

Mündliche Anhörung



gemäß 5. Kapitel § 19 Abs. 2 Verfahrensordnung
des Gemeinsamen Bundesausschusses

hier: Wirkstoff Afatinib

Sitzung im Hause des Gemeinsamen Bundesausschusses in Berlin
am 25. März 2014
von 11.32 Uhr bis 13.04 Uhr

– Stenografisches Wortprotokoll –

Angemeldete Teilnehmer der Firma **Boehringer Ingelheim Pharma GmbH & Co. KG:**

Frau Dr. Drechsler
Herr Dr. Pfannkuche
Herr Dr. Voß
Herr Dr. Lechner

Angemeldete Teilnehmerin der Firma **Medac Gesellschaft für klinische Spezialpräparate mbH:**

Frau Reim

Angemeldete Teilnehmer der Firma **Pierre Fabre Pharma GmbH:**

Frau Pfeiffer
Herr Dr. Neckermann

Angemeldete Teilnehmerinnen der Firma **Bristol-Myers Squibb GmbH & Co. KGaA:**

Frau Plesnila-Frank
Frau Dr. Mala

Angemeldete Teilnehmer der Firma **Lilly Deutschland GmbH:**

Herr Dr. Brück
Herr Borchardt-Wagner

Angemeldete Teilnehmer der Firma **Roche Pharma AG:**

Frau Dr. Roske
Herr Dr. Flückiger

Angemeldeter Teilnehmer für die **Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie e. V. (DGHO):**

Herr Prof. Dr. Wörmann

Angemeldeter Teilnehmer für die **Arbeitsgemeinschaft Internistische Onkologie (AIO):**

Herr Dr. Sebastian

Angemeldete Teilnehmer für den **Verband Forschender Arzneimittelhersteller e. V. (vfa):**

Herr Dr. Dintsios
Herr Dr. Rasch

Beginn der Anhörung: 11.32 Uhr

(Die angemeldeten Teilnehmer betreten den Raum)

Herr Hecken (Vorsitzender): Guten Morgen! Herzlich willkommen zu unserer Anhörung im schriftlichen Stellungnahmeverfahren der frühen Nutzenbewertung von Arzneimitteln. Ausgangspunkt für unsere heutige mündliche Anhörung ist ein Dossierbericht des IQWiG vom 13. Februar 2014 mit einem sehr differenziertem Beurteilungs- und Bewertungsspektrum. Das reicht von einem Hinweis auf einen „erheblichen Zusatznutzen“ über Anhaltspunkte und Hinweise für „geringen Zusatznutzen“ bis „Zusatznutzen nicht belegt in verschiedenen Patientenpopulationen“. Wir sollten heute über diese Dossierbewertung des IQWiG auf der Basis der von Ihnen vorgelegten schriftlichen Stellungnahme diskutieren.

Mir wären zwei Punkte ganz wichtig, über die wir sprechen sollten. Zum einen ist dies die Fragestellung, ob und in welchem Umfang es hier adäquat ist, Subgruppen in der Art und Weise zu bilden, wie es das IQWiG mit einer gewissen Filigranität getan hat. Dann sollten wir noch einmal über die Altersgrenze 65 sprechen und selbstverständlich über alles, was Ihnen darüber hinaus noch am Herzen liegt.

Ich weise Sie – Sie sind alle Profis – gleichwohl vorsorglich darauf hin, dass wir Wortprotokoll führen. Deshalb habe ich die herzliche Bitte, Ihren Namen, Ihre Firma oder Ihre Institution zu benennen, wenn Sie das Wort ergreifen.

Ich begrüße zu unserer heutigen Anhörung – das muss ich für das stenografische Protokoll tun – zunächst Frau Dr. Drechsler, Herrn Dr. Pfannkuche, Herrn Dr. Voß und Herrn Dr. Lechner von Boehringer, Frau Reim von Medac, Frau Pfeiffer und Herrn Dr. Neckermann von Pierre Fabre Pharma GmbH, Frau Plesnila-Frank und Frau Dr. Mala von Bristol-Myers Squibb, Herrn Dr. Brück und Herrn Borchardt-Wagner von Lilly Deutschland, Frau Dr. Roske und Herrn Dr. Flückiger von Roche, Herrn Prof. Wörmann von der Deutschen Gesellschaft für Hämatologie und Medizinische Onkologie – bekannt aus vielen Anhörungen –, Herrn Sebastian von der Arbeitsgemeinschaft Internistische Onkologie sowie Herrn Dr. Dintsios und Herrn Dr. Rasch – die waren eben auch schon da – für den vfa.

