig übriggebliebene Studie ist die GEICAM9906-EndoPredict Studie in die detailliertere Analyse einbezogen worden. Allerdings möchten wir anmerken, dass das IQWiG hier eine falsche Indikation für den Biomarkertest verwendet. Im Vorbericht wird die Frage untersucht, ob Endopredict prädiktiv ist für das Ansprechen auf eine FEC Chemotherapie vs. FEC+P Chemotherapie. Dies ist nicht der Fall, aber es entspricht auch nicht der Indikation des Tests. Denn Endopredict ist ein prognostischer Test zur Identifizierung einer Niedrigrisiko-Gruppe unter alleiniger endokriner Therapie. Die einzige noch verbliebene Analyse im Vorbericht geht somit hier an der klinischen Prognose-Fragestellung vorbei.

Laufende Studien

Im Vorbericht werden sechs aktuell laufende Studien zu Biomarkertests beim Mammakarzinom genannt. Selbstverständlich werden diese Studien sehr wichtige Ergebnisse zur Bewertung der Biomarkertests liefern, erste vielversprechende Ergebnisse aus der TailorX Studie sind ja bereits publiziert worden. Wir möchten aber darauf hinweisen, dass bei einer genauen Durchsicht der Studienprotokolle schnell deutlich wird, dass die Mehrzahl dieser Studien die im Prüfplan vorgegebenen Kriterien des IQWiG offensichtlich nicht oder nicht vollständig erfüllen werden. Es handelt sich teilweise um nicht vollständig randomisierte Studien, die nur nodal-negative Patientinnen einschließen und veränderte Cut-offs verwenden.

Die prospektiven Studien ergeben somit nur in Kombination mit den vorhandenen Daten aus prospektiv-retrospektiven Studien ein vollständiges Bild. Es ist im Interesse der Patientinnensicherheit unerlässlich, hier das Gesamtbild unter Berücksichtigung aller vorhandenen Daten zu betrachten.



Wir halten die Strategie des IQWiG, zuerst alle vorhandenen publizierten Studien zu verwerfen, um dann in Zukunft auf die Ergebnisse prospektiver Studien zu warten, für wissenschaftlich und gesundheitspolitisch fragwürdig. Denn durch die Verzögerung, die bei prospektiven Studien notwendigerweise immer wieder entsteht, würde das deutsche Gesundheitssystem auf lange Sicht vom wissenschaftlichen Fortschritt abgekoppelt. Zudem werden die prospektiven Studien in absehbarer Zeit nur 5-Jahre-Überlebensdaten liefern. Da das hormonrezeptorpositive Mammakarzinom häufig erst spät metastasiert, sind zusätzliche 10-Jahres-Daten unerlässlich, die nur aus prospektiv-retrospektiven Studien sicher gewonnen werden können.

Auch ein Vergleich von verschiedenen Tests in der identischen Patientinnenkohorte lässt sich nur in prospektiv-retrospektiven Studien durchführen, derartige Analysen sind zur vergleichenden Beurteilung der Sicherheit der verschiedenen Assays unerlässlich und werden auch in der klinischen Diskussion immer wieder gefordert.

Wir möchten abschließend darauf hinweisen, dass die vom IQWiG ausgeschlossenen Studien sowohl von den Zulassungsbehörden anderer Länder (FDA - USA, NICE – Großbritannien, INCA-Frankreich, Schweiz, etc.) als auch von praktisch allen nationalen und internationalen Expertengruppen (AGO, St. Gallen, ESMO etc.) als Grundlage für die Bewertung der Biomarkertests herangezogen werden. Auf dieser Grundlage sind derartige Test in vielen Ländern jetzt fester Bestandteil der klinischen Realität, um die Anzahl nicht indizierter Chemotherapien bei zahlreichen am ER+/HER2-Mammakarzinom (ca. 60-70% aller Mammakarzinome) erkrankten Patientinnen zu reduzieren. Dass sich dadurch eine finanzielle Entlastung der Krankenkassen ergibt, sei hier nur am Rande erwähnt. Dieser Entwicklung sollte sich das deutsche Gesundheitssystem nicht verschließen.

Als Vertreter der Deutschen Gesellschaft für Pathologie und des Bundesverbandes der Deutschen Pathologen würden wir es sehr begrüßen, wenn unsere hier geäußerten Vorschläge Gehör finden würden.

Wir sind selbstverständlich sehr gerne bereit, uns in die weitere Diskussion konstruktiv einzubringen.

Mit freundlichen Grüßen



Prof. Dr. med. Werner Schlake
Präsident

Dieses untereinander abgestimmte Schreiben wird parallel in identischer Form als gemeinsame Stellungnahme der Deutschen Gesellschaft für Pathologie und des Bundesverbandes der Deutschen Pathologen übermittelt.

A.1.4 – Genomic Health Deutschland GmbH

Autoren:

- Langenfeld, Matthias
- Morawski, Erwin
- Weidenauer, Holger

Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom

Stellungnahme von Genomic Health, Inc.
zum Vorbericht D14-01 des IQWiG

Genomic Health Deutschland GmbH
Erna Scheffler Straße 1a
51103 Köln

08.12.2015

Das IQWiG hat am 9. November 2015 nach dem Auftrag durch den Gemeinsamen Bundesausschuss (G-BA) vom 17. April 2014 einen Vorbericht zur Bewertung von biomarkerbasierten Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom vorgelegt. Genomic Health, Inc. möchte zu den folgenden Hauptthemen zu diesem Vorbericht Stellung nehmen:

- | | |
|--|-------|
| 1. Umsetzung des Auftrags des G-BA durch das IQWiG | S. 2 |
| • Prädiktion versus Prognose | |
| • Bewertung nicht-biomarkerbasierter Methoden | |
| 2. Bewertung und Interpretation der Studien von Martin et al. (2014), Paik et al. (2006) und TAILORx | S. 5 |
| 3. Benennung einer zusätzlichen, potentiell relevanten Studie | S. 13 |
| 4. Verweise auf andere qualitativ angemessene Unterlagen: Konkordanzstudien | S. 13 |

1 Umsetzung des Auftrags des G-BA durch das IQWiG

Laut Auftrag des G-BA soll die Bewertung von biomarkerbasierten Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom wie folgt durchgeführt werden:

„[...] soweit medizinisch-fachlich geboten und aufgrund der Erkenntnislage möglich, differenziert nach relevanten Patienten- und Interventionscharakteristika in Bezug auf patientenrelevante Zielgrößen und auf der Grundlage von Studien der Evidenzstufe I (2. Kapitel § 11 Absatz 2 der Verfahrensordnung des G-BA) [durchgeführt werden]. Insbesondere sollen bei der Bewertung folgende Aspekte berücksichtigt werden:

- *Patientencharakteristika: z. B. Alter, Rezeptorstatus (Östrogenrezeptor, Progesteronrezeptor, HER2), TNM-Klassifikation*
- *Andere Methoden zur Entscheidungsfindung zur Durchführung der adjuvanten Chemotherapie, insbesondere unter Berücksichtigung des Versorgungskontextes und im Hinblick auf Verfahren ohne Einbeziehung von Biomarkern*
- *Bewertung des Einflusses verfügbarer bzw. empfohlener adjuvanter Chemotherapieregime auf die Methoden der Entscheidungsfindung zur Durchführung der adjuvanten Therapie.*
- *Bewertung der Ergebnisse vergleichender Studien zwischen unterschiedlichen Testverfahren bzw. anderen Methoden der Entscheidungsfindung, insofern diese insbesondere bezogen auf patientenrelevante Zielgrößen sowie in Bezug auf die durch die unterschiedlichen Testverfahren ggf. formulierten Therapieempfehlungen vorliegen.“*

(Auftrag des G-BA vom 17. April 2014 an das IQWiG)

Prädiktion versus Prognose

Gemäß dem oben zitierten Auftrag des G-BA soll das IQWiG biomarkerbasierte Tests hinsichtlich deren Nutzens zur Entscheidung für oder gegen eine systemische Chemotherapie bei Patientinnen mit primärem Mammakarzinom untersuchen. Diese Bewertung bezieht sich somit auf die prädiktive Aussagekraft eines Tests, da eine Aussage über den Nutzen einer *bestimmten* Therapie getroffen werden soll. Im Gegensatz dazu steht die prognostische Aussagekraft eines Tests, also die Schätzung des *allgemeinen Rezidivrisikos* für die Patientinnen unabhängig von einer Therapie [1].

Die Methoden zur Entscheidung ob eine Chemotherapie im Gegensatz zu keiner Chemotherapie einen Nutzen beim primären Mammakarzinom hat, müssen also prädiktive Aussagekraft besitzen [1]. Eine Aussage zur Prognose kann zusätzlich getroffen werden, ist aber nicht von primärem Interesse des vorliegenden Auftrags des G-BA. Dennoch schließt das IQWiG Prognosestudien mit in die Bewertung ein.

Der biomarkerbasierte Test Oncotype DX[®] besitzt nachweislich prädiktive Aussagekraft (z.B. NSABP-B-20-Studie, Paik et al., 2006 [2]). Das IQWiG hat die randomisierte, kontrollierte Studie von Paik et al. im Rahmen seiner Recherche zwar identifiziert und auch als relevant bewertet aber aufgrund der geringen Probenzahl im Vergleich zu den ursprünglich rekrutierten Patientinnen wurden die Ergebnisse letztlich nicht berücksichtigt. Dennoch wurde in dieser Studie eine statistisch signifikante Interaktion zwischen dem ermittelten Recurrence Score[®] und der Behandlung mit Chemotherapie ermittelt ($p=0,038$), sodass der Nutzen einer Chemotherapie nachweislich mittels Bestimmung des Recurrence Score[®] verlässlich vorhergesagt werden kann. Diese Analyse wurde auf Nachfrage des IQWiG auch für die Subgruppe an Patientinnen mit intermediärem Rezidivrisiko nach St. Gallen 2007 und mit HER2-negativem Status durchgeführt. Auch hier wurde eine signifikante Interaktion zwischen Recurrence Score[®] und Behandlung mit Chemotherapie nachgewiesen.

Somit liegt aus unserer Sicht umfassende Evidenz für den prädiktiven Nutzen des Recurrence Score[®] hinsichtlich der Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom vor. Auf weitere Studien zu Oncotype DX[®], die diesen Umstand unterstützen und bestätigen, wird unter Punkt 2 dieser Stellungnahme detailliert eingegangen.

Bewertung nicht-biomarkerbasierter Methoden zur Entscheidung für oder gegen eine adjuvante Chemotherapie

Der Auftrag des G-BA beinhaltet explizit auch die Untersuchung nicht-biomarkerbasierter Tests für die Entscheidung für oder gegen eine adjuvante Chemotherapie:

„[...] Andere Methoden zur Entscheidungsfindung zur Durchführung der adjuvanten Chemotherapie, insbesondere unter Berücksichtigung des Versorgungskontextes und im Hinblick auf Verfahren ohne Einbeziehung von Biomarkern“. (Auftrag des G-BA vom 17. April 2014 an das IQWiG)

Diese Bewertung nicht-biomarkerbasierter Tests hinsichtlich einer Entscheidung für oder gegen eine adjuvante Chemotherapie wurde in dem Vorbericht des IQWiG nicht umgesetzt.

Traditionelle nicht-biomarkerbasierte Tests zur Entscheidung für oder gegen eine Chemotherapie sind klinisch-pathologische Parameter. Dies sind die Tumorgöße, der

Nodalstatus (Tumorbefall der Lymphknoten), das sogenannte Grading des Tumors oder die Bestimmung des Zellproliferationsmarkers Ki-67.

Die Größe des Tumors und dessen Nodalstatus liefern zwar prognostische Informationen, doch diese sind der prognostischen und prädiktiven Aussagekraft des Recurrence Score® weit unterlegen. Das standardmäßig durchgeführte Grading des Tumors hat eine hohe Variabilität je nachdem welche Person die Untersuchung durchgeführt hat (Interobserver-Variabilität) (Abbildung 1).

		Histologisches Grading, zentrales Labor			insgesamt
		1	2	3	
Histologisches Grading, lokales Labor	1	28.0 %	64.6 %	7.3 %	7,2 %
	2	4.4 %	70.8 %	24.8 %	70,0 %
	3	0.8 %	34.7 %	64.5 %	22,8 %
insgesamt		5.2 %	62.2 %	32,6 %	100 %

Abbildung 1: Diskordanz der Ergebnisse des Gradings: Lokale versus zentrale Pathologie (Immunhistochemie). [Quelle: Nitz U et al. SABCS 2014; Poster P4-11-01 [3].]

Ähnlich verhält es sich auch bei der immunhistochemischen Bestimmung des Proteins Ki-67, wie bereits mittels verblindeten, kontrollierten Studien gezeigt wurde [4, 5]. Im Gegensatz zur etablierten prognostischen Aussagekraft dieser herkömmlichen Faktoren zur Beurteilung eines Brustkrebstumors, gibt es keine Evidenz für deren prädiktiven Wert. In der Studie von Yerushalmi et al. (2010) wurde beispielsweise die vorhandene Evidenz für Ki-67 hinsichtlich dessen prognostischer und prädiktiver Qualität zusammengefasst und explizit darauf hingewiesen, dass derzeit keine prädiktive Aussage mittels Ki-67 hinsichtlich eines Vorteils einer Chemotherapie gegenüber einer endokrinen Therapie getroffen werden kann [6].

Dementsprechend werden diese herkömmlichen klinisch-pathologischen Parameter auch nicht von der Arbeitsgemeinschaft Gynäkologische Onkologie (AGO) als prädiktiv anerkannt. Im Gegensatz dazu wird aber Oncotype DX® in der aktuellen Leitlinie der AGO aufgrund dessen prädiktiver Informationen hinsichtlich des Nutzens einer adjuvanten Chemotherapie bereits empfohlen [7]. Ebenso erkennen die internationalen Leitlinien der American Society of Clinical Oncology (ASCO), des National Comprehensive Cancer Network (NCCN), von St. Gallen und der European Society for Medical Oncology (ESMO) die prädiktive Aussagekraft von Oncotype DX® an [8-11].

Das IQWiG schließt selbst nur Patientinnen mit primärem HER2/neu-negativem Mammakarzinom und 0 bis 3 befallenen Lymphknoten ein, da bei diesen Patientinnen eine Entscheidung für oder gegen eine adjuvante Chemotherapie unklar sei. Es liegt also – im Gegensatz zu Oncotype DX® – keine prädiktive Aussagekraft für diese herkömmlichen Parameter hinsichtlich des Nutzens einer Chemotherapie vor und dies wurde vom IQWiG und G-BA bisher noch nicht näher untersucht.

2 Bewertung und Interpretation der im Vorbericht eingeschlossenen Studien

Das IQWiG hat insgesamt vier Studien in seinen Vorbericht eingeschlossen: die zwei randomisierten, kontrollierten Studien von Paik et al., 2006 (Oncotype DX) und Martin et al., 2014 (EndoPredict) sowie die Prognosestudien von Gnant et al., 2014 (PAM50 Risk of Recurrence Score) und Sgroi et al., 2013 (Breast Cancer Index). Lediglich die Studie von Martin et al., 2014 wird vom IQWiG als relevant erachtet, da diese einen Probenanteil von knapp unter 70 % der ursprünglichen Patientinnen (GEICAM-Studie) aufweist.

2.1 Martin et al., 2014

Die Studie von Martin et al., 2014 zu EndoPredict wird zwar vom IQWiG als relevant bewertet, es liegen aber keine signifikanten Unterschiede vor, weshalb mittels der vorliegenden Evidenz zu EndoPredict keine Aussage hinsichtlich einer Entscheidung für oder gegen eine adjuvante Chemotherapie getroffen werden kann.

Generell ist aber zunächst zu beachten, dass in der Studie von Martin et al., 2014, lediglich zwei verschiedene Chemotherapieregime miteinander verglichen werden. Eine potentiell prädiktive Aussage des Tests EndoPredict kann mit diesem Studiendesign somit nur schwer getroffen werden, da lediglich der Nutzen einer *anderen* Chemotherapie untersucht wird, nicht aber der zu untersuchende Nutzen einer Chemotherapie im Vergleich zu *keiner* Chemotherapie.

Wie unter Punkt 1 bereits ausführlich dargestellt wurde, beinhaltet der Auftrag des G-BA die Bewertung der prädiktiven Aussage der vorliegenden Tests. Um sich für oder gegen eine Chemotherapie entscheiden zu können – also um den möglichen Nutzen einer Chemotherapie für Patientinnen mit ER-positivem Brustkrebs schätzen zu können – muss der Vergleich eines Studienarms mit Chemotherapie-Behandlung in Kombination mit endokriner Therapie mit einem Studienarm mit endokriner Therapie allein, zwingend vorliegen. Dies trifft auf das Studiendesign von Martin et al. nicht zu. Daher ist diese Studie für die Beantwortung der Fragestellung nicht geeignet, denn beide Studienarme erhielten eine Chemotherapie.

Wir bitten daher um eine entsprechende Anpassung der Bewertung im Bericht des IQWiG.

2.2 Paik et al., 2006

Die Studie von Paik et al., 2006, die auf der NSABP-B20-Studienpopulation beruht, bildet die Fragestellung des Nutzens einer Chemotherapie ab. Ein Studienarm erhält nur eine endokrine Therapie wohingegen der andere Studienarm die endokrine Therapie in Kombination mit einer Chemotherapie erhält. Bei allen Patientinnen wurde mittels Oncotype DX[®] der Recurrence Score[®] und dessen Interaktion mit der Chemotherapie-Behandlung ermittelt.

Es liegen statistisch signifikante Ergebnisse zur Interaktion vom Recurrence Score[®] und der Anwendung einer Chemotherapie im Vergleich zu keiner Chemotherapie (beides in Kombination mit einer Hormontherapie) vor, wie Genomic Health auf Nachfrage des IQWiG auch im Rahmen einer Subgruppenanalyse hinsichtlich der Zielpopulation dargestellt hat. Diese Ergebnisse werden allerdings aufgrund der geringen Probenanzahl im Vergleich zu den ursprünglich rekrutierten Patientinnen vom IQWiG nicht berücksichtigt.

So werden vom IQWiG nur Studien anerkannt, deren Proben mindestens 70 % der ursprünglichen Patientenpopulation beinhalten. Die wissenschaftliche Rationale hinter dieser Festlegung für biomarkerbasierte Studien ist allerdings unklar und Genomic Health ist keine publizierte Evidenz für einen solchen Grenzwert bekannt. Unsere Position ist daher, dass dieser allgemeine Grenzwert nicht valide ist, sondern verschiedene Voraussetzungen erfüllt sein müssen, unter denen auch ein geringerer Prozentsatz vorhandener Proben ausreichend für eine valide Aussage einer biomarkerbasierten Studie ist.

Simon et al. adressieren in ihrer wegweisenden Publikation von 2009 die Herausforderungen bei der Verwendung von archivierten Proben zur Bestimmung von prognostischen und prädiktiven Biomarkern [12]. Hierbei definieren sie folgende Voraussetzungen, die zutreffen sollten:

- 1) Die verfügbaren Proben des Archivs sollten repräsentativ sein.
 - a) Die korrelierende Biomarker-Studie sollte mindestens zwei Drittel der Gesamtpopulation der Ursprungsstudie beinhalten **oder**
 - b) die Proben sollten so ausgewählt worden sein, dass ein Auswahlbias vermieden wird.

Insbesondere der zweite Teil der Anforderungen (b) ist äußerst relevant und wird daher im Folgenden näher diskutiert. Generell gibt es zwei Faktoren, die in einer prospektiv-retrospektiven Studie zu einem Auswahlbias führen können. Der erste und wichtigste ist die Voraussetzung, dass die randomisierten Studienarme bezüglich der Baseline-Charakteristika der Patienten vergleichbar sind – denn dies ist das Ziel einer Randomisierung. Hierbei ist es allerdings *nicht* notwendig, dass „[...] von einem zufallsbedingten, krankheitsunabhängigen Fehlen der Tumorproben ausgegangen werden [kann].“, wie das IQWiG im Vorbericht postuliert. Denn nicht eine generelle Selektion, sondern eine *unterschiedliche* Selektion zwischen den randomisierten Armen muss vermieden werden. So stellt eine geringe Verfügbarkeit von Tumorproben kein Problem dar, solange die Verfügbarkeit der Proben in beiden randomisierten Armen ähnlich ist. In einem solchen Fall würde die Vergleichbarkeit der Proben hinsichtlich der Baseline-Charakteristika in den beiden Behandlungsarmen, die mittels Randomisierung gewährleistet wurde, aufrechterhalten werden.

Abschließend ist daher festzuhalten, dass eine definierte Grenze von 70 % verfügbarer Proben der Ursprungsstudie nicht sinnvoll ist, sondern dass die Gründe des Ausschlusses der Patientenproben und die Vergleichbarkeit der verfügbaren Proben in den randomisierten Behandlungsarmen untersucht werden müssen. Denn nicht eine generelle Selektion sondern nur eine *unterschiedliche* Selektion der Proben zwischen den Studienarmen würde die Aussagekraft der Biomarker-Studie verzerren.