Mein Vorschlag wäre es, dass Boehringer jetzt kurz einleitend vorträgt. Wie gesagt, die Stellungnahme haben wir gelesen. Sie können die aus Ihrer Sicht wichtigen Punkte adressieren, damit wir dann auf dieser Basis diskutieren können. – Wer möchte es machen? – Herr Pfannkuche, bitte schön.

Herr Dr. Pfannkuche (Boehringer): Sehr geehrte Damen und Herren! Sehr geehrter Herr Vorsitzender! Vielen Dank für die einführenden Worte und die, wie immer bei Ihnen in Ihrem Haus, nette Begrüßung. Bevor ich anfangen möchte, möchte ich kurz meine Kolleginnen und Kollegen vorstellen, die mich heute begleiten. Frau Drechsler ist bei uns im Marktzugang, Gesundheitsökonomie verantwortlich für den Bereich Onkologie und damit auch für das Frühbewertungsdossier, Herr Lechner ist bei uns auf globaler Ebene im Bereich Medizin verantwortlich für Afatinib, daneben haben wir noch Herrn Voß, er ist der verantwortliche Projektstatistiker für das Afatinib-Frühbewertungsdossier.

Was können wir sagen? Boehringer Ingelheim hat vor mehr als zwanzig Jahren angefangen, seine Forschung im Bereich Onkologie zu starten. Wir freuen uns, dass wir heute den ersten Wirkstoff dieser Pipeline, Afatinib mit dem Handelsnamen Giotrif, hier diskutieren und gemeinsam näher beleuchten können, was die wissenschaftliche Datenlage anbelangt. Der Wirkstoff ist mittlerweile von der EMA und der FDA zugelassen worden. Er wird eingesetzt zur Monotherapie bei nicht kleinzelligem Lungenkarzinom, wenn vorher keine EGFR–Thyrosinkinase-Inhibitor-Therapie stattgefunden hat, bei

Patienten mit einem metastasierten oder fortgeschrittenen Lungenkarzinom mit einer aktivierenden Mutation des EGFR-Rezeptors.

Wir sehen den Zusatznutzen auf verschiedenen Ebenen. Zum einen sagen wir, wir haben mit der LUX-Lung-3-Studie sehr positive Ergebnisse gezeigt, was das Gesamtüberleben für die Patienten betrifft, die eine Del-19-Mutation haben. Dort konnten wir einen Überlebensvorteil von über zehn Monaten zeigen, was eine bislang noch nicht dagewesene Verbesserung der Lebensqualität in dieser Therapielinie ist. Weder eine Chemotherapie noch eine Thyrosinkinase-Inhibitor-Therapie konnte dies bislang zeigen. Daneben haben wir auch bei der Gesamtpopulation zeigen können, dass das Fortschreiten der Erkrankung um über vier Monate verlängert werden kann bzw. bei Patienten mit häufigen Mutationen um über sieben Monate. Wir konnten auch zeigen – ich denke, das ist hier besonders –, dass es auch zu einer Symptomverbesserung kommt, das sowohl hinsichtlich der Zeit bis zur Verschlechterung als auch der Zeit bis zur Verbesserung von Symptomen, und gleichzeitig können wir im Hinblick auf die Lebensqualität zeigen, dass es zu einer Verbesserung der körperlichen Funktion kommt. Das heißt zum Beispiel schwere Einkaufstaschen tragen, längere Gehwege hinter sich bringen, da kommt es zu Verbesserungen. Gleichzeitig können wir auch zeigen, dass es bei den Thyrosinkinase-Inhibitor-spezifischen Nebenwirkungen zu einer Ausprägung kommt, die handhabbar ist. Sprich: Diarrhö als Beispiel kann therapiert werden und hat keinen Einfluss auf die Lebensqualität an dieser Stelle.