Dies lässt sich auf die Studie von Paik et al., 2006 übertragen. So wurden in der Ursprungsstudie, NSABP-B-20, von 1988 bis 1993 Patientinnen rekrutiert und damals war es üblich, dass nur von einem kleinen Teil der 2299 Patientinnen Gewebeproben gesammelt wurden. So wurden von der Mehrheit der Patientinnen keine Gewebeproben oder Schnitte in die NSABP-Gewebebank aufgenommen (ungefähr 1450 Fälle bzw. 63 % der 2299 Patienten der Ursprungsstudie). Darüber hinaus wurden 179 Gewebeproben zwar ursprünglich entnommen, dann aber schon in früheren NSABP-Studien verbraucht. Die Identifizierung der noch vorhandenen Gewebeproben wurde von der NSABP ohne Beteiligung von Genomic Health durchgeführt. Letztlich wurden 651 Patientenproben in die B-20-Biomarker-Studie von

Paik et al. eingeschlossen – dies entspricht 79 % der verfügbaren Proben der Ursprungsstudie. Wie wir bereits dem IQWiG dargelegt haben, unterscheiden sich die Baseline-Charakteristika der Patientinnen der Biomarker-Studie, die mit endokriner Therapie behandelt wurden, nicht maßgeblich von denen der Patientinnen im Behandlungsarm mit Chemotherapie in Kombination mit Hormontherapie (s. auch Antwort auf eine Nachfrage des IQWiG vom Juni 2015, Tabelle 5.1 und 5.2). Es also eine ausreichende Vergleichbarkeit der Proben in den beiden Studienarmen gewährleistet und somit ist die Biomarker-Studie von Paik et al. trotz des relativ geringen Anteils an Gewebeproben valide.

Ergänzend kann die Vergleichbarkeit der Biomarker-Studienpopulation mit der Ursprungsstudienpopulation von Relevanz für einen möglichen Selektionsbias sein. In dem vorliegenden Fall ist es aber von größerer Bedeutung, dass die Studienpopulation der Biomarker-Studie repräsentativ für die Patientenpopulation ist, die letztlich von dem Biomarker-Test profitieren soll. Dies trifft auf die B-20-Studie von Paik et al. zu, wie wir bereits in einer früheren Antwort an das IQWiG dargelegt haben (Juni 2015, Tabelle 4.1, 4.2 und 4.3).

Abschließend möchten wir erneut darauf hinweisen, dass die Ergebnisse der Studie von Paik et al. robust und statistisch signifikant sind und Oncotype DX[®] bzw. dessen Recurrence Score[®] dafür geeignet sind, eine valide Entscheidung für oder gegen eine adjuvante Chemotherapie zu treffen. Oncotype DX[®] stellt die beste Evidenz in diesem Gebiet dar. Der Bericht des IQWiG sollte daher dahingehend geändert werden, dass die NSABP-B-20-Studie von Paik et al. in die Bewertung mit einbezogen wird.

2.3 TAILORx

Darüber hinaus verweist das IQWiG auf potentiell relevante Daten der derzeit noch laufenden Studien TAILORx, planB und MINDACT hin, deren Ergebnisse 2016 erwartet werden. Wie das IQWiG auch selbst positiv erwähnt hat, liegen für TAILORx (Oncotype DX[®]) bereits erste Ergebnisse vor, die im Folgenden kurz dargelegt werden.

Studiendesign und Durchführung

Die TAILORx-Studie wird vor der ECOG-ACRIN Cancer Research Group durchgeführt, eine kooperative Gruppe die als Zusammenschluss der Eastern Cooperative Oncology Group (ECOG) und des American College of Radiology Imaging Network (ACRIN) entstanden ist. Das Monitoring der Studie wird von einem unabhängigen ECOG-ACRIN Daten-Monitoring-Komitee (DMC) durchgeführt. Dies ist zusammengesetzt aus einer unabhängigen Expertengruppe, die nicht dem Studienteam angehört und zwei Mal im Jahr zusammenkommt.

Die Studie besteht aus drei Patientenpopulationen (Abbildung 2):

- Patienten mit einem Recurrence Score[®] von < 11 erhalten eine ausschließlich endokrine Therapie
- Patienten mit einem Recurrence Score[®] von 11 – 25 werden randomisiert auf eine ausschließlich endokrine Therapie oder eine Kombinationsbehandlung aus Chemotherapie und Hormontherapie
- Patienten mit einem Recurrence Score[®] von > 25 erhalten eine Kombinations-therapie aus Chemotherapie und Hormontherapie

Entsprechend des Studienprotokolls prüft das ECOG-ACRIN-DMC zunächst die Daten der randomisierten Studiengruppe (Recurrence Score[®] von 11 bis 25) – aber erst wenn mindestens 25 % der geplanten 835 Disease-Free-Survival-Ereignisse aufgetreten sind. Die Ergebnisse dieser geplanten Interimsanalysen werden hinsichtlich der Nichtunterlegenheit gegenüber präspezifizierten Kriterien für einen verfrühten Abbruch der Studie evaluiert.

Das ECOG-ACRIN-DMC prüft außerdem die Ergebnisse der Patientinnen mit niedrigem Rezidivrisiko (Patientinnen mit einem Recurrence Score[®] von 0 bis 10), welche ausschließlich mit Hormontherapie behandelt wurden. Hierbei war das Studienziel zu validieren, dass diese Patientinnen so geringe Ereignisraten haben, dass eine adjuvante Chemotherapie nur mit sehr geringer Wahrscheinlichkeit einen absoluten Nutzen hat. 75 Fernrezidiv-Ereignisse wurden im Studienprotokoll als nötige Information für die Analyse festgelegt.

Beim vierten Treffen des ECOG-ACRIN-DMC waren die Kriterien für eine Auswertung der Ergebnisse der randomisierten Gruppe noch nicht erfüllt. Allerdings empfahl das Komitee die Auswertung und Veröffentlichung der Ergebnisse der Patientinnen mit niedrigem Recurrence Score[®]. Diese Daten wurden anschließend dem Vorstand der TAILORx-Studie präsentiert und letztlich veröffentlicht [13].

Das DMC ist damit beauftragt, die Sicherheit der Studienpatienten als auch die durchgehende Validität und den wissenschaftlichen Wert der Studie zu gewährleisten. Während der laufenden Studie teilt das Komitee die Gründe für seine Empfehlungen nicht vollständig mit, um die Integrität der Studie nicht zu beeinflussen. Daher ist auch die Rationale für die Empfehlung der früheren Veröffentlichung der Ergebnisse der Studienpopulation mit geringem Recurrence Score[®] nicht vollständig bekannt. Es ist aber anzunehmen, dass eine Situation mit folgenden Bedingungen zutrifft:

- Die Kriterien zum frühen Abbruch der randomisierten Gruppe waren nicht erfüllt.
- Das DMC war der Meinung, dass die Ergebnisse der Patientengruppe mit geringem Recurrence Score[®] so umfassend waren, dass eine Veröffentlichung zwingend erforderlich war.
- Das DMC war der Meinung, dass die verfrühte Veröffentlichung der einer Patientengruppe die wissenschaftliche Integrität der randomisierten Studiengruppe nicht beeinflusst.
- Das DMC war der Meinung, dass kein bedeutsames Risiko für die randomisierte Patientengruppe oder für das öffentliche Gesundheitswesen (Oncotype DX[®] ist weit verbreitet verfügbar) vorliegen wird, wenn der randomisierte Teil der Studie weiterläuft und deren Interimsergebnisse nicht mitgeteilt werden.

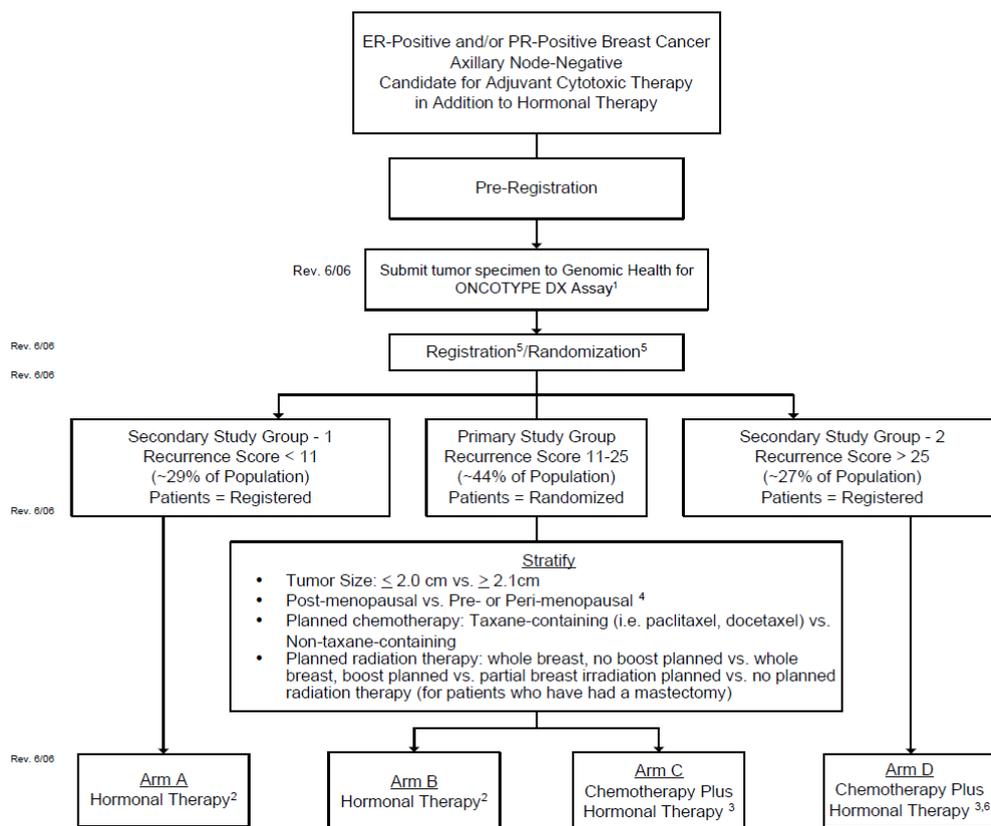


Abbildung 2: Flussdiagramm zum Studiendesign der Studie TAILORx

Veröffentlichung zu TAILORx

Die Publikation der Ergebnisse der Patientinnen mit einem geringen Rezidivrisiko (Recurrence Score[®] von 0 bis 10), die ausschließlich mit endokriner Therapie behandelt wurden, wurde bereits dieses Jahr veröffentlicht [13]. Dort wird anhand der geringen Rezidivrate dieser Patientinnen gezeigt, dass diese Patientengruppe keine Chemotherapie benötigt und eine rein endokrine Therapie ausreichend ist. Hierdurch konnten weitreichende Nebenwirkungen vermieden werden, denn nach herkömmlichen pathologischen Parametern wären die Patientinnen höchstwahrscheinlich mit einer adjuvanten Chemotherapie behandelt worden.

Die Patientengruppe der Studie, die einen mittleren Recurrence Score[®] aufwies, wurde randomisiert: Die eine Hälfte wurde mit endokriner Therapie in Kombination mit Chemotherapie behandelt und die andere Gruppe allein mit endokriner Therapie (Abbildung 2). Wie oben ausgeführt, wurden in dieser Patientenpopulation die Grenzen für eine frühe Auswertung, um Nichtunterlegenheit zu zeigen oder abzulehnen, noch nicht erreicht. Dies deutet aber nicht auf eine geringere Qualität der Ergebnisse hin; die Möglichkeit einer Nichtunterlegenheit besteht weiterhin. Vielmehr liegt eine geringere Ereignisrate vor als erwartet, da die nötige Anzahl der Ereignisse für eine vollständige Auswertung noch nicht aufgetreten ist. Dies deutet darauf hin, dass tatsächlich nicht alle dieser Patientinnen mit einem mittleren Recurrence Score[®] eine Chemotherapie benötigen.

Die Ergebnisse dieser randomisierten Patientengruppe wurden kürzlich zur Veröffentlichung eingereicht. Die Daten der bereits veröffentlichten Patientengruppe werden durch diese neuen Ergebnisse bestätigt und die prädiktive Aussagekraft des Recurrence Score[®] wird somit erneut belegt. Daher wird die Studie TAILORx im Folgenden näher dargestellt.

Definition des Endpunkts Rezidivfreiheit

Wie oben bereits beschrieben, liegen für TAILORx bereits publizierte Ergebnisse der Patientengruppe der rein endokrin behandelten, nicht randomisierten Subgruppe mit einem Recurrence Score[®] von 0 bis 10 vor [13]. Hier liegt die 5-Jahres-Rezidivrate, definiert als ipsilateraler Brustkrebs, lokoregionale und Fernrezidive sowie Tod, bei 1,3 %. Die Rezidivfreiheit beträgt somit 98,7 %.

Im Gegensatz zur Studienauswertung schließt das IQWiG in seinem Vorbericht allerdings auch 15 berichtete Fälle von neuem, kontralateralem Brustkrebs mit in die Berechnung der Rezidivrate ein und gibt daher eine höhere Rate von ungefähr 2 % in 5 Jahren an – was immer noch eine sehr geringe Rezidivrate wäre. Genomic Health ist es nicht möglich, die genaue Ereignisrate für Rezidive und kontralaterale Erkrankungen zu ermitteln, da die Durchführung der TAILORx-Studie uns nicht obliegt. Wir erhalten lediglich die Proben zur Bestimmung des Recurrence Score[®] und sind sonst nicht an der Studiendurchführung und Auswertung beteiligt.

Allerdings steht das Vorgehen des IQWiG im Gegensatz zu den etablierten, standardisierten Definitionen für Wirksamkeitsendpunkte in adjuvanten Brustkrebsstudien (STEEP, *standardized definitions for efficacy end points* [14]). Außerdem wurde bereits gezeigt, dass eine Chemotherapie nur einen marginalen Einfluss auf das spätere Auftreten von kontralateralem Brustkrebs hat. So wurde in einer Studie der *Early Breast Cancer Trialists' Collaborative* (EBCTCG) nach 15 Jahren Beobachtungszeit von nahezu 29.000 Patientinnen (randomisiert für entweder Chemotherapie oder keine Chemotherapie) kein Effekt einer Chemotherapie auf das Risiko eines zweiten primären Tumors oder auf nicht-Brustkrebs-Mortalität beobachtet [15].

Zusammenfassend ist die Freiheit von Fernrezidiven somit der relevante Endpunkt bei der Entscheidung für oder gegen eine adjuvante Chemotherapie, denn dies ist der Endpunkt, auf den die Chemotherapie einen signifikanten Einfluss zeigt. Da die Chemotherapie nur einen marginalen Einfluss auf das Risiko kontralateraler Rezidive beim frühen Mammakarzinom zeigt, sollten diese in der vorliegenden Fragestellung nicht mit zu dem Endpunkt „Rezidive“ hinzugezählt werden.

Beobachtungszeitraum

Das IQWiG schreibt in seinem Vorbericht:

„Wenn sich diese 5-Jahres-Ergebnisse [der Studie TAILORx] als ausreichend stabil erweisen, wären die Voraussetzungen des Einsatzes des Tests zur Identifikation einer Gruppe von Patientinnen, die allenfalls marginal von einer Chemotherapie profitieren können, erfüllt.“ (Vorbericht S. 13)

Im Allgemeinen ist es wichtig, bei der Bewertung der ersten Ergebnisse von TAILORx zu berücksichtigen, dass eine Chemotherapie ihren größten Effekt auf die Therapie von Brustkrebs in den ersten fünf Jahren nach der Behandlung hat. Die umfassende Meta-Analyse der *Early Breast Cancer Trialists' Collaborative* (EBCTCG) mit 8575 Patientinnen hat gezeigt, dass die absolute Reduktion der Rezidivrate mittels Anthrazyklin-basierter Chemotherapie nach fünf Jahren bei 8,5 % und nach zehn Jahren bei

8,0 % im Vergleich zu keiner Chemotherapie lag [15]. Die Rezidivrate ist somit nach zehn Jahren nur marginal geringer als nach fünf Jahren, was darauf zurückzuführen ist, dass die Mehrheit der Rezidive innerhalb der ersten fünf Jahre auftritt. Natürlich würde man bei einer längeren Beobachtungszeit generell auch das Auftreten von Rezidiven erwarten, aber es ist sehr unwahrscheinlich, dass diese späten Rezidive durch die adjuvante Chemotherapie hätten verhindert werden können. Dies wird durch die stabile Rezidivrate nach fünf und zehn Jahren nach Therapie verdeutlicht.

Darüber hinaus werden klinische Medikamentenstudien in der Indikation Brustkrebs ebenfalls mit einem Beobachtungszeitraum von fünf Jahren akzeptiert. Ein aktuelles Beispiel ist die frühe Nutzenbewertung nach § 35a von Pertuzumab¹ vom 1. Dezember 2015, die auf der Studie NeoSphere beruht, welche einen maximalen Beobachtungszeitraum von fünf Jahren nach Randomisierung der letzten Patientin und den sekundären Endpunkt „Rezidive“ aufweist [16]. Die Studiendauer wurde vom IQWiG akzeptiert.

Daher ist der klinisch relevante Zeitraum für die Beurteilung der Wirksamkeit einer Chemotherapie fünf Jahre nach Behandlung erreicht. Studien mit einem Beobachtungszeitraum von fünf Jahren müssen somit bei der Bewertung der biomarkerbasierten Tests eingeschlossen und berücksichtigt werden.

Konsistenz der Ergebnisse

Die ersten Ergebnisse der Studie TAILORx haben sich als stabil und konsistent zu verschiedenen anderen Studiendesigns und Populationen erwiesen, welche im Nachfolgenden kurz dargelegt werden (Tabelle 1). Bei den bereits publizierten Ergebnissen der Studie TAILORx hinsichtlich Patientinnen mit einem Recurrence Score[®] von 0 bis 10 und einer endokrinen Therapie lag die Fernrezidivrate bei 0,7 %, die Rate der Brustkrebsrezidive bei 1,3 % und die Todesrate bei 2% nach 5 Jahren [13].

Eine weitere prospektive, noch laufende Studie zu Oncotype DX[®] ist die von der Westdeutschen Studiengruppe (WSG) durchgeführte Studie planB. Erste Ergebnisse dieser vom IQWiG in dessen Bericht ebenfalls erwähnten Studie wurden bereits auf dem European Cancer Congress (ECC) 2015 präsentiert [17]. In dieser Studie wurde nach drei Jahren eine Ereignisrate von 1,7 % bei Patienten mit hohem Rezidivrisiko (Nodalstatus 0 (NO) und zusätzliche Risikofaktoren oder N1) aber geringem Recurrence Score[®] und endokriner Therapie beobachtet (Tabelle 1). Dies zeigt trotz einer Beobachtungszeit von unter 5 Jahren eine sehr geringe Rezidivrate bei Patienten mit nach herkömmlichen Kriterien hohem Rezidivrisiko aber niedrigem Recurrence Score[®]. Das Manuskript wurde zur Veröffentlichung eingereicht und eine Publikation liegt daher noch nicht vor.

Der Recurrence Score[®] wurde außerdem in bevölkerungsweiten Kohortenstudien untersucht [18, 19]. Diese haben den Vorteil, dass hier die Modalitäten der klinischen Praxis vorliegen und somit die Ergebnisse der kontrollierten Studien (TAILORx und planB [13, 17]) in der klinischen Realität validiert werden können (siehe Tabelle 1).

¹ Beginn des Verfahrens der frühen Nutzenbewertung nach § 35a von Pertuzumab im neuen Anwendungsgebiet am 1. September 2015; das neue Anwendungsgebiet lautet „Neoadjuvante Behandlung von erwachsenen Patienten mit humanem epidermalem Wachstumsfaktorrezeptor-2(HER2)-positivem lokal fortgeschrittenem, entzündlichem oder frühem Brustkrebs mit hohem Rezidivrisiko“.

So lag die Fernrezidivrate der Patientinnen einer Kohorte aus Israel mit einem Recurrence Score[®] von < 18 bei nur 1,4 %, und die Brustkrebs-spezifische Mortalität betrug 0,1% (Stemmer et al., 2015, siehe Tabelle 1). Die Patientinnen wurden entsprechend ihrer Ergebnisse des Recurrence Score[®] behandelt.

In einer noch größeren Kohorte aus den USA, bei der 21.023 Patientinnen mit frühem ER+, HER- und N0-Brustkrebs diagnostiziert wurden und die einen geringen Recurrence Score[®] aufwiesen sowie keine oder keine bekannte Chemotherapie bekamen, lag die Brustkrebs-spezifische Mortalität bei nur 0,4 % in fünf Jahren (Shak et al., 2015; siehe Tabelle 1).

Tabelle 1: Vergleich der Ergebnisse prospektiver, kontrollierter Studien und bevölkerungsweiter Kohortenstudien hinsichtlich des Recurrence Score[®] und der Mortalität.

	RS	Rekrutierungszeitraum	Patienten	n	Behandlung	3 J. ER	5 J. BCSM	5 J. Mortalität
Prospektive kontrollierte Studien								
TAILORx (Sparano 2015 [13])	0-10	2006 – 2010	ER+, HER2-, N0	1.626	endokrin			2,0%
planB (ECC 2015 [17])	0-11	2009 – 2011	ER+, HER2-, N0-1	404	endokrin	1,7%		
Bevölkerungsweite Kohortenstudien								
Stemmer et al. (SABCS 2015 [19])	0-17	2004 – 2010	ER+, HER2-, N0	813	endokrin (99%)		0,1%	
Shak et al. (SABCS 2015 [18])	0-17	2004 – 2011	ER+, HER2-, N0	21.023	keine CT		0,4%	

RS: Recurrence Score[®] Ergebnis; ER: Ereignisrate; BCSM: Brustkrebs-spezifische Mortalität (*breast cancer specific mortality*); CT: Chemotherapie; J.: Jahre

Somit zeigen sich konsistent geringe Brustkrebs-assoziierte Mortalitäts- und Rezidivraten bei Patientinnen mit einem geringen Recurrence Score[®].

Die letztgenannten Ergebnisse sind mit denen der kontrollierten Studien TAILORx und planB vergleichbar, deren Mortalitätsraten ebenfalls in den ersten 5 Jahren, ein Zeitraum auf den eine Chemotherapie nachweislich einen Einfluss hat, weit unter 5 % lagen (Tabelle 1). Diese konsistenten, prospektiven Ergebnisse sowohl aus klinischen Studien als auch von Bevölkerungs-basierten Studien der klinischen Praxis stimmen mit den klinischen Validitätsstudien des Oncotype DX[®] überein und bestätigen auch die ersten Ergebnisse der noch laufenden Studie TAILORx.

Insgesamt sind die Ergebnisse der diversen Studien mit unterschiedlichen Studiendesigns und Populationen konsistent und belegen die hohe prädiktive Aussagekraft von Oncotype DX[®], die daher bei der Bewertung des IQWiG nicht ignoriert werden darf.

3 Benennung von zusätzlichen, im Vorbericht nicht genannten, relevanten Studien

Im Vorbericht des IQWiG wird die Biomarker-Studie der transATAC-Population, die von Sgroi et al. 2013 publiziert wurde, eingeschlossen. Gleichzeitig wurde die Publikation von Dowsett et al. von 2010 ausgeschlossen, die aber auf der gleichen Ursprungstudie beruht [20].

Dieses Vorgehen bleibt daher unklar und wir möchten hiermit um eine Erläuterung bitten, warum letztere Studie der transATAC-Population nicht in die Bewertung miteingeschlossen wurde.

4 Verweise auf andere qualitativ angemessene Unterlagen, einschließlich einer Begründung für ihre jeweilige fragestellungsbezogene Eignung und Validität: Konkordanzstudien

Laut Auftrag des G-BA sollen Studien, die sich mit Fragen der Wirtschaftlichkeit der Methode beschäftigen, ebenfalls dem G-BA zur weiteren Bewertung übermittelt werden. Genomic Health hat einen Vergleich bezüglich der Konkordanz verschiedener biomarkerbasierter Tests anhand von Literatur durchgeführt, der im Folgenden dargestellt wird.

Wie bereits oben detailliert ausgeführt wurde, ist Oncotype DX[®] der einzige derzeit verfügbare Test, der nachweislich eine hohe prädiktive Aussagekraft besitzt. So wurde mittels prospektiver Studien ein signifikanter Langzeitnutzen des Tests für Brustkrebspatientinnen aufgezeigt.

Konkordanzstudien haben gezeigt, dass Oncotype DX[®] sich sowohl von EndoPredict [21], als auch von MammaPrint [22] und Prosigna [23] unterscheidet. Die ersten Ergebnisse der OPTIMA-Studie haben außerdem ergeben, dass Oncotype DX[®] mehr Patienten einem niedrigen Rezidivrisiko zuordnet als andere Tests und diesen Patienten eine Chemotherapie bei unverändertem Rezidivrisiko erspart werden kann [24] (Tabelle 2).

Der EndoPredict-Test war nicht in die OPTIMA-Studie eingeschlossen. Dennoch zeigen die Ergebnisse der Konkordanzstudie, dass die gleichen Patienten mit Oncotype DX[®] und EndoPredict unterschiedlich eingeordnet werden [21]. Oncotype DX[®] ordnet mehr Patienten einem niedrigen Recurrence Score[®] zu, so dass mehr Patienten eine Chemotherapie erspart werden kann.

Unter Berücksichtigung dieser Ergebnisse ist also zu erwarten, dass Oncotype DX[®] zu größeren Kosteneinsparungen für das Gesundheitssystem führen wird als die anderen Tests, da bei mehr Patientinnen auf eine Chemotherapie verzichtet werden kann.

Tabelle 2: Erste Ergebnisse der OPTIMA-Studie

Risikogruppe	Oncotype DX® (≤ 25 vs >25)	Mamma-Print	Prosigna ROR_P	IHC4	IHC4 AQUA
N	301 (99.7%)	298 (99%)	299 (99%)	257 (85%)	271 (90%)
geringes Risiko	247 (82%)	183 (61%)	108 (36%)	62 (24%)	87 (32%)
intermediäres Risiko	-	-	88 (30%)	123 (48%)	80 (30%)
mittleres Risiko	-	-	-	-	55 (20%)
Hohes Risiko	54 (18%)	115 (39%)	103 (34%)	72 (28%)	49 (18%)

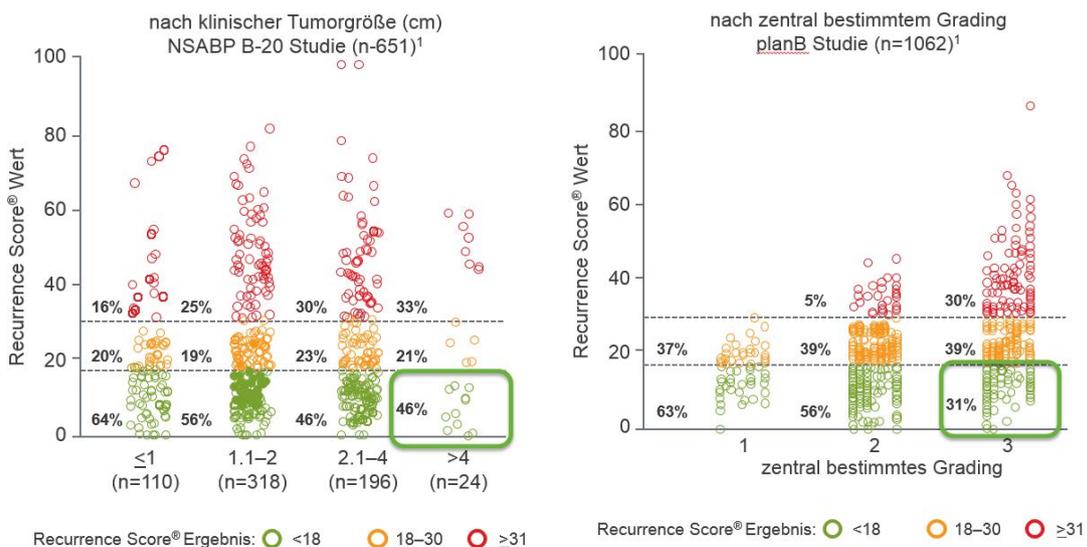
ROR_P: Risk of Recurrence; IHC: Immunhistochemie

5 Schlussfolgerung zur Evidenz von Oncotype DX®

Die Evidenzlage für Oncotype DX® setzt sich aus einer Vielzahl qualitativ hochwertiger analytischer Validitätsstudien, prospektiver-retrospektiver randomisierter kontrollierter Studien ([2, 20, 25-27]), aus über 20 Entscheidungs-Impact-Studien (einschließlich einer deutschen Studie) sowie aus zwei prospektiven Studien (TAILORx, WSG-Plan B) und zwei prospektiv evaluierten, großen Bevölkerungs-basierten Studien (CLALIT-Register, SEER-Datenbank-Analyse) zusammen. Zusammengenommen liegt für Oncotype DX® damit der höchste Grad an Evidenz vor, der von der Validierung genomischer Tests erwartet und erfüllt werden kann [12].

Die Daten der Studien zeigen überzeugend und konsistent, dass es keine klinische Notwendigkeit gibt, eine adjuvante Chemotherapie bei Patientinnen mit niedrigem oder mittlerem Recurrence Score® anzuwenden. Somit kann diese teure und Nebenwirkungs-assoziierte Behandlung bei vielen Patientinnen vermieden werden, die diese nach herkömmlichen Prognosefaktoren erhalten hätten.

Im Gegensatz hierzu gibt es für die prädiktive Aussagekraft herkömmlicher Prognoseparameter, wie Tumorgröße, Nodalstatus, Grading und Ki-67-Expression keinerlei Evidenz (s. Abbildung 3). So wurde für diese traditionellen klinischen und pathologischen Parameter weder gezeigt, dass sie eine prädiktive Vorhersage hinsichtlich des Nutzens einer Chemotherapie machen, noch wurden sie mittels Evidenzgrad I validiert.



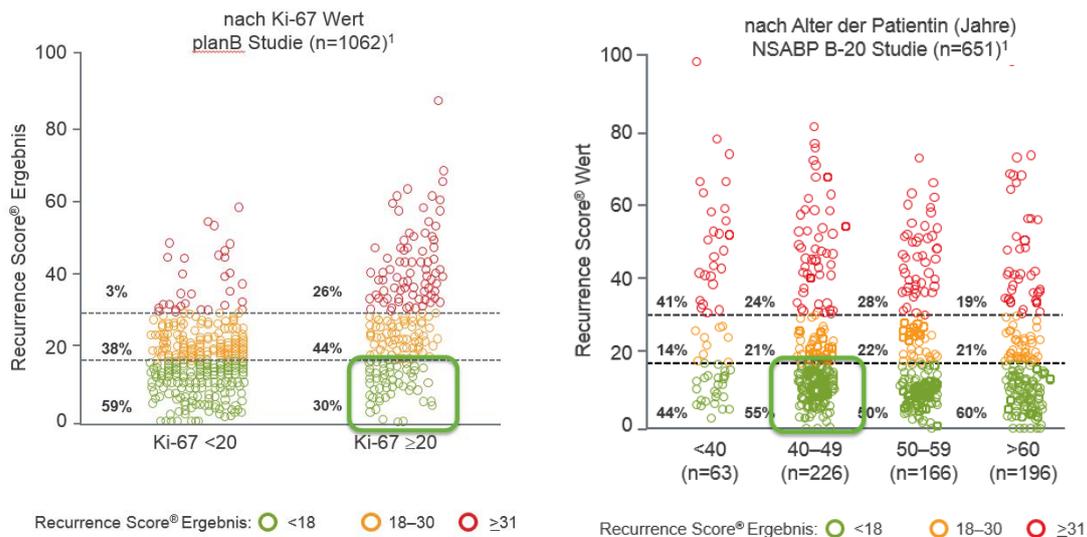


Abbildung 3: Vergleich des Recurrence Score® mit den herkömmlichen, pathologischen Parametern TumorgroÙe, Grading und Ki-67-Wert sowie dem Alter der Patientinnen (NSABP-B20 und planB). Grün umrandet markiert sind diejenigen Patientinnen, die nach herkömmlichen Kriterien eine Chemotherapie erhalten hätten, diese aber nach dem Ergebnis des Recurrence Score® nicht benötigen.

Durch diese Darstellungen wird einmal mehr deutlich, dass Recurrence Score® Informationen generiert, die über die der traditionellen Parameter weit hinausgehen. Letztere werden im Versorgungsalltag jedoch trotz eingeschränkter Evidenz herangezogen, um eine prädiktive Aussage zur Wirksamkeit einer Chemotherapie zu machen.

Ohne ausreichende Evidenz dieser einzelnen oder kombinierten Parameter, sollte im Versorgungsalltag diejenige Diagnostik herangezogen werden, die für die Patientin die beste Prädiktion mit dem besten Evidenzgrad der verfügbaren Methoden und Unterstützung durch Langzeit-Daten aus der klinischen Praxis hat.

Demnach stellt letztlich Oncotype DX® die derzeit beste und einzige Evidenz mit prädiktiver Aussagekraft bei der Entscheidung für oder gegen eine adjuvante Chemotherapie beim primären Mammakarzinom dar und muss daher bei der Bewertung des IQWiG berücksichtigt werden.

6 Literatur

1. Ballman KV (2015): Biomarker: Predictive or Prognostic? *J Clin Oncol*; 33(33):3968-71.
2. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. (2006): Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol*; 24(23):3726-34.
3. Nitz U, Gluz O, Kates RE, Hofmann D, Kreipe HH, Christgen H, et al. [P4-11-01] Prognostic impact of discordance between different risk assessment tools in early breast cancer (recurrence score, central grade, Ki67): Early outcome analysis from the prospective phase III WSG-PlanB trial SABCS 2014; San Antonio.
4. Varga Z, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, et al. (2012): How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of Breast- and Gynecopathologists. *PLoS One*; 7(5):e37379.
5. Polley MY, Leung SC, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. (2013): An international Ki67 reproducibility study. *J Natl Cancer Inst*; 105(24):1897-906.
6. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA (2010): Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol*; 11(2):174-83.
7. AGO (2015): Diagnostik und Therapie primärer und metastasierter Mammakarzinome: Prognostische und prädiktive Faktoren. [Zugriff: 30.11.2015]. URL: http://www.ago-online.de/fileadmin/downloads/leitlinien/mamma/maerz2015/de/2015D_05_Prognostische_und_praediktive_Faktoren_NEU_150216.pdf.
8. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. (2013): Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*; 24(9):2206-23.
9. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. (2007): American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol*; 25(33):5287-312.
10. National Comprehensive Cancer Network (2010): NCCN Breast Cancer Guidelines Version 1.2011/3.2010. [Zugriff: 15.11.2010]. URL: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp.
11. Senkus E, Kyriakides S, Penault-Llorca F, Poortmans P, Thompson A, Zackrisson S, et al. (2013): Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*; 24 Suppl 6:vi7-23.
12. Simon RM, Paik S, Hayes DF (2009): Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst*; 101(21):1446-52.
13. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. (2015): Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*; 373(21):2005-14.
14. Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JA, et al. (2007): Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol*; 25(15):2127-32.

15. Early Breast Cancer Trialists' Collaborative Group (2005): Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*; 365(9472):1687-717.
16. IQWiG. Pertuzumab (neues Anwendungsgebiet) - Nutzenbewertung gemäß § 35a SGB V (IQWiG-Berichte - Nr. 343). 2015.
17. Gluz O, Nitz U, Kreipe HH, Christgen H, Kates RE, Hofmann D, et al. Clinical impact of risk classification by central/local grade or luminal-like subtype vs. Oncotype DX®: First prospective survival results from the WSG phase III planB trial (Abstract number 1937). European Cancer Congress (ECC) 2015; Wien2015.
18. Shak S, Petkov VI, Miller DP, Howlader N, Gliner N, Howe W, et al. [P5-15-01] Breast cancer specific survival in 38,568 patients with node negative hormone receptor positive invasive breast cancer and oncotype DX recurrence score results in the SEER database. SABCs 2015; San Antonio.
19. Stemmer SM, Steiner M, Rizel S, Soussan-Gutman L, Geffen DB, Nisenbaum B, et al. [P5-08-02] Real-life analysis evaluating 1594 N0/Nmic breast cancer patients for whom treatment decisions incorporated the 21-gene recurrence score result: 5-year KM estimate for breast cancer specific survival with recurrence score results ≤ 30 is $>98\%$. SABCs 2015; San Antonio.
20. Dowsett M, Cuzick J, Wale C, Forbes J, Mallon EA, Salter J, et al. (2010): Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol*; 28(11):1829-34.
21. Varga Z, Sinn P, Fritzsche F, von Hochstetter A, Noske A, Schraml P, et al. (2013): Comparison of EndoPredict and Oncotype DX test results in hormone receptor positive invasive breast cancer. *PLoS One*; 8(3):e58483.
22. Stein RC, Makris A, Hughes-Davies L, Bartlett JMS, Marshall A, Hall PS, et al. Results of the OPTIMA (Optimal Personalized Treatment of early breast cancer using Multi-parameter Analysis) prelim study. The European Cancer Congress 2015 (ECC); Wien: *European Journal of Cancer*; 2015. p. 296.
23. Dowsett M, Sestak I, Lopez-Knowles E, Sidhu K, Dunbier AK, Cowens JW, et al. (2013): Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol*; 31(22):2783-90.
24. Bartlett JM, Stein RC, Bayani J, Marshall A, Dunn JA, Campanati AF, et al. [P4-11-07] Comparison of multiparameter tests in the UK OPTIMA-Prelim trial. SABCs 2014; San Antonio.
25. Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, et al. (2010): Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol*; 11(1):55-65.
26. Mamounas EP, Tang G, Paik S, Baehner FL, Liu JE, Jeong J-H, et al. Prognostic impact of the 21-gene recurrence score (RS) on disease-free and overall survival of node-positive, ER-positive breast cancer patients (pts) treated with adjuvant chemotherapy: Results from NSABP B-28. ASCO Breast Cancer Symposium 2012: *J Clin Oncol*.

27. Penault-Llorca FM, Filleron T, Asselain B, Baehner FL, Fumoleau P, Lacroix-Triki M, et al. Prediction of recurrence with the Oncotype DX recurrence score in node-positive, HR-positive, breast cancer patients treated with adjuvant chemotherapy: Results from PACS01 trial. ASCO Annual Meeting 2014: J Clin Oncol. p. abstr. 11052.

A.1.5 – Mamma Mia! Das Brustkrebsmagazin

Autoren:

- Schumacher-Wulf, Eva



Mamma Mia! · Altkönigstraße 31 · 61476 Kronberg

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IQWiG Stellungnahme zum Vorbericht D14-01
Prof. Dr. med. Jürgen Windeler
Im Mediapark 8
50670 Köln

Mamma Mia!

Eva Schumacher-Wulf
Chefredaktion

Tel: +49 (0) [REDACTED]

Mobil: +49 (0) [REDACTED]

Fax: +49 (0) [REDACTED]

IQWiG Stellungnahme zum Vorbericht D14-01

5. Dezember 2015

Sehr geehrte Damen und Herren,

mit Entsetzen habe ich festgestellt, dass im Vorbericht des IQWiGs zum Thema „Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom“ kein erwiesener Nutzen für den Einsatz von Genexpressionstests gesehen wird. In der begleitenden Pressemitteilung geht das IQWiG noch weiter und suggeriert, dass die Personen, die den Einsatz der Biomarkertests fordern, auf die PR-Kampagnen der Hersteller hereingefallen seien. Und dass die betroffenen Frauen zu Unrecht glauben, mit einem Genexpressionstest besser behandelt zu werden.

Ich persönlich setze mich seit mehr als zehn Jahren professionell mit dem Thema Mammakarzinom auseinander. In dieser Zeit habe ich es gelernt, wissenschaftliche Studien zu lesen und einzuschätzen. Die Studienlage zu Genexpressionstests ist in meinen Augen überzeugend. In meiner Meinung diesbezüglich werde ich nicht zuletzt durch die Fachgesellschaften, die sich noch viel länger als ich schwerpunktmäßig mit diesem Thema befassen, bestärkt. Denkt das IQWiG wirklich, dass nationale und internationale Fachgremien auf die PR der Hersteller hereingefallen sind und sich nicht mit den Studien auseinandergesetzt haben?

Ich habe das Gefühl, dass die Personen, die den Bericht geschrieben haben, immer noch nicht verstanden haben, warum Genexpressionstests ein großer Gewinn für die behandelnden Ärzte und die Patientinnen sind und in welchen Fällen sie eingesetzt werden sollten: Bei Patientinnen, bei denen die Leitlinien beide Handlungsoptionen (mit oder ohne Chemotherapie) offen lassen. Eben weil es für diese Patientinnen bisher überhaupt keine validen und eindeutigen Entscheidungskriterien gab.

Nationale und internationale Experten berücksichtigen bei der Evaluierung von neuen Diagnostika auch - von Ihnen als unzureichend abgetan - prospektiv-retrospektive Daten. Dies geschieht im allgemeinen Konsens unter der Voraussetzung, dass mehrere Studien zum gleichen Ergebnis kommen, um eine Verfälschung der Ergebnisse durch die Studienauswahl auszuschließen. Dies ist auch im Falle der Genexpressionstest geschehen.

Rechnungsanschrift:
GeKo Verlag GbR
Altkönigstraße 31
61476 Kronberg

www.mammamia-online.de
USt-IdNr.: [REDACTED]

[REDACTED] Sparkasse, BLZ: [REDACTED]
Konto: [REDACTED]
SWIFT: [REDACTED]
IBAN: [REDACTED]

Sie hingegen fordern– wie bei Arzneimitteln nötig, denn diese lassen sich in keinem Fall retrospektiv validieren – prospektiv-randomisierte Studien.

Wie lebensfern diese Forderung ist, zeigte sich bereits in der Diskussion im Vorfeld des Vorberichts. Stefan Lange, der stellvertretender Leiter des IQWiG, sagte hier: „Immer dann, wenn wir eine Frage haben, können wir auch randomisieren. Das ist immer ganz wichtig. Selbst bei Hochrisikopatientinnen mit Mammakarzinom scheint es so zu sein, dass noch die eine oder andere Frage offen ist, wo man dann theoretisch eine randomisierte Studie machen kann.“

Bedeutet das, dass man einem Teil der Hochrisikopatientinnen eine Chemotherapie vorenthalten würde, um neue Daten zu generieren? Dies wäre eindeutig eine Gefährdung der Patientinnen. Um diesen Frauen guten Gewissens von der Therapie abzuraten, ist die Datenlage für die Genexpressionstests einfach viel zu gut. Hier randomisierte Studien zu fordern, ist in meinen Augen schlichtweg unethisch.

Zudem reicht laut Ihrem Bericht bei den prospektiv-randomisierte Studien eine Nachbeobachtungszeit von einem Jahr. Auch das ist für mich nicht nachvollziehbar. Bei den Patientinnen, um die es hier geht, entwickeln sich kaum Metastasen innerhalb des ersten Jahres. Etwa die Hälfte der Metastasierungen findet sogar erst nach dem fünften Jahr nach Diagnose statt. Was bringt den Patientinnen da eine Nachbeobachtungszeit von einem Jahr? Die Frauen wollen auf lange Sicht gesund und am Leben bleiben!