Wenn wir uns die Arzneimittel-Nutzenbewertungsverordnung ansehen, § 3 Abs. 2 – ich zitiere jetzt auch nicht weiter; ich denke, sie ist allen im Raum bekannt –, können wir sagen, dass wir alle die Kriterien, die dort genannt sind, mit dem Präparat Giotrif, Afatinib, erfüllen. Was wir an dieser Stelle gern festhalten wollen, ist, dass wir insgesamt den Zusatznutzen für alle im Dossier dargelegten Patientengruppen sehen. Hier haben wir einen kleinen Unterschied zum IQWiG, über dessen Bericht wir uns im ersten Augenblick gefreut haben, das den Zusatznutzen genau wie wir in wesentlichen Aspekten gesehen hat. In ein paar Punkten haben wir aber Unterschiede, und die wollen wir heute mit Ihnen gerne diskutieren. Im Wesentlichen sind es aus unserer Sicht vier zentrale Punkte.

Zum einen wäre das die Frage: Welchen Stellenwert hat die zweite Studie, die LUX-Lung-6-Studie, bei der Bewertung des Zusatznutzens? Die zweite Frage wäre: Welchen Stellenwert haben die Symptome bei der Bewertung des Zusatznutzens? Beides wären Punkte, auf die Herr Lechner, wenn Sie erlauben, gleich näher eingehen würde. Der dritte Punkt, der für uns sehr wichtig wäre, wäre die Frage, die auch Sie eben schon angesprochen haben, Herr Vorsitzender, die Frage des Alterssplits. Dazu wird Herr Voß noch etwas sagen. Der letzte Punkt, der uns sehr am Herzen liegt, ist das Thema: Wie kann eine Saldierung des Zusatznutzens stattfinden, sprich: Wie kann man die Endpunkte gewichten? Wir haben im Hinblick darauf gearbeitet und eine Patientenpräferenzstudie durchgeführt, wozu wir auch noch mehr sagen können. Das wird Frau Drechsler übernehmen.

Ich wäre an dieser Stelle mit der kurzen Einführung fertig, und wenn Sie erlauben, würde ich gerne an Herrn Lechner übergeben, der zu den ersten beiden Punkten noch mehr sagen kann.

Herr Hecken (Vorsitzender): Herzlichen Dank. – Bitte schön, Herr Lechner.

Herr Dr. Lechner (Boehringer): Ich möchte zusätzlich zu der Einführung meines Kollegen ausführen, warum wir denken, dass die LUX-Lung-6-Studie einbezogen werden sollte. Wir bitten Sie auch, diese zu berücksichtigen. Wir finden, dass einfach in einer solchen frühen Phase der Bewertung eines Arzneimittels nach der Markteinführung jede verfügbare Evidenz herangezogen werden sollte, um die Aussagesicherheit, die Effektmodifikation, aber auch die Nebenwirkungen zu beurteilen. Wir

haben zwischen diesen beiden Studien auch eine sehr hohe Konsistenz in den Effekten gesehen und bitten Sie darum, diese Studie zu berücksichtigen.

Ein wichtiges Anliegen sind uns auch die krankheitsbedingten Symptome. Da wir wissen, dass in dieser Palliativsituation die Vermeidung, Reduktion und Verzögerung dieser krankheitsbedingten Symptome – wir sprechen hier hauptsächlich von Husten, von Atemnot, sogenannter Dyspnoe, und Schmerzen in der Brust – für diese Patienten sehr belastend sind und natürlich auch auf die allgemeine Lebensqualität dieser Patienten sehr starken Einfluss haben, glauben wir, dass eine Verbesserung und Verzögerung dieser Symptome wichtig ist.

Bevor ich, wenn ich darf, an meinen Kollegen Herrn Voß übergebe, um über die Alterssubgruppen zu reden, möchte ich sagen: Aus medizinischer Sicht sind wir primär der Meinung, dass mehr der Allgemeinzustand und der ECOG-Performance-Standard der Patienten zählt, nicht das biologische Alter.

Herr Hecken, wenn Sie erlauben, würde ich an meinen Kollegen Herrn Voß übergeben.

Herr Hecken (Vorsitzender): Bitte schön, Herr Voß.

Herr Dr. Voß (Boehringer): Wie Sie in der Bewertung des IQWiG wahrscheinlich gelesen haben, sieht das IQWiG bei einzelnen Endpunkten – das sind hier Rollenfunktion, körperliche Funktion und Husten – eine bessere Wirksamkeit bei jüngeren Patienten. Die Frage ist nun: Ist das ein realer Effekt, oder ist das eher ein statistisches Zufallsergebnis? Wir haben diese Bedenken ernst genommen und es uns genauer angeschaut und auch Methoden verwendet, die die EMA in ihrer Guideline für Subgruppenanalysen empfiehlt.

Zuerst haben wir in der LUX-Lung-3-Studie den primären Endpunkt PFS und auch das Gesamtüberleben angeschaut. Hier sieht man keinerlei Effekte des Alters auf diese Endpunkte. Genauso sieht man bei pharmakokinetischen Parametern keinen Effekt des Alters. Aus dieser Sicht sehen wir keine Hinweise, dass Afatinib bei jüngeren und älteren Patienten unterschiedlich wirkt.

Nun haben wir für unser Frühbewertungsdossier für Symptome und Lebensqualität umfangreiche Subgruppenanalysen geliefert. Bei einer solch großen Anzahl von Analysen ist die Wahrscheinlichkeit groß, dass man zufällig Subgruppeneffekte sieht, die aber nicht real sind. Wir haben also die LUX-Lung-6-Studie verwendet, um zu schauen, ob sich diese Effekte dort wiederholen. Wiederholen sich die Effekte dort, ist ein realer Effekt sehr wahrscheinlich, wiederholen sich die Effekte jedoch nicht, muss man wahrscheinlich von einem Zufallsbefund ausgehen. Wenn man sich die Ergebnisse zu den drei Endpunkten körperliche Funktion, Husten und Rollenfunktion in der LUX-Lung-6-Studie anschaut, sieht man keinerlei Effekte des Alters. Hier wirkt Afatinib bei jüngeren und älteren Patienten gleichmäßig gut. Also gibt es hier keine Hinweise auf eine schlechtere Wirksamkeit bei älteren Patienten. Wir haben dann noch eine zweite Analyse durchgeführt. Wir haben den Alterstrennwert von 65 Jahren variiert und bei 60 und 70 Jahren angesetzt, um zu schauen, ob man hier konsistente Effekte des Alters sieht. Auch diese sieht man nicht.

Deshalb sind wir in der Gesamtschau der Ergebnisse der Meinung, dass das Zufallsbefunde sind und Afatinib bei jüngeren und älteren Patienten gleichmäßig gut wirkt.

Vielen Dank. Wenn Sie erlauben, würde ich an meine Kollegin Frau Drechsler weitergeben.

Herr Hecken (Vorsitzender): Ja, gerne. Bitte schön, Frau Dr. Drechsler.

Frau Dr. Drechsler (Boehringer): Wir haben uns bei der Dossiererstellung recht früh gefragt: Welche Endpunkte sind für den Patienten wichtig, und vor allem, wie gewichtet der Patient die Endpunk-

te? Aus diesem Grund haben wir gemeinsam mit Professor Mühlbacher eine Patientenpräferenzstudie bei Patienten mit metastasiertem Lungenkarzinom durchgeführt. Es handelt sich hierbei um ein Discrete-Choice-Experiment. Herr Mühlbacher ist, glaube ich, in Deutschland der Experte für Discrete-Choice-Experimente. Dieses Experiment haben wir nach dem aktuellsten Stand der internationalen Wissenschaft durchgeführt. Die Methode der Discrete Choice wird auch seit Jahren im Gesundheitswesen anerkannt und findet unter anderem auch im Methodenpapier des IQWiG eine Benennung.