Genexpressionstests bieten zusätzliche Informationen in einer unsicheren Situation. Es gibt zurzeit keine besser untersuchte Möglichkeit, diese Frauen bei einer sinnvollen Entscheidungsfindung zu unterstützen. Es ist nicht zu verantworten, den betroffenen Frauen diese Möglichkeit zu verweigern und sie dazu zu zwingen, die Entscheidung aufgrund ihres eigenen Bauchgefühls oder das der Ärzte zu treffen.

Ich bitte Sie daher, die Genexpressionstests neu zu bewerten und Ärzten und Patientinnen die Möglichkeit zu geben, die bestmögliche Therapieentscheidung zu treffen.

Mit freundlichen Grüßen



Eva Schumacher-Wulf
Chefredaktion

**A.1.6 – Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen (MDS)
e. V.**

Autoren:

- Janatzek, Sandra
- Preukschat, Daniela

MDS | Postfach 10 02 15 | 45002 Essen

Institut für Qualität und Wirtschaftlichkeit
im Gesundheitswesen
– Stellungnahme zum Vorbericht D14-01 –
Prof. Dr. med. Jürgen Windeler
Im Mediapark 8
50670 Köln

IBAN: [REDACTED]
BIC: [REDACTED]

Servicezeiten:
Montag-Donnerstag 8.30-16.00 Uhr
Freitag 8.30-14.00 Uhr

Datum
8. Dezember 2015

Ihre Ansprechpartnerin
Dr. Daniela Preukschat
Telefon 0201 8327 [REDACTED]
Telefax 0201 832 [REDACTED]
[REDACTED]@mds-ev.de

**Stellungnahme zum Vorbericht D14-01 (Version 1.0 vom 02.11.2015)
„Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische
Chemotherapie beim primären Mammakarzinom“**

Sehr geehrter Herr Professor Windeler,

gerne nutzen wir die Möglichkeit, zum Vorbericht „Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom“ (Auftrag D14-01) Stellung zu nehmen.

Anmerkungen zur projektspezifischen Methodik

1. Nicht berücksichtigte Studienergebnisse und ausgeschlossene Studien

1.1. Nichtberücksichtigungsanteil

Im Vorbericht wurden die Ergebnisse des RCTs von Paik (2006) und der Prognosestudien Gnant (2014) und Sgroi (2013) aufgrund eines Anteils der Patientinnen ohne jegliche Berücksichtigung in der Auswertung (Nichtberücksichtigungsanteil) von größer als 30 % nicht in die Nutzenbewertung einbezogen. Dieses Vorgehen, Ergebnisse solcher (insbesondere prospektiv-retrospektiver) Analysen **alleine aufgrund** eines Nichtberücksichtigungsanteils von größer als 30 % nicht in die Nutzenbewertung einfließen zu lassen, ist aus unserer Sicht nicht angemessen. Wir empfehlen mit Verweis auf Simon (2009) (Referenz Nr. 36 im Vorbericht) die Überprüfung, ob die Auswahl der retrospektiv ausgewählten Proben in der jeweiligen Studie in einer Art und Weise erfolgte, die einen Selektionsbias ausschließen kann („we suggest that the correlative study should include **at least two-thirds** of the total accrued patients or

*that the patients be selected in a way that strives to **avoid selection bias***“ (Simon, 2009)).

Die im Vorbericht beschriebenen Berechnungen der Nichtberücksichtigungsanteile lassen erkennen, dass in die Festlegungen, welche Anzahlen in der jeweiligen Studie als Zähler bzw. Nenner herangezogen werden, Überlegungen hinsichtlich eines potentiell auszuschließenden Selektionsbias eingeflossen sind. Dieses Vorgehen ist grundsätzlich zu begrüßen und sollte durch das Aufheben der festen 30%-Grenze ergänzt werden.

1.2. Population

Ziel der vorliegenden vorläufigen Nutzenbewertung ist die Population “Patientinnen mit primärem Hormonrezeptor-positivem, HER2/neu-negativem Mammakarzinom und 0 bis 3 befallenen Lymphknoten“. Wir empfehlen detaillierter auszuführen, warum in mehreren Fällen ein Studienausschluss von potentiell relevanten RCTs (z.B. **Tang 2011** (Nr. 15 in Abschnitt A.6.1.3 – E1), **Albain 2011** (Nr. 1 in Abschnitt A.6.1.3 – E1)) und Prognosestudien (z.B. **Paik 2004** (Nr. 80 in Abschnitt A.6.2.3 – Nicht E1), **Paik 2006** (Nr. 81 in Abschnitt A.6.2.3 – Nicht E1), **Tang 2011** (Nr. 107 in Abschnitt A.6.2.3 – Nicht E1)) mit der Angabe des Ausschlussgrundes „E1: Population“ erfolgte. Ferner empfehlen wir darzustellen, warum anhand dieser Studien außerdem auch keine vergleichende Nutzenbewertung (bzw. eine Bewertung des Mehrwertes durch die prognostische Information des Biomarkers) **für entsprechende Subgruppen** möglich ist.

2. Vergleichende Bewertung von RCTs und Prognosestudien / Notwendigkeit einer Vergleichsintervention

2.1. Berücksichtige Studientypen: RCTs

Wie bereits in unserer Stellungnahme zum vorläufigen Berichtsplan angemerkt und in der wissenschaftlichen Erörterung am 15.12.2014 ausdrücklich betont (siehe auch Seite 53 der Dokumentation und Würdigung der Anhörung; Tagesordnungspunkt 4 (Verschiedenes)), ist nach unserer Einschätzung – für mindestens eine biomarkerbasierte Strategie („Referenz-Biomarker“) – ein Vergleich der biomarkerbasierten Therapieempfehlung mit der bisherigen Therapieempfehlung notwendig. Auch wenn die bisherige Therapieempfehlung bei Teilen des hier betrachteten Patientenkollektivs unter Unsicherheit und vermutlich teilweise uneinheitlich erfolgt, ist es dennoch erforderlich sicherzustellen, dass die biomarkerbasierten Therapieempfehlungen einen Zusatznutzen (vergleichbare Mortalität bei gleichzeitig reduzierten Nebenwirkungen durch Verzicht auf adjuvante Chemotherapien) gegenüber den bisherigen Therapieempfehlungen

aufweisen. Andernfalls bliebe eine Unsicherheit darüber bestehen, ob der biomarkerbasierte Verzicht auf Chemotherapien sich möglicherweise – im Vergleich zu den bisherigen Therapieempfehlungen – ungünstig auf das Überleben auswirken könnte.

Auch in Abschnitt 2 „Fragestellung“ des Vorberichtes ist als Ziel der Untersuchung „die Nutzenbewertung einer biomarkerbasierten Strategie zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie **im Vergleich zu** einer biomarkerunabhängigen Entscheidungsstrategie oder einer zweiten biomarkerbasierten Entscheidungsstrategie“ formuliert.

Die in den Vorbericht eingeschlossene randomisierte Studie von Paik 2006 (Interaktionsdesign mit prospektiv-retrospektiver Analyse) sieht jedoch **keinen Vergleich** zu einer „biomarkerunabhängigen Entscheidungsstrategie oder einer zweiten biomarkerbasierten Entscheidungsstrategie“ vor.

Die zweite in den Vorbericht eingeschlossene randomisierte Studie Martin 2014 könnte zwar grundsätzlich als **vergleichendes** Interaktionsdesign aufgefasst und ausgewertet werden („EP-score“ vs. „EPclin-score“). Jedoch erfolgt zum einen ein solcher Vergleich nicht und zum anderen würde sich auch dieser Vergleich von dem aus unserer Sicht notwendigen Vergleich zu den bisherigen Therapieempfehlungen unterscheiden.

2.2. Berücksichtigte Studientypen: Prognosestudien

Im Vorbericht, Abschnitt „Einordnen des Arbeitsergebnisses“, S.13, werden die kürzlich publizierten Ergebnisse (Sparano 2015, Referenz Nr. 31 im Vorbericht) der TAILORx-Studie zur nicht randomisierten Subgruppe der Patientinnen mit einem Recurrence Score von 0 bis 10 perspektivisch positiv gewertet. Nach unserer Einschätzung ist – **auch für die Beurteilung der prognostischen Güte der Biomarker – ein Vergleich zu den bisher eingesetzten Prognosefaktoren/Biomarkern** erforderlich. Aus unserer Sicht ist dieser vergleichende Bewertungsansatz in den Daten von Sparano 2015 nicht erkennbar. Ein solcher Vergleich sollte zeigen, dass ein neuer Biomarker beziehungsweise ein neues Testergebnis eine Verbesserung der Güte der erlangten prognostischen Information gegenüber der Güte der prognostischen Information auf Basis der bisher verwendeten (Kombination von) Prognosefaktoren bewirkt.

Wir empfehlen somit **auch für die Bewertung der prognostischen Güte** (und darauf aufbauende Schlussfolgerungen im Hinblick auf den Nutzen) einen **vergleichenden Bewertungsansatz**. Dies wird aus unserer Sicht auch in Ihrem Methodenpapier Allgemeine Methoden, Version 4.2 vom 22.04.2015, Seite 65, be-

stätigt: „Wenn eine Studie aussagekräftige Daten zum Nutzen, zur diagnostischen Güte oder zur prognostischen Wertigkeit liefern soll, ist es wesentlich, dass hierbei ein Vergleich zur bisherigen diagnostischen Vorgehensweise erfolgt. Nur auf diese Weise lässt sich der **Mehrwert der diagnostischen oder prognostischen Information** sicher bestimmen.“

3. Einordnung der Ergebnisse gegenüber anderen systematischen Übersichten

Im Abschnitt A.4.1 „Einordnung der Ergebnisse gegenüber anderen systematischen Übersichten“, S.70, empfehlen wir die Darstellung des HTA-Berichtes der BlueCross BlueShield Association („*Gene expression profiling in women with lymph node-negative breast cancer to select adjuvant chemotherapy*“), welcher in Abschnitt A.6.1.2. aufgelistet ist.

4. Berücksichtigung der Ergebnisse relevanter laufender Studien

Im Vorbericht wird angemerkt, dass die Ergebnisse von 3 der relevanten laufenden Studien (MINDACT, TAILORx, planB), für die bereits Anfang 2016 Ergebnisse erwartet werden, **gegebenenfalls im Rahmen des Abschlussberichts berücksichtigt werden können**. Wir empfehlen, die Berücksichtigung neuer Ergebnisse nur in Verbindung mit der Erstellung einer weiteren (neuen) Version des Vorberichtes und mit der Durchführung eines erneuten Stellungnahmeverfahrens umzusetzen.

5. Einschluss der Studie Martin (2014)

Die eingeschlossene randomisierte prospektiv-retrospektive Studie von Martin (2014) untersucht die Interaktion zwischen dem Ergebnis des Biomarkers (EndoPredict) und der Art des Chemotherapie-Regimes (FEC mit oder ohne Paclitaxel). Hingegen wird die Interaktion zwischen dem Biomarker-Ergebnis und der adjuvanten Therapie mit oder ohne Chemotherapie in dieser Studie nicht untersucht.

Aus unserer Sicht stellt sich zunächst die Frage, ob es sinnvoll ist, unter der Zielsetzung des Vorberichtes „Nutzenbewertung einer biomarkerbasierten Strategie für oder gegen eine adjuvante systemische Chemotherapie im Vergleich zu [...]“ (S. 4 Vorbericht) auch die Nutzenbewertung einer biomarkerbasierten Strategie **für oder gegen die zusätzliche Gabe von Paclitaxel zum Chemotherapie-Regime FEC** zu inkludieren.

Sofern dieses Vorgehen als sinnvoll bewertet wird, sollte nach unserer Einschätzung bei der Darstellung der Ergebnisse der Studie von Martin (2014) sorgfältig transparent gemacht werden, dass es sich um die Interaktion zwischen dem Biomarker-Ergebnis und der Art der Chemotherapie (Regime mit vs ohne Paclitaxel) handelt. Eine Verwechslung

mit der Frage, ob der Biomarker geeignet ist, die Patientinnen zu identifizieren, die auf eine adjuvante Chemotherapie zugunsten einer rein endokrinen Therapie verzichten können, und die Patientinnen zu identifizieren, die von einer adjuvanten Chemotherapie zusätzlich zur endokrinen Therapie profitieren, ist zu vermeiden.

In Abschnitt A.4 „Kommentar“ des Vorberichts (S. 69) ist zutreffend dargelegt, dass „die Frage, welche Patientinnen komplett auf eine Chemotherapie verzichten und rein endokrin behandelt werden können, [...] im Rahmen dieser Studie nicht beantwortet werden“ kann. Eine solche Klarstellung wäre auch in den Ergebnisabschnitten hilfreich. Dort findet sich bislang die aus unserer Sicht missverständliche Formulierung „Interaktion des EndoPredict-Ergebnisses mit der Variable Behandlung mit (einer) Chemotherapie“ (S. 11, 66, 67). Auch in den Abschnitten 5 „Einordnung des Arbeitsergebnisses“ und 6 „Fazit“ empfehlen wir, die Datenlage hinsichtlich der Therapieentscheidung für oder gegen den Verzicht auf eine adjuvante Chemotherapie und die Datenlage hinsichtlich der Therapieentscheidung für eines von zwei Chemotherapie-Regimen getrennt darzustellen.

Wir bedanken uns für die Möglichkeit, unsere Überlegungen einbringen zu können und verbleiben

mit freundlichen Grüßen

Dr. Daniela Preukschat
Team Methodik/Methodenberatung
Bereich Evidenzbasierte Medizin

Dr. Sandra Janatzek
Leiterin Team Methodik/Methodenberatung
Bereich Evidenzbasierte Medizin

A.1.7 – NanoString Technologies GmbH

Autoren:

- Faull, Iris
- Mellert, Uta
- Stapf, Martin

Stellungnahme der NanoString Technologies GmbH zum Vorbericht (D14-01) – Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mamma-Karzinom

Dr. Uta Mellert, Iris Faull, Dr. Martin Stapf¹

1 Bewertung und Interpretation der im Vorbericht eingeschlossenen Studien

Abschnitt A.3 – Details der Ergebnisse

- ≠ Die Ergebnisse der ABCSG-8- und TransATAC-Studie sowie der kombinierten Analyse von Gnant et al 2015 für Prosigna sollten in die Tabellen in diesem Abschnitt (**Tabelle 8, 11, 12, 16**) aufgenommen und nicht von dieser Analyse ausgeschlossen werden. Gnant et al lieferten einen klaren Vergleich von Basischarakteristika zwischen allen randomisierten und in die PAM50-Analyse eingeschlossenen Patientinnen. Es wurden keine signifikanten Unterschiede zwischen beiden Gruppen beobachtet. Daher sollten die Ergebnisse der prognostischen Marker für die gesamte Studienpopulation repräsentativ sein. Die Vergleichbarkeit zwischen allen randomisierten und analysierten Patientengruppen erscheint wichtiger für die Ableitung relevanter Informationen, wenn ein wesentlicher Teil der Patientinnen in die Analyse eingeschlossen wird. PAM50 liefert außerdem noch wesentliche weitere prognostische Informationen, die über die klassischen prognostischen klinisch-pathologischen Faktoren hinausgehen und Grundlage für die Entscheidung für oder gegen eine Chemotherapie sind.
- ≠ Die klinische Validierungsstudie von Martin et al 2014 für Endopredict wurde in diesen Bericht eingeschlossen, doch eine Analyse mit PAM50/Prosigna für dieselbe Studie wurde nicht erwähnt bzw. nicht eingeschlossen.

2 Benennung von zusätzlichen, im Vorbericht nicht genannten, relevante Studien

- ≠ **In Abschnitt A.2.1.7.2** werden Einschlusskriterien für Meta-Analysen bewertet, allerdings werden Kriterien für die Bewertung der kombinierten Analyse nicht eingeschlossen. Für Prosigna gibt es veröffentlichte Daten zur kombinierten Analyse (Gnant et al 2015) aus zwei RCTs mit follow-up Daten über 15 Jahre, die in dieser Bewertung berücksichtigt werden sollten.
 - a. **Literatur:** Gnant M, Sestak I, Filipits M, et al. (2015) Identifying Clinically Relevant Prognostic Subgroups of Postmenopausal Women with Node-positive Hormone Receptor Positive Early Stage Breast Cancer Treated with Endocrine Therapy: A combined analysis of ABCSG-8 and ATAC using the PAM50 risk of

¹ Formblätter des IQWiG zur schriftlichen Stellungnahme zum Vorbericht anbei

recurrence score and Intrinsic Subtype. 2015 Ann Oncol.
doi:10.1093/annonc/mdv215

Die TransATAC-Studie sollte ebenso wie spätere Analysen in den Bericht eingeschlossen werden. Als Begründung für die Einbeziehung dieser Studie und späterer Analysen verweisen wir auf Punkt 4 „Anmerkungen zur projektspezifischen Methodik“.

Literatur:

- b. Dowsett M, Sestak I, Lopez-Knowles E, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol* 2013; 31: 2783-2790.
- c. Sestak I, Cuzick J, Dowsett M et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol* 2015; 33: 916-922.

Die folgenden Studien bestätigen die Validierungsdaten:

DBCG-Kohortenanalyse: Diese Analyse umfasste 1466 Proben aus 1480 Tumorpräparaten (90%) mit einer medianen Dauer der Nachbeobachtung von 9,25 Jahren. Dabei verbesserte Prosigna die Beurteilung der Prognose bei nodal positiven (N1) Patientinnen und identifizierte mindestens 37% Patientinnen mit einem und 15% mit zwei positiven Lymphknoten, die eine ausgezeichnete Prognose besitzen und denen somit unter realen Bedingungen eine adjuvante Chemotherapie erspart werden kann. Diese Studie bestätigte insgesamt die Ergebnisse der klinischen Studien Trans ATAC und ABCSG8.

Literatur:

1. Ejlertsen B, Jensen M, Eriksen J et al. Validation of prediction of distant recurrence (DR) by Prosigna (PAM50) in subgroups of a Danish Breast Cancer Cooperative Group (DBCG) cohort of node-positive (N1), hormone receptor positive (HR+), postmenopausal early breast cancer (EBC) patients allocated 5yr of endocrine therapy (ET). *J Clin Oncol* 33, 2015 (suppl; abstr 513) 2015.

In einer anderen Studie mit 99% bewerteten Tumorproben mit einer medianen Dauer der Nachbeobachtung von 9,25 Jahren erbrachte Prosigna eine korrekte Differenzierung von Risikogruppen. Einem wesentlichen Anteil der Patientinnen könnte demnach unter realen Bedingungen eine Chemotherapie erspart werden.

Literatur:

2. Lænkholm A, Jensen M, Eriksen J et al. Prediction of 10yr distant recurrence (DR) using the Prosigna (PAM50) assay in a Danish Breast Cancer Cooperative Group (DBCG) cohort of postmenopausal Danish women with hormone receptor-

positive (HR+) early breast cancer (EBC) allocated to 5yr of endocrine therapy (ET) alone. J Clin Oncol 33, 2015 (suppl; abstr 546) 2015.

In einer dritten Studie wurde bestätigt, dass Prosigna das Risiko für späte Fernrezidive (DR) für Patientinnen mit hormonrezeptor-positivem (HR+) frühem Mammakarzinom (EBC) unabhängig vom Lymphknotenstatus in einer Kohorte unter realen Bedingungen ohne Selektionsfehler durch Ärzte bestimmen konnte.

Literatur:

3. Lænkholm A, Jensen M, Eriksen J et al. Prediction of late distant recurrence (DR) using the Prosigna (PAM50) test in a Danish Breast Cancer Cooperative Group (DBCG) cohort of postmenopausal women diagnosed with hormone receptor-positive (HR+) early breast cancer (EBC) allocated to 5yr of endocrine therapy (ET). J Clin Oncol 33, 2015 (suppl; abstr 544) 2015.

Studie C9741

Im Hinblick auf RCTs, Prädiktion und das Ansprechen auf Chemotherapie sollten die folgenden Studien in der Bewertung berücksichtigt werden, da hier der Nachweis erbracht wird, dass Prosigna zusätzliche prognostische Informationen liefert, die über klassischen klinisch-pathologische Prognoseparameter hinausgehen.

Die Studie CALGB 9741 untersuchte den Zusammenhang zwischen BMI und Prognose bei nodal positivem Brustkrebs. Bei 1272 von 1909 in diese prospektiv geplante randomisierte Studie aufgenommenen Patientinnen wurde der intrinsische Subtyp mittels PAM50 bestimmt. Rezidivfreies Überleben (RFS) und Gesamtüberleben (OS) wurden bewertet. Alle Patientinnen erhielten Anthrazyklin, Cyklophosphamat und Taxan.

Literatur:

Ligibel JA, Cirrincione CT, Liu M et al. Body mass index, PAM50 subtype and outcomes in node-positive breast cancer: CALGB 9741(Alliance). JNCI 2015, 107(9):djv179

Klinische Studie NCIC- MA.5: Phase 3-Studie, die zwei Chemotherapieregime bei prämenopausalen Frauen mit nodal positivem Brustkrebs untersuchte; durch die Analyse mit PAM50/Prosigna wurden prädiktive Informationen nachgewiesen, wonach Patientinnen mit Tumoren vom Subtyp Luminal A und Basal-like keinen Nutzen von Anthrazyklin-haltigen Chemotherapien besitzen und potenziell auf eine Therapie mit Anthrazyklinen verzichten können.