Mit 211 Patienten handelt es sich in dieser Indikation bis dato um die größte Patientenpräferenzstudie, die durchgeführt worden ist. Als Ergebnis erhält man eine Rangfolge an Therapieeigenschaften, die für den Patienten wichtig sind. Konkret hatten wir hier an erster Stelle für den Patienten die Lungenkrebssymptome Atemnot, Husten, Schmerz sowie das progressionsfreie Überleben, gefolgt von Übelkeit und Erbrechen, geringere Relevanz schrieben die Patienten dem Hautausschlag, dem Durchfall sowie Müdigkeit und Mattheit zu. Aus unserer Sicht können diese Ergebnisse benutzt werden, um den Zusatznutzen zu saldieren. Das heißt, Eigenschaften, die vom Patienten höher gewichtet werden, könnten auch beim Zusatznutzen höher gewichtet werden, und Eigenschaften, die ein geringeres Gewicht erhalten, wären weniger zu gewichten.

Das heißt in unserem Fall, wenn man sich die Effekte von Afatinib anschaut: Wir haben positive und gleichgerichtete Effekte für die Gesamtpopulation bei Atemnot, Husten, Schmerz sowie PFS. Damit haben wir die Eigenschaften auch in der Studie bewiesen, die dem Patienten am wichtigsten sind.

Weitere positive Effekte haben wir, wie wir schon erläutert haben, auch bei der Lebensqualität, der körperlichen Funktion, bei Müdigkeit, Mattheit, Haarausfall, Übelkeit und Erbrechen und bei den Patienten mit Deletion 19 noch zusätzlich ein Gesamtüberleben von fast elf Monaten.

Herr Hecken (Vorsitzender): Herr Pfannkuche, machen Sie weiter?

Herr Dr. Pfannkuche (Boehringer): Ich denke, wir haben jetzt einen kurzen Durchritt durch die vier Themen, die uns sehr wichtig sind, gegeben. Ich würde an dieser Stelle gern in die Diskussion einsteigen und dann auf Ihre Fragen weiter antworten.

Herr Hecken (Vorsitzender): Herzlichen Dank. – Dann fange ich sofort mit einer Frage an Professor Wörmann, vielleicht auch an Herrn Dr. Sebastian an. Das IQWiG hat die LUX-Lung-6-Studie ausgeschlossen und nicht in die Bewertung einbezogen, denn die Grundaussage des IQWiG war, die Dosierung der zweckmäßigen Vergleichstherapie entspricht nicht der Fachinformation, ist nicht fachinformationsgerecht, und insofern kann hier keine belastbare und verwertbare Evidenz generiert werden. Der pharmazeutische Unternehmer argumentiert dagegen und sagt, das hier eingesetzte Dosierschema entspricht der Versorgungsrealität und ist darüber hinaus eben auch von europäischen Zulassungsbehörden als adäquat adressiert und beschrieben worden. Das ist natürlich ein ganz wichtiger Punkt, den wir in unsere Erwägungen einbeziehen müssen. Herr Wörmann, können Sie mir da aus meiner Not helfen und vielleicht zwei, drei Takte dazu sagen? Dass von der Fachinformation andere Dosierungen vorgegeben sind, ist klar. Aber wie sieht die Versorgungsrealität aus, und wie sieht es im europäischen Kontext aus?

Prof. Dr. Wörmann (DGHO): Ich kann vielleicht zu drei Punkten etwas sagen. Wenn Sie sie vorschlagen, nehme ich die drei Punkte auf. Das erste ist: Cisplatin-Pemetrexed würden wir als Standard in Deutschland sehen und nicht Cisplatin-Gemcitabin.

Der zweite Punkt: Die Dosierung bei Gemcitabin ist in Deutschland 1.250 mg und in der Studie 1.000 mg. Jetzt kann man lange diskutieren, ob das einen großen Unterschied macht. Aber es macht einen Unterschied, was zum Beispiel Nebenwirkungen angeht. Deswegen denken wir nicht, dass es ein perfekter Match ist. Es passt eben nicht genau. Es passt so ein bisschen das, was wir schon öfter diskutiert haben. Wenn man in einer Studie den Vergleichsarm ein bisschen schlechter macht, dann kann dadurch der Verumarm, der neue Arm, gerade die Stufe zur nächsten Signifikanz überschreiten oder hier vielleicht sogar von „beträchtlich“ auf „erheblich“. Da reicht ganz wenig, um das zu tun. Es ist für uns grundsätzlich ein kleines Problem in randomisierten Studien, dass wir denken, der Vergleichsarm muss absolut optimal sein und auf keinen Fall ein Geschmäcke kriegen, dass der Vergleichsarm ein bisschen suboptimal ist. Ich glaube nicht, dass dadurch das Gesamtergebnis beeinträchtigt ist, aber in Nuancen schon. Möglicherweise wäre hier genau erreicht, dass durch die strikte Operationalisierung mit Grenzen gerade die nächste Stufe überschritten worden ist. Dieser Eindruck ist ein bisschen dabei.