Literatur:

Cheang MC, Voduc KD, Tu D et al. Responsiveness of intrinsic subtypes to adjuvant anthracycline substitution in the NCIC.CTG MA.5 randomized trial. Clin Cancer Res 2012; 18: 2402-2412

Klinische Studie NCIC-MA.21: Phase 3-Studie, in der PAM50/Prosigna zur Untersuchung der Prognose und zum Prädiktionsvergleich zwischen zwei Chemotherapiearmen eingesetzt wurde; dabei waren höhere PAM50/Prosigna Rezidivrisiko (ROR)-Scores mit schlechterem rezidivfreien Überleben (RFS) assoziiert ($P=0,03$) und wurde es wurde festgestellt, dass der intrinsische Subtyp (Luminal A, Luminal B, etc.) einen signifikanten prognostischen Effekt auf RFS hat ($P=0,002$).

Literatur

Liu S, Chapman JA, Burnell MJ et al. Prognostic and predictive investigation of PAM50 intrinsic subtypes in the NCIC CTG MA.21 phase III chemotherapy trial. *Breast Cancer Res Treat* 2015; 149: 439-448).

3. Verweise auf andere qualitativ angemessene Unterlagen, einschließlich einer Begründung für ihre jeweilige fragestellungsbezogene Eignung und Validität.

- A. Gemäß der jüngsten St. Gallen-Konsensuskonferenz von 2015 wird eine Einteilung von HR+/Her2--Brustkrebs in intrinsische Subtypen zur Entscheidung über die weitere adjuvante Therapie dringend empfohlen. Im IQWiG-Bericht wird ebenfalls eine Einteilung in intrinsische Subtypen auf der Grundlage einer immunhistochemischen Analyse von ER, PR HER2 und Ki-67 befürwortet. Diese immunhistochemische Analyse ist der Versuch einer Übertragung in molekulare Signaturen, die jedoch wesentliche Beschränkungen aufweist. Die St. Gallen-Konferenz „stellte fest, dass IHC-Tests die wirklichen intrinsischen Subtypen nicht korrekt messen“ (Coates *Ann Oncol* 2015). Eine korrekte Klassifizierung mit der wichtigen Unterscheidung zwischen den Subtypen Luminal A und Luminal B sind ausschließlich durch PAM-50/Prosigna gegeben.

Literatur:

Coates AS, Winer EP, Goldhirsch A et al. Tailoring therapies - improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer *Ann Oncol*. 2015 Aug;26(8):1533-46

- B. In der „Optima-Studie“ ist Prosigna aufgrund der besten Kosteneffizienz im Vergleich mit anderen Gensignaturen der Genexpressionstest der Wahl.
- C. Darüber hinaus werden auch prospektiv geplante retrospektive Analysen mit Prosigna für die Kohorte der Brustkrebspatientinnen mit triple-negative-Status für die Plan B- und ADAPT-Studie durchgeführt, da der medizinische Bedarf dieser Patientengruppe sehr hoch ist; hierbei sollen ebenfalls Untergruppen definiert werden, denen unnötige Therapien erspart werden können, und verbesserte individuell zugeschnittene Therapien erreicht werden.

Literatur:

Hofmann D, Nitz U, Gluz O et al: WSG ADAPT Study protocol, <http://www.trialsjournal.com/content/14/1/261>

Barlett J, Canney P, Campbell A et al. Selecting breast cancer patients for chemotherapy: The opening of the UK-OPTIMA trial. *Clinical Oncology* 25 (2013) 109-116.

Hall PS, Smith AF, Vargas-Palacios A et al. UK OPTIMA-prelim study demonstrates economic value in more clinical evaluation of multi-parameter prognostic tests in early breast cancer. *SABCS 2014*, P4-11-07

4. Anmerkungen zur projektspezifischen Methodik unter jeweiliger Angabe wissenschaftlicher Literatur zur Begründung der Anmerkung

Abschnitt 1-6 (Hintergrund bis Fazit)

1. Unklar ist, weshalb das Kriterium des Einschlusses von mindestens 70% der ursprünglichen Patientenpopulation als entscheidend für den Einschluss von Studien in diesen Bericht gilt, auch wenn es sich um Studien handelt, die teilweise über 1000 Patientinnen retrospektiv einschließen und analysieren konnten. Zulassungsbehörden wie die FDA wenden bei ihrer Beurteilung von Studien offenbar andere Kriterien an, wie aus der Zulassung von Studien für Prosigna und BCI zu ersehen ist. Dies gilt offenbar auch für die Experten, welche für die AGO-Therapieleitlinien ebenso wie für die anderen relevanten europäischen und amerikanischen Therapieleitlinien wie ESMO, St. Gallen, SEOM, NCCN 2015, zuständig sind und Genexpressionstests bei entsprechender Indikation in ihre Empfehlungen aufgenommen haben.

Literatur:

- ≠ Senkus E, Kyriakides S, Ohno S et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 2015; Sept 26: Suppl5v8-v30.
 - ≠ Garcia-Saenz JA, Bermejo B, Estevez LG et al. SEOM clinical guidelines in early-stage breast cancer 2015. *Clin Trans Oncol* 2015 Oct 26 (Epub ahead of print).
 - ≠ Senn HJ, Thurlimann B, Gnant M. St.Gallen-2015 in Vienna: News from a successful breast cancer conference transposition! *Breast* 2015 Nov 24 Suppl 2: S1.doi.10.1016/breast.2015.07.030 Epub 2015 Oct 4
 - ≠ NCCN Breast Cancer Guidelines. In 3.2015. NCCN 2015.
2. Unklar ist, weshalb große Kohortenanalysen aus dem Bericht ausgeschlossen wurden. Die FDA-Zulassung von Mammaprint erfolgte aufgrund der Vorlage einer großen Kohortenanalyse. Wir reichen außerdem die DBCG-Kohortenanalyse ein, die den Effekt von Prosigna an 3000 Frauen, einschließlich knapp 1500 nodal positiven Frauen, unter

realen Bedingungen untersuchte, um zu bestätigen, ob die in klinischen Studien beobachteten Ergebnisse unter realen Bedingungen replizierbar sind, in denen es keine Selektionsfehler gibt.

3. Weshalb wurde Martin et al 2014 in den Bericht eingeschlossen, obwohl lediglich 555/1246 (45%) Tumoren bewertet wurden? Weshalb konnte der Biomarker nur in 555/800 (69%) der untersuchten Fälle erfolgreich bestimmt werden?
4. Bei der Bewertung der Ergebnisse der patientenrelevanten Endpunkte ist der primäre Endpunkt für alle auf dem Markt befindlichen Genexpressionstests die Rezidivfreiheit nach 10 Jahren. Dieser Endpunkt wurde von medizinischen Fachkreisen so gewählt, um gerade für frühe Mammakarzinome mit geringer Rezidivwahrscheinlichkeit einen langen Nachbeobachtungszeitraum zu überblicken.. Dies gilt ebenfalls für Prosigna, denn die Ergebnisse der ABCSG8- und der TransATAC-Studie zeigen, dass die Fähigkeit des Tests zur Differenzierung zwischen geringem, intermediärem und hohem 10-Jahres-Rezidivrisiko signifikante Unterschiede aufdeckt. Die prognostischen Daten sind wichtig für die Identifikation von Patientinnen, die ein geringes Rezidivrisiko und einen geringen Nutzen von einer Chemotherapie besitzen. Diesen Patientinnen kann somit eine Chemotherapie erspart werden, gemäßdem Kriterium der deutschen AGO-Leitlinien von 2015, wonach „ein geringes absolutes Risiko einen geringen absoluten Nutzen impliziert“. Die Daten der akzeptierten Studien erreichten ein Evidenzlevel 1. Siehe: DKG e.V. 2015 Guidelines Breast Version 2015 1.D AGO
5. In der Optima-„per-lim“, Vorstudie zum Vergleich der verschiedenen prognostischen Tests wurde festgestellt, dass Prosigna den größten Gesamtwert erbrachte und somit derjenige Test ist, der in der prospektiven Optima-Studie verwendet wird. Die Optima-Studie wurde in diesem Vorbericht leider nicht berücksichtigt?
6. Im Fazit in **Abschnitt 6** heißt es: „Für einen Nutzen bzw. Schaden einer biomarkerbasierten Entscheidung für oder gegen eine adjuvante Chemotherapie beim primären Mammakarzinom liegt kein Anhaltspunkt vor, da die Daten von 3 der 4 eingeschlossenen Studien aufgrund der hohen Unvollständigkeit nicht für die vorliegende Nutzenbewertung verwendet werden können.“

Tatsächlich gibt es in Fachkreisen sowie bei den Zulassungsbehörden eine andere Interpretation, was den Nutzen einer biomarkerbasierten Entscheidung für oder gegen eine adjuvante Chemotherapie mithilfe dieser Tests betrifft. Gemäß den ESMO-, St. Gallen- und AGO-Leitlinien besteht der Hauptnutzen der Genexpressionstests in deren Fähigkeit zu bestimmen, welchen Frauen eine adjuvante Chemotherapie und die damit verbundenen Nebenwirkungen erspart werden kann. Darüber hinaus wird in onkologischen Fachkreisen sowohl BCI als auch Prosigna die Fähigkeit bestätigt, eine Patientenpopulation mit geringem Risiko zu identifizieren, bei der auf eine verlängerte adjuvante endokrine Therapie verzichtet werden kann.

7. **Abschnitt A.3.3:** Verzerrungspotenzial der Studien. Es wird darauf hingewiesen, dass für die in dieser Studie verwendeten Genexpressionstests (PAM50 und Endopredict) nicht signifikant nachgewiesen wurde, dass diese für die Bestimmung eines Nutzens einer bestimmten Chemotherapie verwendet werden können. Tatsächlich ist dies auch nicht der Hauptzweck der Genexpressionstests. Gemäß den deutschen AGO-Leitlinien sowie anderen internationalen Leitlinien werden Genexpressionstests eingesetzt, um zu entscheiden, ob bestimmten Patientinnen die Nebenwirkungen einer Chemotherapie erspart werden können.

8. In **Abschnitt A.4.2** heißt es: „Da es keinen Konsens dazu gibt, unterhalb welchen Rezidivrisikos den Patientinnen eine Chemotherapie erspart werden kann, wurde für eine Nutzenbewertung auf Basis von Prognosestudien unter Zuhilfenahme verschiedener Annahmen und Daten aus der Literatur ein möglicher Grenzwert festgelegt.“ Wir weisen darauf hin, dass gemäß den Leitlinien St. Gallen 2013 und 2015 sowie den AGO-Leitlinien bei Patientinnen mit "geringem Risiko" dieses Risiko als ausreichend niedrig gilt, um auf eine Chemotherapie zu verzichten, was, wie bereits erwähnt, den Hauptzweck der Genexpressionstests darstellt.

Literatur:

- ≠ Goldhirsch A, Winer EP, Coates AS et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* 2013; 24: 2206-2223).
- ≠ Coates AS, Winer EP, Goldhirsch A et al. Tailoring therapies - improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer. *Ann Oncol.* 2015 Aug;26(8):1533-46
- ≠ Senn HJ. St. Gallen consensus 2013: optimizing and personalizing primary curative therapy of breast cancer worldwide. *Breast Care (Basel)* 2013; 8: 101.
- ≠ Untch M, Gerber B, Harbeck N et al. 13th st. Gallen international breast cancer conference 2013: primary therapy of early breast cancer evidence, controversies, consensus - opinion of a german team of experts (zurich 2013). *Breast Care (Basel)* 2013; 8: 221-229.
- ≠ Ahn HJ, Jung SJ, Kim TH et al. Differences in Clinical Outcomes between Luminal A and B Type Breast Cancers according to the St. Gallen Consensus 2013. *J Breast Cancer* 2015; 18: 149-159.
- ≠ Jackisch C, Harbeck N, Huober J et al. 14th St. Gallen International Breast Cancer Conference 2015: Evidence, Controversies, Consensus - Primary Therapy of Early Breast Cancer: Opinions Expressed by German Experts. *Breast Care (Basel)* 2015; 10: 211-219.
- ≠ Gnant M, Thomssen C, Harbeck N. St. Gallen/Vienna 2015: A Brief Summary of the Consensus Discussion. *Breast Care (Basel)* 2015; 10: 124-130.

9. Ärztekammern erkennen an, dass ein zwingender medizinischer Bedarf zur Bereitstellung weiterer prognostischer Daten besteht, um individuelle Therapieentscheidungen für unklare Fälle zu definieren, bei denen klassische klinisch-pathologische Parameter nicht ausreichen. In den deutschen (AGO), EU (ESMO) und USA (NCCN) Leitlinien wird der Nutzen von Gensignaturen für Patientinnen anerkannt, für die keine klare Entscheidung für oder gegen eine adjuvante Therapie möglich ist und für die somit eine individuelle Therapieentscheidung verbessert wird. Bezeichnend ist hier, dass die Bedeutung von intrinsischen Subtypen für Therapieentscheidungen in Fachkreisen breite Anerkennung findet und bereits 2011 in die St. Gallen-Leitlinien aufgenommen wurde. Diese Information liefert der Prosigna zusätzlich zu einem Risikoscore, der das 10-Jahres Rezidivrisiko angibt und in dieser Form auch von anderen Tests bestimmt wird.

Literatur:

- ≠ Gnant M, Harbeck N, Thomssen C. St. Gallen 2011: Summary of the Consensus Discussion. Breast Care (Basel) 2011; 6: 136-141.
 - ≠ Goldhirsch A, Wood WC, Coates AS et al. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol 2011; 22: 1736-1747.
10. In **Abschnitt A.2.1** heißt es unter dem Index-Parameter, dass der Test folgende Kriterien erfüllen muss: „Sowohl Testentwicklung als auch -validierung müssen an einer unabhängigen Stichprobe durchgeführt worden sein. Anhand des berichteten Testergebnisses lässt sich die Zielpopulation dichotom einer Therapieempfehlung für oder gegen adjuvante Chemotherapie zuordnen.“
- Anders als bei der Entwicklung einiger anderer Tests, wie z.B. Oncotype Dx, ist zu beachten, dass Prosigna in Bezug auf seine Entwicklung diesen Parameter erfüllt, da die Testentwicklung und -validierung anhand von unabhängigen Kohorten und unabhängigen Reviewern durchgeführt wurde.
- In diesem Abschnitt ist unklar, weshalb die Konkordanzstudien für Prosigna nicht in diese Bewertung eingeschlossen wurden, obwohl sie sowohl bei Therapieleitlinien als auch Zulassungsbehörden wie der FDA der USA, der FDA Kanada und der FDA Australien eingereicht und akzeptiert wurden.
11. In **Abschnitt A.2.1.7.1** – Gegenüberstellung der Ergebnisse der Einzelstudien heißt es: „Ergebnisse fließen in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen basieren.“

Wir weisen darauf hin, dass dieses Kriterium bei der Bewertung zur Nutzenbewertung durch Zulassungsbehörden in Deutschland, Europa, den USA und anderen Fachkreisen, die Behandlungsleitlinien festlegen, nicht zugrunde gelegt wurde. Im Übrigen müßte aufgrund dieses Kriteriums die Endopredict-Studie von Martin et al 2014 ebenfalls nicht in diese Bewertung eingeschlossen werden.

12. Private Krankenversicherungen in Europa und den USA erstatten die Kosten für Prosigna, doch es gibt auch staatliche Krankenversicherungen, die die Kostenerstattung für Prosigna leisten, beispielsweise in Spanien, der Schweiz, Israel und den USA (Medicare).

13. Abschnitt A.10

Es ist nicht nachzuvollziehen, weshalb der Grenzwert für einen Nutzen der Chemotherapie bezogen auf das 10-Jahres-Rezidivrisiko aufgrund von nicht veröffentlichten Daten auf 5% festgelegt wurde.

Fernrezidive sind mit einem Mortalitätsrisiko von 65% assoziiert; dies gilt für jede Art von Rezidiv (lokal/lokoregional, kontralateral und fern). Sämtliche verfügbaren Daten weisen darauf hin, dass das langfristige Mortalitätsrisiko bei Fernrezidiven bei mindestens 90% liegt. Die Toxizität einer bestimmten Chemotherapie korreliert nicht mit dem Rezidivrisiko.

Die Zahlenangaben bezüglich der SUEs werden als zu niedrig angenommen. Es wurden keine UEs in Verbindung mit dem Test berichtet. Vielmehr weisen alle aktuellen Daten auf ein SUE-Risiko moderner Chemotherapieregimes von 20-30% (z.B. Lück et al. Senology 2015). Hinzu kommen die Langzeitrisiken, wie Myelodysplasien, periphere Neuropathien, Sekundärmalignome, etc.

Bei Einsatz hoch dosierter Chemotherapie in diesem Setting (als nicht-Standardtherapie) liegt das Hospitalisierungsrisiko bei bis zu 100%.

Literatur

Wolff AC, Blackford AI, Visvanathan K et al. Risk of marrow neoplasms after adjuvant breast cancer therapy: The national Comprehensive cancer network experience. JCO 2014 Dec 22(ahead of print)

Enright K, Grenfell E, Yung L. et al. Population-based assessment of emergency room visits and hospitalizations among women with early breast cancer. J Oncol Practice Vol 11(2).

Hasset MJ, O'Malley J, Pames JR et al. Frequency and cost of chemotherapy serious adverse effects in a population sample of women with breast cancer. J Nat Cancer Institute, Vol 98(16) 2006

Barcnas CH, Niu J, Zhang N et al. Risk of hospitalization according to chemotherapy regimen in early stage breast cancer. JCO 2014 (May ahead of publication)

Anlagen

Abstract #544

Prediction of late distant recurrence (DR) using the Prosigna® (PAM50) test in a Danish Breast Cancer Cooperative Group (DBCCG) cohort of postmenopausal women diagnosed with hormone receptor-positive (HR+) early breast cancer (EBC) allocated to 5-year of endocrine therapy (ET)



1. Department of Pathology, Region Zealand, Denmark; 2. The Danish Breast Cancer Cooperative Group, DBCC Secretariat, Rigshospitalet, Copenhagen, Denmark; Anne-Vibeke Lærnkholm^{1,2}, Maj-Britt Jensen^{1,3}, Jens Ole Eriksen^{1,3}, Torben Kibøll^{1,3}, Birgitte Bruun Rasmussen^{1,3}, Ann S. Knoop⁴, Sean Ferree⁵, Taryn Hafner⁶, Carl Schaper⁶, Bent Ejlersten^{1,2}

3. Department of Pathology, Herlev Hospital, Herlev, Denmark; 4. Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; 5. NanoString Technologies, Inc., Seattle, WA

Background

A number of tests have been developed to estimate prognosis in HR+ breast cancer^{1,2,3}; however, the prognostic significance of these assays are largely confined to the early recurrence period.⁶ HR+ EBC is a chronic disease with disease recurrence continuing beyond 10 years and accurate assessment of the risk of late DR (5-10 year after surgery) may be used to tailor the duration of ET for patients with HR+ EBC.

Prosigna (PAM50) is a decentralized test that uses a 50-gene expression profile from FFPE samples to assign breast cancer to one of four intrinsic subtypes: Luminal A, Luminal B, HER2-enriched, or Basal-like.⁴ The test also calculates a Prosigna Score (also known as ROR or 'Risk of Recurrence' Score), which has been clinically validated as a continuous predictor of risk of DR at 10 years in node-negative and node-positive (1-3 nodes) EBC patients treated with 5yr of endocrine therapy alone.^{4,5} The ROR score is used to construct ROR risk groups (low, intermediate, high) by using cut points based on pre-defined clinical criteria.

The Prosigna ROR score was shown to predict late DR for EBC patients treated with 5yr of ET in two randomized clinical trials.^{6,7,8} Here we assess the ability of PAM50 to predict late DR in a comprehensive nationwide cohort from Denmark consisting of postmenopausal women diagnosed with HR+ EBC allocated to 5-year of ET alone.⁹

Objective

To demonstrate that the ROR score provides clinically useful prognostic information in a multivariate analysis late recurrence for all HR+ postmenopausal patients treated with 5yr of ET (regardless of nodal status) who are distant recurrence-free at 5 years, both (a) as a continuous measure, and (b) using the pre-defined risk-groups.

Materials & Methods

2749 blocks were available from 2971 eligible patients diagnosed between 2000 and 2003 and PAM50 results were obtained in 2722 cases (99%). At 5yr 2164 patients (80%) were alive and disease free and analyzed for risk of late DR. Median follow-up was 4.58 years from completion of ET. Multivariate analyses tested the ability of PAM50 to predict late DR. Patients were categorized as Low, Intermediate, or High risk based upon pre-specified ROR cutoffs varied by number of positive nodes.

Table 1. Study population characteristics

Characteristic	Study Population N (%)	Late Cohort N (%)
Age		
50-59	1013 (37%)	847 (99%)
60-69	1142 (42%)	927 (43%)
≥70	567 (21%)	390 (18%)
Tumor Size		
≤10	260 (10%)	222 (10%)
11-20	1166 (43%)	972 (45%)
21-30	913 (34%)	687 (32%)
>30	383 (14%)	283 (13%)
Number of Pos Nodes		
0	1256 (46%)	1029 (48%)
1	809 (30%)	643 (30%)
2	426 (16%)	326 (15%)
3	231 (9%)	166 (8%)
Histological Subtype		
Ductal	2277 (84%)	1817 (84%)
Lobular	351 (13%)	275 (13%)
Other	94 (3%)	72 (3%)
Histological Grade		
1	644 (24%)	547 (25%)
2	1452 (53%)	1156 (53%)
3	355 (13%)	250 (12%)
Not Graded	268 (10%)	211 (10%)

Results

High risk patients had a late DR risk of 10.2% [95%CI: 8.0-12.7], compared to 6.1% [4.2-8.6] for intermediate and 2.4% [1.3-4.1] for Low risk patients. When ROR was added to a multivariable model including standard clinical and pathological variables it improved the prediction of late DR (likelihood ratio: p<0.0001; HR for a 20-point change =1.5 [1.2-1.9]). When the ROR risk groups were added to a multivariable model, they also improved the prediction of late DR (likelihood ratio P=0.0006).