Der dritte Punkt ist ein übergeordneter Punkt. Die Vergleichsstudie ist nur in Asien, in drei Ländern, durchgeführt worden, in sicher hochversierten Zentren. Es gibt eine große Diskussion, ob eine pharmakokinetisch identische Übertragung von asiatischen zu europäischen Patienten möglich ist. Vielleicht geht das. Trotzdem sehen wir in der vorherigen Studie – wir haben das in der Stellungnahme aufgeführt – ziemliche Unterschiede zwischen nur in Asien und nur bei uns durchgeführten Studien. Es kann sein, dass das nicht in der Pharmakokinetik der Substanz liegt, sondern daran, dass die molekularen Grundlagen etwas anders sind. Wir sehen aus der Subgruppenanalyse auch, dass die molekularen Untergruppen etwas anders ansprechen. Deswegen glauben wir nicht, dass das perfekt übertragbar ist. Wir finden, es ist eine Bestätigung des Ergebnisses, wir finden aber nicht, dass man das als Vergleichstherapie so eins zu eins übernehmen kann.

Herr Hecken (Vorsitzender): Danke schön. – Herr Pfannkuche dazu.

Herr Dr. Pfannkuche (Böhringer): Ganz kurz zur LUX-Lung-6-Studie noch einen wichtigen Punkt. Wenn man sich die Dosierung anschaut: Ja, sie entspricht nicht genau der Fachinformation. Wir sehen jetzt: Wenn wir uns die Wirksamkeit anschauen, so könnte das eine Verbesserung zugunsten von Afatinib bedeuten, wenn wir uns die Nebenwirkungen anschauen, so könnte es eher zu einer Verzerrung zuungunsten von Afatinib kommen. Deswegen ist unsere Position an dieser Stelle, dass man die LUX-Lung-6-Studie für spezifische Aspekte berücksichtigen sollte. Das Erste wäre zur Beurteilung von Effektmodifikationen, das Zweite wäre zur Beurteilung der Aussagesicherheit, und das Dritte wäre zur Beurteilung der Nebenwirkungen, da eben zwei Studien vorliegen, was zu einem solch frühen Zeitpunkt auch nicht unbedingt immer der Fall ist.

Das zweite Thema, das eben angeklungen ist, ist die Ethnie. Auch dazu würden wir gerne noch etwas sagen. Da kann sowohl der Herr Lechner als auch Herr Voß noch mehr dazu sagen, wie unsere Einschätzung hier ist.

Herr Hecken (Vorsitzender): Bitte, sofort dazu.

Herr Dr. Lechner (Boehringer): Bezüglich der Ethnien: Ausschlaggebend für uns und unseren Standpunkt ist, dass in diesem Krankheitsbild, also in dem EGFR-Mutation-positiven Lungenkrebs, primär die Mutation der Treiber ist, dass hier also die Tumorbiologie eine Rolle spielt. Es wurde angesprochen, dass durchaus eventuelle Unterschiede zwischen den einzelnen Mutationssubgruppen bestehen, dass die Ethnie hier aber eine untergeordnete Rolle spielt. Es stimmt zwar, dass in asiatischen Populationen, also an asiatischen Patienten, eine höhere Prävalenz der EGFR-Mutationen vorliegt.

Wir haben aber keine Hinweise, dass hier Unterschiede in der Tumorbiologie bestehen. Daher glauben wir, dass es durchaus vergleichbar ist. Dies wäre jetzt mein Hauptpunkt zur ethnischen Unterteilung. Auch in unseren Analysen haben wir keine Unterschiede gesehen, als wir uns den Studienausgang in Bezug auf die Ethnie angesehen haben. Mein Kollege kann hierzu vielleicht noch weiter ausführen.

Herr Hecken (Vorsitzender): Danke schön. – Herr Voß, bitte.