Table 2. Late recurrence rates for Prosigna risk groups and subtypes

Patient Risk Group	N (%)	Prob of late DR [95%CI]	P values	
			Any Diff.	Diff. from Int.
High	870 (40%)	10.2% [8.0-12.7]		0.0091
Intermediate	650 (30%)	6.1% [4.2-8.6]	<0.0001	
Low	644 (30%)	2.4% [1.3-4.1]		0.0074
Subtype	N (%)	Prob of late DR [95%CI]	Any Diff. Diff. from LumA	
			Any Diff.	Diff. from LumA
Luminal A	1281 (59%)	4.5% [3.3-5.9]		-
Luminal B	733 (34%)	10.3% [7.8-13.1]	0.0002	<0.0001
Her2-Enriched	132 (6%)	8.8% [4.4-15.0]	-	0.0347
Basal-Like	18 (1%)	-	-	-

In all subgroups, the incidence curves were significantly different and in particular, the incidence curves for Luminal A and Luminal B patients were significantly different. As shown in table 2 and in figure 2, Luminal B and Her2-enriched patients had significantly worse outcomes than Luminal A patients. Unlike the analysis across the entire period, the late recurrence incidence for women with Her2-enriched tumors was lower than for women with Luminal B tumors. Outcomes for the basal-like subtype tumors were not evaluated due to the very small number of samples.

Conclusions

To our knowledge, this is the first genomic study of breast cancer on a comprehensive nationwide population. We have confirmed the ability of Prosigna (PAM50) to predict late DR for HR+ EBC patients regardless of nodal status in a real world cohort devoid of physician selection bias. PAM50 can reliably be utilized to identify patients who need, or most importantly may be safely spared, extended ET beyond 5 years.

Acknowledgements

The authors would like to thank NanoString Technologies, Inc. for support of this study.

References

1. Pak S, Shak S, Tang G, et al. N Engl J Med. 2011;364:2272-81.
2. Van de Vijver IJ, He YD, Witte V, et al. N Engl J Med. 2002;347:1999-2009.
3. Cuzick J, Dowsett M, Pineda S, et al. J Clin Oncol. 2011;29:4272-8.
4. Dowsett M, Sestak I, Lopez-Knowles E, et al. J Clin Oncol. 2013;31:2783-90.
5. Grant M, Ellis M, Goff F, et al. Ann Oncol. 2014;25:339-45.
6. Sestak I, Dowsett M, Zago L, et al. J Natl Cancer Inst. 2013;105:1504-11.
7. Fliedrick T, Nielsen TO, Rudis M, et al. Clin Cancer Res. 2011;17:398-405.
8. Sestak I, Cuzick J, Dowsett M, et al. J Clin Oncol. 2015;33:916-22.
9. Ejlersten B, Jensen MB, Mouridsen H. Acta Oncol. 2014;53:174-85.

This presentation is the intellectual property of the authors. Contact at nl@nanostring.com for permission to reprint and/or distribute.

Prediction of 10-year distant recurrence (DR) using the Prosigna® (PAM50) test in a Danish Breast Cancer Cooperative Group (DBCG) cohort of postmenopausal Danish women with hormone receptor-positive (HR+) early breast cancer (EBC) allocated to 5-year of endocrine therapy (ET) alone.



Anne-Vibeke Lønning, Maj-Britt Jensen, Jens Ole Eriksen, Birgitte Bruun Rasmussen, Ann S. Knopik, Steen Friese, Taryn Hafner, Carl Schrøder, Bent Ejlersen, 1. Department of Pathology, Region Zealand, Denmark; 2. The Danish Breast Cancer Cooperative Group, Decis Sekretariat, Rigshospitalet, Copenhagen, Denmark; 3. Department of Pathology, Herlev Hospital, Herlev, Denmark; 4. Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; 5. NanoString Technology, Inc., Seattle, WA

Background

A number of approaches have been developed to predict distant recurrence in ER-positive breast cancer patients. The Prosigna (PAM50) test, which is indicated for use with FFP6 tissue. The PAM50 algorithm uses a 50-gene expression profile to assign breast cancer to one of four molecular classes or intrinsic subtypes: Luminal A, Luminal B, HER2-enriched, or Basal-like.^{1,2}

The Prosigna Score (also known as ROR or Risk of Recurrence Score) has been clinically validated as a continuous predictor of risk of Distant Recurrence (DR) at 10 years in node-negative and node-positive breast cancer patients treated with tamoxifen or anastrozole.^{3,4} The ROR score is used to construct ROR risk groups (low, intermediate, high) by using cut points based on pre-defined clinical risk groups.

Accurate assessment of the DR risk may be used to tailor adjuvant treatment for patients with HR+ EBC. Here we assess the ability of PAM50 to predict DR in a comprehensive nationwide cohort from the DBCG. We evaluate whether the ROR score is able to predict DR in patients with HR+ EBC allocated to 5-year of ET alone.⁵

Objective

To evaluate whether the ROR score provides additional prognostic information for time to distant recurrence (TDR) at 10 years as compared to standard clinical variables in a real-world cohort of early stage breast cancer patients, both (a) as a continuous measure, and (b) using predefined risk-groups.

Materials & Methods

Using the population-based DBCG database we identified 2,749 patients with a confirmed diagnosis of primary breast cancer and follow-up data from 2000-2003. 82.5% of postmenopausal patients diagnosed from 2000-2003 (N = 2749/2971) who by nationwide guidelines were allocated to 5yr of ET alone. Patient characteristics are shown in Table 1. PAM50 was conducted using the NanoString nCounter Analysis System. Univariate and multivariate analyses tested the ability of PAM50 to predict DR. Patients were categorized as node-negative or node-positive based upon the number of lymph nodes found upon pre-specified ROR cutpoints varied by number of positive nodes.

Table 1. Patient characteristics

Characteristic	Study Population
Age (yr)	N (%)
50-59	1013 (37%)
60-69	1147 (42%)
≥70	567 (21%)
Tumor Size (mm)	
≤19	266 (10%)
20-29	853 (31%)
30-39	912 (34%)
≥40	385 (14%)
Number of Pos Nodes	
0	1229 (47%)
1	809 (30%)
2	476 (18%)
3	231 (8%)
Histological Subtype	
Ductal	2277 (84%)
Lobular	351 (13%)
Other	94 (3%)
Histological Grade	
1	644 (24%)
2	1452 (55%)
3	356 (13%)
Not graded	266 (10%)

Results

Blacks from 2,749 patients were identified and data from 2,722 samples (1,255 NO, 1,466 NA) were included in the analysis (95%). Median follow-up was 9.25 years. Adding ROR to a Fine and Gray's proportional sub-hazards model containing clinical and pathological variables significantly improved the model (likelihood ratio: p<0.0001; HR for a 20-point change in ROR = 1.7 [1.5 - 1.9]). As shown in Table 2, High-risk patients (n=1,200) had a significantly DR risk compared to Low risk patients (n=733). These figures were consistent across nodal status when risk group cutpoints were defined based upon the number of positive nodes. Analyses were also performed for node-negative and node-positive patients grouped as stage N1 (1-3 positive nodes) using the risk groups of the US cleared Prosigna test.

In this population, over 90% of the patients were Luminal A or Luminal B. The percentage of Luminal A patients was lower in the node-negative group, likely due to ineligibility of the lowest risk node-negative patients for endocrine therapy according to guidelines at the time. Luminal A subtype tumors had a significantly better outcome than Luminal B regardless of nodal status as shown in Table 3.

Figure 1. Cumulative incidence by risk group: all patients

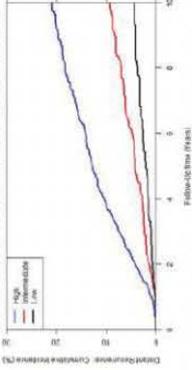


Figure 2. Cumulative incidence by luminal status: all patients

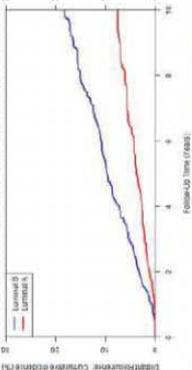


Table 2. Cumulative incidence by risk group and nodal status

Nodal Status	Risk Category	10-Year DR [95% CI]	P values
Node-Negative	High	18.5 [14.9-22.4]	<0.0001
	Intermediate	7.9 [4.6-10.5]	
Node-Positive	Low	4.9 [2.6-7.8]	0.1543
	High	21.9 [19.9-25.1]	
			4.8 [3.1-6.9]

* High-risk patients are grouped into all, based on the conventional Prosigna test in the US

Conclusions

To our knowledge, this is the first genomic study of breast cancer on a nationwide scale. The Prosigna (PAM50) accurately discriminates postmenopausal patients with ER positive breast cancer according to risk of recurrence. Using information from Prosigna and other clinicopathological factors a substantial proportion of postmenopausal patients in a real-world setting could be spared overtreatment with chemotherapy following appropriate endocrine therapy.

Table 3. Description of intrinsic subtypes by number of positive nodes

Intrinsic Subtype*	N [%]	Prob of 10-yr DR (95% CI)
Node-Positive (1-3 Nodes)		
Luminal A	632 (50.3%)	6.3% [4.4-8.8]
Luminal B	502 (40.0%)	14.1% [10.9-17.7]
Luminal A/B P-value		<0.0001

* Intrinsic subtypes is not reported by the Prosigna Assay in the United States.

References

1. Park S, Shuk S, Tang Q, et al. N. Engl J Med. 2017;376:2081-2091.
2. Sparano A, Gray R, Finn S, et al. J Clin Oncol. 2010;28:2091-2099.
3. Ditsch N, Finn S, Gray R, et al. J Clin Oncol. 2010;28:2091-2099.
4. Ditsch N, Finn S, Gray R, et al. J Clin Oncol. 2010;28:2091-2099.
5. Sparano A, Finn S, Gray R, et al. J Clin Oncol. 2010;28:2091-2099.

Acknowledgements

The authors would like to thank NanoString Technology, Inc. for supported this study.

The presentation is the intellectual property of the authors. Contact at anne.vibeke@region Zealand.dk for permission to reprint and/or distribute.

Validation of prediction of distant recurrence (DR) by Prosigna® (PAM50) in subgroups of a Danish Breast Cancer Cooperative Group (DBCG) cohort of node-positive (N1), hormone receptor positive (HR+), postmenopausal early breast cancer (EBC) patients allocated 5-year of endocrine therapy (ET)

Bent Ejlerisen^{1,2}, Maj-Britt Jensen¹, Jens Ole Eriksen³, Torben Kibjøl¹, Birgitte Bruun Rasmussen⁴, Ann S. Kroop², Sean Ferrée⁵, Taryn Halford⁶, Carl Schaper⁶, Anne-Vibeke Lærkeholm¹
 1. Danish Breast Cancer Cooperative Group (DBCG) Secretariat; 2. Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; 3. Department of Pathology, Region Zealand, Denmark; 4. Department of Pathology, Herlev Hospital, Herlev, Denmark; 5. NanoString Technologies, Inc., Seattle, WA

Background

A number of approaches have been developed to predict distant recurrence in HR+ breast cancer, but identification of N1 patients with sufficiently good outcome who may be spared chemotherapy remains an elusive goal.^{1,2} Prosigna (PAM50) is a decentralized test which is indicated for use with FFPE tissue. The algorithm uses a 50-gene expression profile to assign breast cancer to one of four intrinsic subtypes: Luminal A, Luminal B, HER2-enriched, or Basal-like.^{3,4} The Prosigna Score (also known as ROR or 'Risk of Recurrence' Score) has been clinically validated as a continuous predictor of risk of Distant Recurrence (DR) at 10 years in node-negative and node-positive (1-3 nodes) early stage breast cancer (ESBC) patients treated only with tamoxifen or anastrozole.^{3,4} The ROR score is used to construct ROR risk groups (low, intermediate, high) by using cut points based on pre-defined clinical criteria and was recently shown to identify a N+ patient population with less than a 10% probability of DR at 10 years.⁵ Here we assess the ability of PAM50 to predict 10-year DR for patients with 1, 2 or 3 positive nodes in a comprehensive nationwide cohort from Denmark to identify N1 postmenopausal women who may be spared overtreatment with CT. These patients were allocated to 5 yrs of an aromatase inhibitor (AI), tamoxifen or a sequence of these as the only systemic therapy.⁶

Objective

To demonstrate that the ROR score provides additional prognostic information for time to distant recurrence (TDR) at 10 years over and above standard clinical variables in patients with each of 1, 2 and 3 positive nodes, both (a) as a continuous measure, and (b) using predefined risk-groups representing patients with <10% (low), 10-20% (intermediate) and >20% (high) risk of distant recurrence.

Materials & Methods

Using the population based DBCG database primary FFPE tumor blocks and follow-up data were collected from >90% of Danish women diagnosed from 2000-2003 with HR+, postmenopausal, N1 EBC treated with ET alone (N=1,480). PAM50 on the NanoString nCounter® Analysis System categorized patients as Low, Intermediate, or High risk using pre-specified cutoffs varied by number of positive nodes. Multivariate analyses were performed to assess the ability of PAM50 to predict DR in patients with 1, 2, or 3+ nodes.

Table 1. Patient characteristics by # of positive nodes

Characteristic	Number of Positive Nodes		
	1	2	3
Age			
50-59	317 (35%)	172 (46%)	53 (40%)
60-69	328 (41%)	166 (39%)	55 (41%)
≥70	164 (20%)	88 (21%)	43 (19%)
Tumor Size			
≤1.0	81 (10%)	32 (8%)	11 (8%)
1.1-2.0	415 (51%)	185 (45%)	94 (41%)
2.1-3.0	209 (26%)	136 (32%)	76 (33%)
>3.0	104 (13%)	73 (17%)	50 (22%)
Histological Subtype			
Ductal	663 (82%)	352 (85%)	191 (83%)
Lobular	114 (14%)	62 (15%)	34 (15%)
Other	32 (4%)	12 (3%)	6 (3%)
Histological Grade			
1	274 (34%)	129 (30%)	77 (33%)
2	364 (45%)	198 (46%)	101 (44%)
3	85 (11%)	52 (12%)	28 (12%)
Not done	86 (11%)	47 (11%)	25 (11%)

Table 2. Risk Group Cutoffs by Number of Positive Nodes

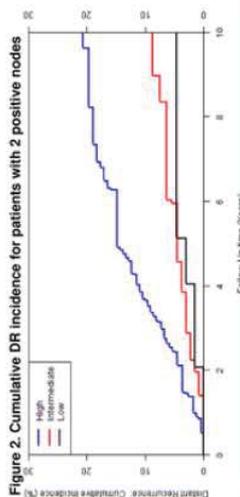
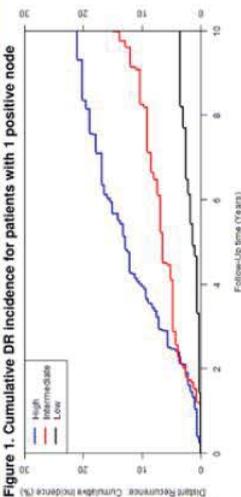
Number of Positive Nodes	Low-Risk Group	Intermediate-Risk Group	High-Risk Group
1	≤ 35	36 - 55	> 55
2	≤ 25	26 - 45	> 45
3	N/A	≤ 25	> 25

Results

Blocks from 1,480 patients were identified and data from 1,466 samples were included in the analysis (99%). Median follow-up was 9.25 years. A sequential primary analysis showed that both the continuous ROR score and the ROR risk categories added statistically significant prognostic information in multivariate cause-specific Cox proportional hazard regression models as well as competing-risks regression models using Fine and Gray's proportional sub-hazards models for patients with 1, 2 and 3 nodes. Table 3 and Figures 1 & 2 show the cumulative DR incidence from the competing risk models. Of note, the population with 3-positive nodes did not have a pre-specified low-risk group, but no DR events were observed in the intermediate-risk group.

Table 3. DR incidence by ROR-Based Risk Groups for N+ patients

Nodal Status (1-3 Positive Nodes)	Risk Category	N	10-Year DR [95% CI]	P values	
				Any Diff.	Diff. from Int.
All Patients (1-3 Positive Nodes)	High	703	22.4% [19.1-25.9]	<0.0001	-
	Intermediate	400	11.8% [8.4-15.9]	<0.0001	0.0007
	Low	363	3.6% [2.0-6.3]	0.0007	0.0202
1-Positive Node	High	274	21.0% [15.9-26.6]	<0.0001	-
	Int.	237	14.9% [9.9-20.9]	<0.0001	0.0034
	Low	298	3.6% [1.7-6.5]	0.0001	0.4165
2-Positive Nodes	High	224	20.7% [15.2-26.8]	0.0007	-
	Int.	137	8.7% [4.4-15.0]	0.0007	0.0086
	Low	65	4.6% [1.2-11.8]	-	-
3-Positive Nodes	High	205	26.1% [19.5-33.0]	0.0086	-
	Int.	26	0% [NA]	-	-
	Low	N/A	-	-	-



Conclusions

To our knowledge, this is the first genomic study of breast cancer on a comprehensive nationwide population. We have shown here that Prosigna improves outcome prediction in N1 patients over standard variables. Prosigna can identify at least 37% of patients with 1 and 15% with 2 positive nodes who have an excellent prognosis and may be spared adjuvant CT in a real world setting. This study confirms the observations made in the joined analysis of the TransATAC and ABCSG8 clinical studies⁶.

Acknowledgements

The authors would like to thank NanoString Technologies, Inc. for support of this study.



References

- Paik S, et al. N Engl J Med 2004;351:2817-26.
- Van de Vijver M, et al. N Engl J Med 2002;347:1999-2009
- Dowsett M, et al. J Clin Oncol 2013;31:2789-90
- Grant M, et al. Ann Oncol 2014;25:339-45
- Grant M, et al. Ann Oncol 2015 May 1 [epub ahead of print]
- Ejlertsen B, et al. Acta Oncol 2014; 53:174-85

This presentation is the intellectual property of the authors. Contact at bent.ejlerisen@dbccg.dk for permission to reprint and/or distribute.

UK OPTIMA-prelim study demonstrates economic value in more clinical evaluation of multi-parameter prognostic tests in early breast cancer

Dr Peter S Hall, Ms Alison F Smith, Mr Armando Vargas-Palacios, Dr Robert C. Stein, Prof John M. S. Bartlett, Dr Andrea Marshall, Dr Leila Rooshenas, Ms Amy Campbell, Prof Christopher Poole, Prof David A. Cameron, Dr Helena Earl, Miss Adele Francis, Dr Iain MacPherson, Prof Dan Rea, Prof Nigel Stallard, Dr Adrienne Morgan, Prof Sarah E. Pinder, Dr Claire Hulme, Dr Andreas Makris, Dr Luke Hughes-Davies, Prof Janet A Dunn, Prof Christopher McCabe.

Introduction

- In the UK adjuvant chemotherapy is routinely offered to patients with lymph node positive (LN+) early breast cancer.
- A sizeable portion of patients with oestrogen receptor positive (ER+) HER2 negative disease may not benefit from adjuvant chemotherapy due to biological insensitivity.
- Various multi-parameter tests have been developed to help identify patients unlikely to benefit from adjuvant chemotherapy.
- OPTIMA-prelim is the 2-year feasibility phase of an ongoing 5-year UK randomised controlled trial (RCT) which aims to assess the optimal strategy using multi-parameter tests, to guide the use of adjuvant chemotherapy in women aged 50-69 with LN+, ER+, HER2-negative early breast cancer.

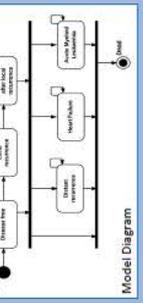
Objective

The aim of the OPTIMA-prelim economic analysis was to confirm value in an ongoing RCT and determine which tests should be included in the ongoing trial.

Methods

- A model-based cost-effectiveness analysis was conducted in line with UK National Institute for Health and Care Excellence (NICE) standards.
- Uncertainty was explored using probabilistic analysis: distributions were assigned to model parameters and 10,000 Monte Carlo simulations were run to estimate lifetime incremental cost and quality-adjusted life-year (QALY) results.
- Cost-effectiveness and the value of future research for each test was assessed using the following statistics:
 - Expected Net Health Benefit. This measures cost-effectiveness by expressing the overall benefit of each test in terms of QALYs, assuming a societal willingness to pay threshold value of £20,000 per QALY.
 - Expected value of perfect information (EVPI). This estimates the value of future research by assuming perfect total parameter information. It represents the maximum cost the NHS should be willing to spend on future research.
 - Expected value of perfect parameter information (EVppi). This measures the value of future research for a particular model parameter. It represents the maximum cost the NHS should be willing to spend on future research for that particular parameter.

Model Diagram



Selected Model Input Parameter Summaries

Parameter	Value	Source
Age (years)	55.5	UK Cancer Registry, 2010
Stage (I-IV)	0.25, 0.35, 0.25, 0.15	UK Cancer Registry, 2010
ER+ (yes/no)	0.85/0.15	UK Cancer Registry, 2010
HER2+ (yes/no)	0.15/0.85	UK Cancer Registry, 2010
LN+ (yes/no)	0.35/0.65	UK Cancer Registry, 2010
Survival (months)	12.5, 15.5, 18.5, 21.5, 24.5, 27.5, 30.5, 33.5, 36.5, 39.5, 42.5, 45.5, 48.5, 51.5, 54.5, 57.5, 60.5, 63.5, 66.5, 69.5, 72.5, 75.5, 78.5, 81.5, 84.5, 87.5, 90.5, 93.5, 96.5, 99.5, 102.5, 105.5, 108.5, 111.5, 114.5, 117.5, 120.5, 123.5, 126.5, 129.5, 132.5, 135.5, 138.5, 141.5, 144.5, 147.5, 150.5, 153.5, 156.5, 159.5, 162.5, 165.5, 168.5, 171.5, 174.5, 177.5, 180.5, 183.5, 186.5, 189.5, 192.5, 195.5, 198.5, 201.5, 204.5, 207.5, 210.5, 213.5, 216.5, 219.5, 222.5, 225.5, 228.5, 231.5, 234.5, 237.5, 240.5, 243.5, 246.5, 249.5, 252.5, 255.5, 258.5, 261.5, 264.5, 267.5, 270.5, 273.5, 276.5, 279.5, 282.5, 285.5, 288.5, 291.5, 294.5, 297.5, 300.5, 303.5, 306.5, 309.5, 312.5, 315.5, 318.5, 321.5, 324.5, 327.5, 330.5, 333.5, 336.5, 339.5, 342.5, 345.5, 348.5, 351.5, 354.5, 357.5, 360.5, 363.5, 366.5, 369.5, 372.5, 375.5, 378.5, 381.5, 384.5, 387.5, 390.5, 393.5, 396.5, 399.5, 402.5, 405.5, 408.5, 411.5, 414.5, 417.5, 420.5, 423.5, 426.5, 429.5, 432.5, 435.5, 438.5, 441.5, 444.5, 447.5, 450.5, 453.5, 456.5, 459.5, 462.5, 465.5, 468.5, 471.5, 474.5, 477.5, 480.5, 483.5, 486.5, 489.5, 492.5, 495.5, 498.5, 501.5, 504.5, 507.5, 510.5, 513.5, 516.5, 519.5, 522.5, 525.5, 528.5, 531.5, 534.5, 537.5, 540.5, 543.5, 546.5, 549.5, 552.5, 555.5, 558.5, 561.5, 564.5, 567.5, 570.5, 573.5, 576.5, 579.5, 582.5, 585.5, 588.5, 591.5, 594.5, 597.5, 600.5, 603.5, 606.5, 609.5, 612.5, 615.5, 618.5, 621.5, 624.5, 627.5, 630.5, 633.5, 636.5, 639.5, 642.5, 645.5, 648.5, 651.5, 654.5, 657.5, 660.5, 663.5, 666.5, 669.5, 672.5, 675.5, 678.5, 681.5, 684.5, 687.5, 690.5, 693.5, 696.5, 699.5, 702.5, 705.5, 708.5, 711.5, 714.5, 717.5, 720.5, 723.5, 726.5, 729.5, 732.5, 735.5, 738.5, 741.5, 744.5, 747.5, 750.5, 753.5, 756.5, 759.5, 762.5, 765.5, 768.5, 771.5, 774.5, 777.5, 780.5, 783.5, 786.5, 789.5, 792.5, 795.5, 798.5, 801.5, 804.5, 807.5, 810.5, 813.5, 816.5, 819.5, 822.5, 825.5, 828.5, 831.5, 834.5, 837.5, 840.5, 843.5, 846.5, 849.5, 852.5, 855.5, 858.5, 861.5, 864.5, 867.5, 870.5, 873.5, 876.5, 879.5, 882.5, 885.5, 888.5, 891.5, 894.5, 897.5, 900.5, 903.5, 906.5, 909.5, 912.5, 915.5, 918.5, 921.5, 924.5, 927.5, 930.5, 933.5, 936.5, 939.5, 942.5, 945.5, 948.5, 951.5, 954.5, 957.5, 960.5, 963.5, 966.5, 969.5, 972.5, 975.5, 978.5, 981.5, 984.5, 987.5, 990.5, 993.5, 996.5, 999.5, 1002.5, 1005.5, 1008.5, 1011.5, 1014.5, 1017.5, 1020.5, 1023.5, 1026.5, 1029.5, 1032.5, 1035.5, 1038.5, 1041.5, 1044.5, 1047.5, 1050.5, 1053.5, 1056.5, 1059.5, 1062.5, 1065.5, 1068.5, 1071.5, 1074.5, 1077.5, 1080.5, 1083.5, 1086.5, 1089.5, 1092.5, 1095.5, 1098.5, 1101.5, 1104.5, 1107.5, 1110.5, 1113.5, 1116.5, 1119.5, 1122.5, 1125.5, 1128.5, 1131.5, 1134.5, 1137.5, 1140.5, 1143.5, 1146.5, 1149.5, 1152.5, 1155.5, 1158.5, 1161.5, 1164.5, 1167.5, 1170.5, 1173.5, 1176.5, 1179.5, 1182.5, 1185.5, 1188.5, 1191.5, 1194.5, 1197.5, 1200.5, 1203.5, 1206.5, 1209.5, 1212.5, 1215.5, 1218.5, 1221.5, 1224.5, 1227.5, 1230.5, 1233.5, 1236.5, 1239.5, 1242.5, 1245.5, 1248.5, 1251.5, 1254.5, 1257.5, 1260.5, 1263.5, 1266.5, 1269.5, 1272.5, 1275.5, 1278.5, 1281.5, 1284.5, 1287.5, 1290.5, 1293.5, 1296.5, 1299.5, 1302.5, 1305.5, 1308.5, 1311.5, 1314.5, 1317.5, 1320.5, 1323.5, 1326.5, 1329.5, 1332.5, 1335.5, 1338.5, 1341.5, 1344.5, 1347.5, 1350.5, 1353.5, 1356.5, 1359.5, 1362.5, 1365.5, 1368.5, 1371.5, 1374.5, 1377.5, 1380.5, 1383.5, 1386.5, 1389.5, 1392.5, 1395.5, 1398.5, 1401.5, 1404.5, 1407.5, 1410.5, 1413.5, 1416.5, 1419.5, 1422.5, 1425.5, 1428.5, 1431.5, 1434.5, 1437.5, 1440.5, 1443.5, 1446.5, 1449.5, 1452.5, 1455.5, 1458.5, 1461.5, 1464.5, 1467.5, 1470.5, 1473.5, 1476.5, 1479.5, 1482.5, 1485.5, 1488.5, 1491.5, 1494.5, 1497.5, 1500.5, 1503.5, 1506.5, 1509.5, 1512.5, 1515.5, 1518.5, 1521.5, 1524.5, 1527.5, 1530.5, 1533.5, 1536.5, 1539.5, 1542.5, 1545.5, 1548.5, 1551.5, 1554.5, 1557.5, 1560.5, 1563.5, 1566.5, 1569.5, 1572.5, 1575.5, 1578.5, 1581.5, 1584.5, 1587.5, 1590.5, 1593.5, 1596.5, 1599.5, 1602.5, 1605.5, 1608.5, 1611.5, 1614.5, 1617.5, 1620.5, 1623.5, 1626.5, 1629.5, 1632.5, 1635.5, 1638.5, 1641.5, 1644.5, 1647.5, 1650.5, 1653.5, 1656.5, 1659.5, 1662.5, 1665.5, 1668.5, 1671.5, 1674.5, 1677.5, 1680.5, 1683.5, 1686.5, 1689.5, 1692.5, 1695.5, 1698.5, 1701.5, 1704.5, 1707.5, 1710.5, 1713.5, 1716.5, 1719.5, 1722.5, 1725.5, 1728.5, 1731.5, 1734.5, 1737.5, 1740.5, 1743.5, 1746.5, 1749.5, 1752.5, 1755.5, 1758.5, 1761.5, 1764.5, 1767.5, 1770.5, 1773.5, 1776.5, 1779.5, 1782.5, 1785.5, 1788.5, 1791.5, 1794.5, 1797.5, 1800.5, 1803.5, 1806.5, 1809.5, 1812.5, 1815.5, 1818.5, 1821.5, 1824.5, 1827.5, 1830.5, 1833.5, 1836.5, 1839.5, 1842.5, 1845.5, 1848.5, 1851.5, 1854.5, 1857.5, 1860.5, 1863.5, 1866.5, 1869.5, 1872.5, 1875.5, 1878.5, 1881.5, 1884.5, 1887.5, 1890.5, 1893.5, 1896.5, 1899.5, 1902.5, 1905.5, 1908.5, 1911.5, 1914.5, 1917.5, 1920.5, 1923.5, 1926.5, 1929.5, 1932.5, 1935.5, 1938.5, 1941.5, 1944.5, 1947.5, 1950.5, 1953.5, 1956.5, 1959.5, 1962.5, 1965.5, 1968.5, 1971.5, 1974.5, 1977.5, 1980.5, 1983.5, 1986.5, 1989.5, 1992.5, 1995.5, 1998.5, 2001.5, 2004.5, 2007.5, 2010.5, 2013.5, 2016.5, 2019.5, 2022.5, 2025.5, 2028.5, 2031.5, 2034.5, 2037.5, 2040.5, 2043.5, 2046.5, 2049.5, 2052.5, 2055.5, 2058.5, 2061.5, 2064.5, 2067.5, 2070.5, 2073.5, 2076.5, 2079.5, 2082.5, 2085.5, 2088.5, 2091.5, 2094.5, 2097.5, 2100.5, 2103.5, 2106.5, 2109.5, 2112.5, 2115.5, 2118.5, 2121.5, 2124.5, 2127.5, 2130.5, 2133.5, 2136.5, 2139.5, 2142.5, 2145.5, 2148.5, 2151.5, 2154.5, 2157.5, 2160.5, 2163.5, 2166.5, 2169.5, 2172.5, 2175.5, 2178.5, 2181.5, 2184.5, 2187.5, 2190.5, 2193.5, 2196.5, 2199.5, 2202.5, 2205.5, 2208.5, 2211.5, 2214.5, 2217.5, 2220.5, 2223.5, 2226.5, 2229.5, 2232.5, 2235.5, 2238.5, 2241.5, 2244.5, 2247.5, 2250.5, 2253.5, 2256.5, 2259.5, 2262.5, 2265.5, 2268.5, 2271.5, 2274.5, 2277.5, 2280.5, 2283.5, 2286.5, 2289.5, 2292.5, 2295.5, 2298.5, 2301.5, 2304.5, 2307.5, 2310.5, 2313.5, 2316.5, 2319.5, 2322.5, 2325.5, 2328.5, 2331.5, 2334.5, 2337.5, 2340.5, 2343.5, 2346.5, 2349.5, 2352.5, 2355.5, 2358.5, 2361.5, 2364.5, 2367.5, 2370.5, 2373.5, 2376.5, 2379.5, 2382.5, 2385.5, 2388.5, 2391.5, 2394.5, 2397.5, 2400.5, 2403.5, 2406.5, 2409.5, 2412.5, 2415.5, 2418.5, 2421.5, 2424.5, 2427.5, 2430.5, 2433.5, 2436.5, 2439.5, 2442.5, 2445.5, 2448.5, 2451.5, 2454.5, 2457.5, 2460.5, 2463.5, 2466.5, 2469.5, 2472.5, 2475.5, 2478.5, 2481.5, 2484.5, 2487.5, 2490.5, 2493.5, 2496.5, 2499.5, 2502.5, 2505.5, 2508.5, 2511.5, 2514.5, 2517.5, 2520.5, 2523.5, 2526.5, 2529.5, 2532.5, 2535.5, 2538.5, 2541.5, 2544.5, 2547.5, 2550.5, 2553.5, 2556.5, 2559.5, 2562.5, 2565.5, 2568.5, 2571.5, 2574.5, 2577.5, 2580.5, 2583.5, 2586.5, 2589.5, 2592.5, 2595.5, 2598.5, 2601.5, 2604.5, 2607.5, 2610.5, 2613.5, 2616.5, 2619.5, 2622.5, 2625.5, 2628.5, 2631.5, 2634.5, 2637.5, 2640.5, 2643.5, 2646.5, 2649.5, 2652.5, 2655.5, 2658.5, 2661.5, 2664.5, 2667.5, 2670.5, 2673.5, 2676.5, 2679.5, 2682.5, 2685.5, 2688.5, 2691.5, 2694.5, 2697.5, 2700.5, 2703.5, 2706.5, 2709.5, 2712.5, 2715.5, 2718.5, 2721.5, 2724.5, 2727.5, 2730.5, 2733.5, 2736.5, 2739.5, 2742.5, 2745.5, 2748.5, 2751.5, 2754.5, 2757.5, 2760.5, 2763.5, 2766.5, 2769.5, 2772.5, 2775.5, 2778.5, 2781.5, 2784.5, 2787.5, 2790.5, 2793.5, 2796.5, 2799.5, 2802.5, 2805.5, 2808.5, 2811.5, 2814.5, 2817.5, 2820.5, 2823.5, 2826.5, 2829.5, 2832.5, 2835.5, 2838.5, 2841.5, 2844.5, 2847.5, 2850.5, 2853.5, 2856.5, 2859.5, 2862.5, 2865.5, 2868.5, 2871.5, 2874.5, 2877.5, 2880.5, 2883.5, 2886.5, 2889.5, 2892.5, 2895.5, 2898.5, 2901.5, 2904.5, 2907.5, 2910.5, 2913.5, 2916.5, 2919.5, 2922.5, 2925.5, 2928.5, 2931.5, 2934.5, 2937.5, 2940.5, 2943.5, 2946.5, 2949.5, 2952.5, 2955.5, 2958.5, 2961.5, 2964.5, 2967.5, 2970.5, 2973.5, 2976.5, 2979.5, 2982.5, 2985.5, 2988.5, 2991.5, 2994.5, 2997.5, 3000.5, 3003.5, 3006.5, 3009.5, 3012.5, 3015.5, 3018.5, 3021.5, 3024.5, 3027.5, 3030.5, 3033.5, 3036.5, 3039.5, 3042.5, 3045.5, 3048.5, 3051.5, 3054.5, 3057.5, 3060.5, 3063.5, 3066.5, 3069.5, 3072.5, 3075.5, 3078.5, 3081.5, 3084.5, 3087.5, 3090.5, 3093.5, 3096.5, 3099.5, 3102.5, 3105.5, 3108.5, 3111.5, 3114.5, 3117.5, 3120.5, 3123.5, 3126.5, 3129.5, 3132.5, 3135.5, 3138.5, 3141.5, 3144.5, 3147.5, 3150.5, 3153.5, 3156.5, 3159.5, 3162.5, 3165.5, 3168.5, 3171.5, 3174.5, 3177.5, 3180.5, 3183.5, 3186.5, 3189.5, 3192.5, 3195.5, 3198.5, 3201.5, 3204.5, 3207.5, 3210.5, 3213.5, 3216.5, 3219.5, 3222.5, 3225.5, 3228.5, 3231.5, 3234.5, 3237.5, 3240.5, 3243.5, 3246.5, 3249.5, 3252.5, 3255.5, 3258.5, 3261.5, 3264.5, 3267.5, 3270.5, 3273.5, 3276.5, 3279.5, 3282.5, 3285.5, 3288.5, 3291.5, 3294.5, 3297.5, 3300.5, 3303.5, 3306.5, 3309.5, 3312.5, 3315.5, 3318.5, 3321.5, 3324.5, 3327.5, 3330.5, 3333.5, 3336.5, 3339.5, 3342.5, 3345.5, 3348.5, 3351.5, 3354.5, 3357.5, 3360.5, 3363.5, 3366.5, 3369.5, 3372.5, 3375.5, 3378.5, 3381.5, 3384.5, 3387.5, 3390.5, 3393.5, 3396.5, 3399.5, 3402.5, 3405.5, 3408.5, 3411.5, 3414.5, 3417.5, 3420.5, 3423.5, 3426.5, 3429.5, 3432.5, 3435.5, 3438.5, 3441.5, 3444.5, 3447.5, 3450.5, 3453.5, 3456.5, 3459.5, 3462.5, 3465.5, 3468.5, 3471.5, 3474.5, 3477.5, 3480.5, 3483.5, 3486.5, 3489.5, 3492.5, 3495.5, 3498.5, 3501.5, 3504.5, 3507.5, 3510.5, 3513.5, 3516.5, 3519.5, 3522.5, 3525.5, 3528.5, 3531.5, 3534.5, 3537.5, 3540.5, 3543.5, 3546.5, 3549.5, 3552.5, 3555.5, 3558.5, 3561.5, 3564.5, 3567.5, 3570.5, 3573.5, 3576.5, 3579.5, 3582.5, 3585.5, 3588.5, 3591.5, 3594.5, 3597.5, 3600.5, 360	

A.1.8 – Sividon Diagnostics GmbH

Autoren:

- Kronenwett, Ralf
- Petry, Christoph

Sividon Diagnostics GmbH Nattermannallee 1 50829 Köln

 Institut für Qualität und Wirtschaftlichkeit
 im Gesundheitswesen
 - IQWiG Stellungnahme zum Vorbericht D14-01
 Prof. Dr. med. Windeler
 Im Mediapark 8

50829 Köln

 Sividon Diagnostics GmbH
 Nattermannallee 1 | S19
 50829 Köln

 T +49 (0)221 669 561 00
 F +49 (0)221 669 561 99

www.sividon.com

info@sividon.com

08. Dez. 2015

04.12.2015

Ansprechpartner
 Dr. Christoph Petry
 Durchwahl: - [REDACTED]

Ressort	Info	Rückspr.	Bearb.	Verbleib	Ressort	Info	Rückspr.	Bearb.	Verbleib
L					GI				
AM					Komm				
NMW					VW				
VQ					GO				
Biometrie					Stb. IM				
Stb. Recht					Stb. QS				

D14-01: Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom

Stellungnahme zum Vorbericht

Sehr geehrter Herr Professor Windeler,

anbei finden Sie unsere schriftliche Stellungnahme zum Vorbericht D14-01 vom 02.11.2015 („Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom“). Unserer Meinung nach werden zentrale, international anerkannte klinische Studien im Vorbericht nicht oder nicht ausreichend berücksichtigt, ohne dass uns der Grund für die Auswahl des IQWiGs klar geworden ist. Des Weiteren kann die projektspezifische Methodik teilweise nicht nachvollzogen werden und weist unserer Auffassung nach an zentralen Stellen wissenschaftliche Schwächen auf. Daher halten wir eine grundlegende Überarbeitung des Vorberichts für geboten. Im Folgenden legen wir Ihnen unsere Einwände im Detail dar.

 Geschäftsführer:
 Dr. Christoph Petry
 Dr. Georg Kox

Amtsgericht Köln

USt-Id. [REDACTED]

 IBAN: [REDACTED]
 BIC: [REDACTED]

1. Bewertung und Interpretation der im Vorbericht eingeschlossenen Studien

Von den vier klinischen Validierungsstudien zu unserem Test EndoPredict ist lediglich eine in die Bewertung eingeschlossen worden. Es handelt sich dabei um die prospektiv-retrospektive Studie von Martin et al. 2014 (GEICAM/9906) [1]. Leider wurde das Ergebnis dieser Studie im Vorbericht nicht korrekt interpretiert. Während das primäre Studienziel, nämlich der Nachweis einer prognostischen Leistungsfähigkeit des EndoPredict-Tests bei prä- und postmenopausalen Patientinnen mit Mammakarzinom, bei der Nutzenbewertung keine Berücksichtigung fand, wurde ein exploratives Ergebnis der Studie nicht korrekt wiedergegeben und dadurch entsprechend fehlerhaft interpretiert: Im Vorbericht steht auf S. 11 „Die Interaktion des EndoPredict-Ergebnisses mit der Variable Behandlung mit Chemotherapie war nicht signifikant ($p = 0,71$).“ Korrekterweise lautet das Ergebnis jedoch: „The interaction test results between the chemotherapy arm and the EP score were not significant“.

Das ist ein elementarer Unterschied. Die Interpretation des IQWiG impliziert, dass der EndoPredict basierend auf den Studiendaten keinen generellen Vorteil von einer Chemotherapie vorhersagen kann, während in der Studie jedoch gezeigt wurde, dass EndoPredict keinen Unterschied zwischen zwei Chemotherapie-Regimen (FEC vs. FEC+Paclitaxel) prädizierte. Dieses Ergebnis ist auch nicht verwunderlich, da der Test für die Fragestellung, ob Paclitaxel wirkt, nicht entwickelt wurde und dies auch nicht dem vorgesehenen Verwendungszweck des Tests entspricht.

Daher bitten wir Sie, das primäre Studienziel der GEICAM/9906-Biomarkerstudie bei der Bewertung zu berücksichtigen, und die Ergebnisse der explorativen Analyse zur Prädiktivität eines Paclitaxel-Benefits entsprechend dem Inhalt der Publikation darzustellen, um irreführende Schlussfolgerungen bezüglich der Leistungsfähigkeit des EndoPredict-Tests zu vermeiden.

2. Benennung von zusätzlichen, im Vorbericht nicht genannten, relevanten Studien

Neben der o.g. GEICAM/9906-Studie gibt es drei weitere prospektiv-retrospektive klinische Validierungsstudien für EndoPredict, die im Vorbericht nicht oder nicht ausreichend berücksichtigt wurden. Alle drei Studien inkludierten die Patientinnen-Zielgruppe der zur Begutachtung stehenden Biomarker-basierten Tests: Patientinnen mit

ER+, HER2- primärem Mammakarzinom und 0 bis 3 befallenen Lymphknoten. Bei den drei Studien handelt es sich um die **ABCSG-6**, **ABCSG-8** und die **ATAC-Studie**.

Alle drei Studien, in denen die Patientinnen systemisch nur mit endokriner Therapie behandelt wurden, zeigten konsistent die langfristige (10 Jahre) prognostische Leistungsfähigkeit des EndoPredict sowohl in nodal-negativen als auch in nodal-positiven Patientinnen (siehe auch Abschnitt 4b unserer Stellungnahme). Die Ergebnisse der **ABCSG-6-** und **ABCSG-8-Studien** sind in der „peer-reviewed“ Literatur publiziert [2,3,4]. Außerdem haben wir als Hersteller des EndoPredict dem IQWiG verschiedene Subgruppenanalysen zur Verfügung gestellt. Leider wurden die beiden Studien im Vorbericht nicht erwähnt. Die o.g. drei zentralen Publikationen zu diesen Studien wurden in die „**Liste der ausgeschlossenen Dokumente mit Ausschlussgründen (Prognosestudien)**“ aufgenommen aber nicht begutachtet. Als Begründung für den Ausschluss wird „Nicht E4 (Studientyp)“ angeführt, was nicht nachzuvollziehen ist, da es sich bei beiden Studien um prospektiv geplante Kohorten-Studien (prospektiv-retrospektive Prognosestudien) handelt, die laut Berichtsplan (E4b) prinzipiell zugelassen sind.

Beide Studien wurden durch nationale (AGO) und internationale (ESMO) Leitlinienkommissionen für die Bewertung der klinischen Validität und Nützlichkeit prognostischer Tests anerkannt, was zur Anwendungsempfehlung von EndoPredict bei Patientinnen mit ER+, HER2- Mammakarzinom führte, bei denen die konventionellen Risikoparameter keine Entscheidung für oder gegen Chemotherapie ermöglichen (siehe Abbildung 1; [5,6]).

Die dritte nicht berücksichtigte klinische prospektiv-retrospektive Validierungsstudie für den EndoPredict-Test ist die **ATAC-Studie**, in der die prognostische Leistungsfähigkeit von EndoPredict erneut gezeigt und seine Überlegenheit gegenüber dem Oncotype DX direkt nachgewiesen werden konnte. Diese Studie ist noch nicht publiziert, die Ergebnisse werden jedoch bei der diesjährigen San Antonio-Brustkrebskonferenz in einem Hauptvortrag am 10.12.2015 präsentiert werden [7]. Ein Manuskript zur Publikation wurde eingereicht. Die Daten wurden dem IQWiG bereits für den Vorbericht zur Verfügung gestellt.

Da ein Ausschluss der drei Studien ABCSG-6, ABCSG-8 und ATAC aus dem Vorbericht nicht nachzuvollziehen ist, bitten wir, diese Studien bei der Nutzenbewertung zu berücksichtigen.

Table 6. Summary of biomarkers used in treatment decision making

Biomarker	Prognostic	Predictive	Technical validation	Clinical validation	Test and scoring recommendations	Patient selection
ER	++	+++	Yes LOE IB	Yes	IHC	Hormonal treatment
PgR	+++	+	Yes LOE IB	No	IHC	If negative, chemotherapy in some cases
HER2	++	+++	Yes LOE IB	Yes	IHC $\geq 10\%$ c/dls with complete membrane staining ISH: number of HER2 gene copies ≥ 6 or the ratio HER2/chromosome 17 ≥ 2	Anti-HER2 treatment
Ki67	++	+	No	No	IHC no final consensus on cut-off but values below 10% are considered low and above 30% are high	Chemotherapy if elevated
Intrinsic subtypes	++	++	Yes	Yes	Gene expression profile (not for IHC surrogates)	Different responses to neoadjuvant chemotherapy according to the subtype
First generation signatures (MammaPrint, Oncotype Dx)	+++	++	Yes	Validated retrospectively in prospective clinical trials, prospective clinical validation ongoing	Gene expression profile, RT-PCR	Chemotherapy if high risk or high score
Second generation signatures (Prosigna, Endopredict [®])	++	++	Yes	Validated retrospectively in prospective clinical trials	N-Counter TM technology, RT-PCR	Prognosis, chemotherapy if high risk or high score

ER, oestrogen receptor; IHC, immunohistochemistry; RT-PCR, reverse transcription polymerase chain reaction; LOE, level of evidence; PgR, progesterone receptor; ISH, *in situ* hybridisation; HER2, human epidermal growth factor 2 receptor.

Abbildung 1: Anerkennung prospektive-retrospektiver klinischer Validierungsstudien bei Genexpressionstests und Empfehlung zur Verwendung bei der Entscheidung für oder gegen Chemotherapie durch die Leitlinien der ESMO (European Society of Clinical Oncology) [6].

3. Verweise auf andere qualitativ angemessene Unterlagen

Keine

4. Anmerkungen zur projektspezifischen Methodik

Die projektspezifische Methodik, die im Vorbericht zur Anwendung kam, weist nach unserer Auffassung in den folgenden zentralen Punkten wissenschaftliche Schwächen auf.

a) Studiendauer bei RCTs (A.2.1.1.5 Studiendauer)

Als Einschlusskriterium für RCTs wird eine Mindestdauer der Nachbeobachtung von einem Jahr gefordert. Aufgrund der Tatsache, dass beim Hormonrezeptor-positiven Mammakarzinom ein Großteil der Metastasen erst im Zeitintervall nach 5 Jahren auftreten [8], ist eine Mindestdauer der Nachbeobachtung von nur einem Jahr nicht ausreichend und gefährdet die Sicherheit der Patientinnen.

Daher schlagen wir für RCTs als Einschlusskriterium eine Mindestdauer der Nachbeobachtung von 10 Jahren vor (Kaplan-Meier-Schätzung mit Konfidenzintervallen nach 10 Jahren).

b) Studiendauer bei Prognosestudien (A.2.1.2.5 Studiendauer)

Als Einschlusskriterium für Prognosestudien wird im Vorbericht eine Mindestdauer der Nachbeobachtung von 10 Jahren gefordert, wobei die Mindestdauer über die mediane Dauer der Nachbeobachtung von 10 Jahren definiert wird. Während wir wegen der im Abschnitt 4a) genannten Gründe zustimmen, dass eine Mindestdauer der Nachbeobachtung von 10 Jahren im Sinne der Patientinnensicherheit notwendig ist, stimmen wir nicht zu, dass dies sinnvoll über die mediane Dauer der Nachbeobachtung definiert werden kann. Um mit einer Kaplan-Meier-Schätzung eine sichere Aussage zum Überleben nach 10 Jahren machen zu können, ist nicht die mediane Nachbeobachtungszeit, sondern das Konfidenzintervall nach 10 Jahren entscheidend. Darauf beruht ja das Prinzip der Kaplan-Meier-Analyse.

Daher schlagen wir für eine sichere Nutzenbewertung der biomarkerbasierten Tests wie für die RCTs auch für Prognosestudien eine Mindestdauer der Nachbeobachtung von 10 Jahren basierend auf der Kaplan-Meier-Schätzung mit Konfidenzintervallen vor. Die Forderung einer definierten medianen Nachbeobachtungszeit ist jedoch nicht sinnvoll, was auch in der Literatur bestätigt wird („Median follow-up is not a valid or useful scientific term, and should not appear in cancer research publications“ [9]).

c) Bei Prognosestudien Vorhersage des Rezidivrisikos mit großer Wahrscheinlichkeit unter 5 %, (A.2.1.2, S. 21 des Vorberichts)

Laut Anforderung des IQWiG an die biomarkerbasierten Tests muss das Rezidivrisiko nach 10 Jahren unter 5% (oberes Ende des Konfidenzintervalls) liegen. Dieser Grenzwert ist im Vorbericht nach unserer Meinung nicht sinnvoll begründet und widerspricht dem internationalen wissenschaftlichen Konsens von 10% und dem Vorschlag der Stellungnehmenden bei der Erörterung des Berichtsplans (ebenfalls 10%). Es gibt im Vorbericht keinen Literaturverweis, mit dem der Grenzwert von 5% begründet werden kann. Stattdessen verwendet das IQWiG ein eigenes Rechenmodell und kommt zu dem Schluss, dass 10% zu hoch sei, um auf eine Chemotherapie zu verzichten. In den Ausführungen wird nicht klar, in welcher Höhe die Chemotherapie-assoziierte Mortalität (akut und als Folge von

Langzeitnebenwirkungen innerhalb von 10 Jahren) in das Modell eingerechnet wird. Klinische Studien haben gezeigt, dass die akute Mortalität durch adjuvante Chemotherapie beim Mammakarzinom etwa 0,3% beträgt [10] und das Risiko für schwere Langzeitnebenwirkungen innerhalb von 10 Jahren größer als 6% ist (schwere kardiale Ereignisse 3%, kongestive Herzinsuffizienz Grad 3 und 4 3%, sekundäre AML/MDS 0.5%) [11]. Zu beachten ist, dass das mediane Alter der Patientinnen in dieser Studie bei 49 lag. Der Großteil der Patientinnen mit Brustkrebs ist jedoch über 60 Jahre, so dass aufgrund von Komorbiditäten das akute Mortalitätsrisiko durch Chemotherapie bei den meisten Patientinnen höher sein wird. In einer eigenen Studie, in der wir die Nebenwirkungen von Chemotherapie bei Patientinnen mit Brustkrebs innerhalb des ersten Jahres an einem deutschen Brustzentrum erfasst haben, traten akute Grad 3 und 4 Toxizitäten mit notwendiger Hospitalisierung bei 37% der Patientinnen auf, die Sterblichkeit lag bei 2% (wegen Pneumonie und Sepsis) [12]. Dem gegenüber steht bei 5% Metastasenrisiko ein erwarteter Benefit durch die Chemotherapie von etwa 1,5% [13].

Daher bleiben wir bei unserer während der Erörterung vertretenen Position, dass 10% ein wissenschaftlich und medizinisch sinnvoller Cut-off für die Chemotherapie-Indikation ist, da sich hier der erwartete Benefit durch die Chemotherapie mit dem möglichen Schaden die Waage hält. 10% sind darüber hinaus ein publizierter internationaler Konsens [14].

d) Bei Prognosestudien muss bei einem relevanten Anteil aller Patientinnen (mindestens 15 %) der Test negativ sein. (A.2.1.2, S. 21 des Vorberichts)

Diese Anforderung an einen biomarkerbasierten Test war im vorläufigen Berichtsplan nicht enthalten, sondern erst in der finalen Version, so dass wir das nicht früher kommentieren konnten. Im finalen Berichtsplan und im Vorbericht gibt es keine Begründung und keine Referenz für die Anforderung, dass mindestens 15% der Patientinnen Test-negativ sein müssen. Wir halten diesen Wert von 15% weder aus gesundheitsökonomischen noch aus medizinischen Gründen für angezeigt. Nicht zuletzt würde ein Test, der nur einen so kleinen Anteil der Patientinnen von der Chemotherapie-Indikation aussparen würde, tatsächlich einen Rückschritt gegenüber den derzeitigen Stratifizierungsmethoden darstellen. Da in der Zielgruppe für die biomarkerbasierten Tests durch klinische Leitlinien bereits 19% klar als Niedrigrisiko klassifiziert werden können [3] und in der klinischen Praxis je nach Brustzentrum 30-50% der Zielgruppe keine adjuvante Chemotherapie bekommen [12,15], würde eine Größe der Niedrigrisiko-Gruppe von 15% die Anwendung von Chemotherapie und damit die Übertherapie von Brustkrebs in Deutschland erheblich erhöhen – geschätzt etwa verdoppeln.

Darüber hinaus macht aus medizinischer Sicht eine isolierte Vorgabe einer Risikogruppengröße überhaupt keinen Sinn. Ihre Größe hängt notwendigerweise stark vom untersuchten Kollektiv ab. So wird der Anteil an Niedrigrisikopatientinnen innerhalb der Gruppe der nodal-positiven Patientinnen deutlich kleiner sein, als bei nodal-negativen. Innerhalb der nodal-negativen Patientinnen wird beispielsweise eine Patientinnengruppe mit einer großen Zahl gut differenzierter Tumore (G1) bei einer Analyse mit den Biomarkern naturgemäß eine größere Gruppe mit niedrigem Tumorrisiko ergeben als wenn eine Patientinnenkohorte mit einer repräsentativeren Verteilung der Differenzierung untersucht wird. Aus diesem Grund sollten eher standardisierte international anerkannte (Clinical Laboratory Standards Institute [CLSI], Global Harmonization Task Force [GHTF], siehe Absatz 4f unten) und von diagnostischen Unternehmen bei der Testentwicklung erfassten Leistungsparameter wie diagnostische Sensitivität und diagnostische Spezifität beurteilt und mit der derzeitigen klinischen Praxis verglichen werden, als eine willkürlich gewählte Größe der Risikogruppe, die nicht notwendigerweise primär von der Leistungsfähigkeit des Tests abhängt.

e) Bewertung des Verzerrungspotenzials von Prognosestudien (A.2.1.6.2, S. 31 des Vorberichts) und Gegenüberstellung der Ergebnisse der Einzelstudien (A.2.1.7.1)

Das IQWiG referenziert bei der Bewertung des Verzerrungspotenzials von Prognose-Studien auf drei Publikationen (Wolf et al. 2010, Hayden et al. 2013, Altman et al. 2001) und lässt Studien in der Regel bei der Nutzenbewertung unberücksichtigt, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen basieren. Die Zahl von 70% wird nicht begründet. Sie findet sich in keiner der drei o.g. Publikationen. Bei der Arbeit von Wolf et al. (2010) handelt es sich im Übrigen um keine Vollpublikation, sondern um einen Abstract, aus dem außer der Ankündigung, dass Ergebnisse auf einem Kongress präsentiert werden sollen, keine Informationen hervorgehen. Somit ist diese Arbeit als Referenz ungeeignet. In der im Vorbericht genannten weiteren Publikation von Schulz&Grimes (2002), mit der eine Grenze von 80% begründet wird, ist für prognostische Biomarker-Studien nicht anwendbar, da sich die Arbeit auf prospektive randomisierte Therapiestudien bezieht. Als weitere Referenz wird im Vorbericht die Arbeit von Simon et al. (2009) angegeben. Leider wird diese Referenz unvollständig zitiert. Denn in dieser Arbeit werden zwar 66% als Richtschnur genannt, es wird jedoch auch gesagt, dass eine Minimalanforderung nicht universell anwendbar sei und dass

man alternativ zu den 66% durch andere Methoden einen Selektions-Bias vermeiden könne. Eine Möglichkeit hierfür ist zu zeigen, dass die klinischen Basischarakteristika der mit dem Biomarker untersuchten Patientinnen den Charakteristika der Gesamtstudie entsprechen. Dies konnten wir für unsere beiden klinischen Validierungsstudien für EndoPredict ABCSG-6 und ABCSG-8 zeigen und publizieren ([2], Supplementary table S7).

Im Übrigen ist es bei Prognosestudien für Biomarker entscheidend, dass die untersuchte Kohorte der Zielgruppe für den biomarkerbasierten Test entspricht und nicht, ob sie den Charakteristika der Patientinnen in der gesamten klinischen Therapie-Studie entspricht, die für ein ganz anderes Studienziel „designed“ wurde. Dies wird durch die im Vorbericht zitierte Arbeit von Hayden et al. unterstützt. Dort wird als Kriterium für eine nichtverzerrte Prognosestudie genannt, dass die Studienproben adäquat die „population of interest“ repräsentieren.

Da bei beiden publizierten klinischen Validierungsstudien von EndoPredict (ABCSG-6, ABCSG-8) (1) die untersuchten Populationen in ihren klinischen Charakteristika den Gesamtstudien entsprechen, (2) die untersuchten Populationen der Zielpopulation entsprechen und (3) beide Studien ein konsistentes Ergebnis lieferten, das inzwischen durch eine weitere Studie (ATAC) bestätigt wurde, ist das Risiko für eine Verzerrung bei den EndoPredict Validierungsstudien sehr gering. Daher sollten diese drei Studien für die Nutzenbewertung des EndoPredict Tests berücksichtigt werden.

f) Analytische Leistungsfähigkeit

Im Vorbericht wird ein wesentlicher Aspekt der Leistungsbewertung, Qualitätssicherung und Patientensicherheit für ein *in vitro* Diagnostikum (IVD) nicht berücksichtigt, nämlich dessen analytische Leistungsfähigkeit und Konformität mit dem Medizinproduktegesetz / Anhang I der IVDD 98/79/EG.

Zum Nachweis der Erfüllung der Grundlegenden Anforderungen von Anhang I der IVDD 98/79/EG für die CE-Kennzeichnung von EndoPredict orientierten wir uns bei der Erhebung der Leistungsdaten an den Empfehlungen des Clinical Laboratory Standards Institute (CLSI), das ein internationaler Standard für IVD Hersteller darstellt. Für die Leistungsbewertung zur CE-Kennzeichnung des EndoPredict Tests orientierten wir uns an dem definierten Verfahren der „Global Harmonization Task Force“ (GHTF; inzwischen ersetzt durch IMDRF [International Medical Device Regulators Forum]) zur Sammlung von Daten und Erstellung eines klinischen Leistungsbewertungsberichts von *in-vitro*-Diagnostika (Clinical Evidence for

IVD medical devices – Key Definition and Concepts sowie Scientific Validity and Performance Evaluation [16,17]). Die GHTF/IMDRF hat das Ziel, die regulatorischen Anforderungen, eingereichten Datenformate, Konformitätsbewertungen und Qualitätssysteme zu harmonisieren, um Herstellern und Behörden die Beurteilung der Leistungsfähigkeit eines Medizinprodukts zu erleichtern. Dieses Verfahren entspricht den Anforderungen des MPG (Medizinproduktegesetz) und berücksichtigt die Grundlegenden Anforderungen des Anhang I der IVDD 98/79/EG. Elementarer Bestandteil für eine Leistungsbewertung ist nach GHTF/IMDRF-Empfehlung neben der wissenschaftlichen Validität, der klinischen Leistungsdaten und der klinischen Nützlichkeit auch die analytische Leistungsfähigkeit eines IVD. Umso verwunderlicher ist, dass bei der Nutzenbewertung der biomarkerbasierten Testverfahren analytische Leistungsdaten überhaupt nicht berücksichtigt werden, zumal dies für eine sichere und reproduzierbare Diagnostik absolut notwendig ist. Um reproduzierbar korrekte diagnostische Ergebnisse und damit die Sicherheit für Patientinnen zu gewährleisten, sollte aus unserer Sicht bei der Nutzenbewertung auch die Qualität des Medizinprodukts basierend auf analytischen Leistungsdaten und der Konformität mit dem Medizinproduktegesetz berücksichtigt werden. Entsprechende Daten aus der Technischen Dokumentation des CE-markierten IVD EndoPredict können dem IQWiG zur Verfügung gestellt werden.

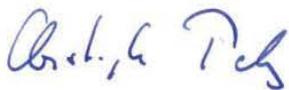
Zusammenfassung

Zusammenfassend ist aus unserer Sicht die Überarbeitung folgender Aspekte im Vorbericht notwendig:

- Bewertung des primären Studienziels der GEICAM/9906-Studie mit EndoPredict
- Berücksichtigung der drei prospektiv-retrospektiven Prognosestudien für EndoPredict (ABCSG6, ABCSG8, ATAC)
- Verlängerung der Nachbeobachtungszeit bei RCTs auf 10 Jahre basierend auf einer Kaplan-Meier-Schätzung mit Konfidenzintervall, um eine Langzeitsicherheit für Patientinnen zu gewährleisten
- Definition der Mindestdauer einer Prognosestudie nicht über die mediane Dauer der Nachbeobachtung, sondern über die Kaplan-Meier-Schätzung mit Konfidenzintervall nach 10 Jahren
- Erhöhung des akzeptablen Rezidivrisikos in der Niedrigrisikogruppe von 5% auf den medizinisch und gesundheitsökonomisch sinnvollen Wert von 10%.

- Bewertung des Verzerrungspotenzials von Prognosestudien nicht über einen willkürlich festgelegten Anteil an untersuchten Patientinnen, sondern über von Simon et al., 2009 genannte valide Möglichkeiten
- Bewertung der analytischen Leistungsfähigkeit der biomarkerbasierten Tests entsprechend internationaler Standards

Mit freundlichen Grüßen



Dr. Christoph Petry
Geschäftsführer



Priv.-Doz. Dr. Ralf Kronenwett
Leiter Forschung und Entwicklung

Anlagen:

- Referenzliste (Anlage 1)
- Kopien der Referenzen, die nicht im Vorbericht zitiert sind (Anlage 2; nur elektronisch per Email)
- Formblatt zur schriftlichen Stellungnahme zum Vorbericht mit Unterschriften von Dr. C. Petry und Priv.-Doz. Dr. R. Kronenwett
- Erklärung von potenziellen Interessenskonflikten von Dr. C. Petry
- Erklärung von potenziellen Interessenskonflikten von Priv.-Doz. Dr. R. Kronenwett