Herr Dr. Voß (Boehringer): Wir haben für das Frühbewertungsdossier umfangreiche Daten auch zu der Lebensqualität und den Symptomen geliefert und haben immer auch die Ethnie als Subgruppe analysiert und haben da wirklich eine große Zahl von Analysen gemacht. Da sieht man eigentlich so gut wie keine Hinweise, dass die Ethnie hier eine Rolle spielt. Der Vorteil unserer Studie ist, dass wir hier sowohl Kaukasier als auch Asiaten haben und das eben analysieren können. Wenn man sich die Bewertung des IQWiG ansieht, so stellt man fest, das IQWiG hat auch keine Hinweise bezüglich der Ethnie gesehen.

Herr Hecken (Vorsitzender): Herzlichen Dank. – Herr Dr. Sebastian.

Herr Dr. Sebastian (AIO): Ich wollte auf die Dosierung des Cisplatin-Gemcitabin eingehen. Ich halte das Cisplatin-Gemcitabin für eine durchaus gängige und im klinischen Alltag benutzte Kombination. Wir hatten uns gewundert, dass wegen dieser sogenannten Unterdosierung diese Studie nicht zur Bewertung herangezogen wurde. Es ist richtig, die meisten Studien und auch die Zulassungen sind mit 1.200 und 1.250 mg/m² gemacht worden. Man muss sich die Studien dann aber ansehen. Eine der letzten großen Studien ist die AVAPERL-Studie. Da haben wir Therapieabbrüche wegen Nebenwirkungen in dem Chemotherapiearm von über 20 Prozent. Dabei muss man sich schon fragen, ob er das dann eine adäquate Therapie in der palliativen Therapiesituation ist. Klar, 1.000 mg sind möglicherweise etwas unterdosiert oder am unteren Ende dosiert. Wir kennen es mit anderen Kombinationen auch. Bei Caroplatin-Paclitaxel variiert die Dosis von 175 bis 225 mg. Ich glaube, Ähnliches kann man für diese Dosis annehmen. Es ist im niedrigen, aber noch wirklich wirksamen und üblichen Bereich. Deswegen hätten wir gar kein Problem, diese Studie als „zur Bewertung heranzuziehen“ zu bewerten.

Herr Hecken (Vorsitzender): Danke schön, Herr Sebastian. – Herr Kaiser, bitte.

Herr Dr. Kaiser: Ich möchte zwei Punkte in Zusammenhang bringen. Das sind die LUX-Lung-6-Studie und das Thema Effektmodifikationen. Ich denke, die Zulassung hat schon einen hohen Stellenwert. Ich würde das auch nicht zu niedrig hängen. Die Zulassung hat eine andere Dosierung. Es wäre Ihnen möglich gewesen, eine Studie mit einer adäquaten Dosierung durchzuführen, dann eben nicht in Asien, sondern in Europa. Das haben Sie einfach nicht gemacht. Deswegen muss man sich fragen, ob diese Studie, die eben in Asien mit einer dort gültigen Zulassung durchgeführt worden ist – das mag so sein, aber hier geht es nicht um Patienten in China, sondern um Patienten in Deutschland –, eine angemessene Studie für die Beantwortung der Fragestellung ist.

Wenn Sie sagen, dass Sie bestimmte Dinge in die eine oder andere Richtung verzerrt haben, so muss man sich natürlich fragen, ob man diese Studie überhaupt für Effektmodifikationen gebrauchen kann. Denn sie überdecken möglicherweise Effektmodifikationen dadurch, dass sie Verzerrungen in bestimmten Endpunkten haben. Das heißt, Sie können an dieser Stelle die Studie dann auch nicht verwenden, um von Konsistenz oder Nichtkonsistenz von Effektmodifikationen zu sprechen. Das funktioniert schlicht und einfach nicht. Diese Studie beantwortet einfach erst einmal eine ganz andere

Fragestellung als die andere Studie. Manchmal haben Sie Konsistenz, manchmal haben Sie keine Konsistenz in den Ergebnissen. Sie sagen zum Beispiel, bei Overall Survival haben Sie Konsistenz, das werten Sie dann, bei Effektmodifikationen haben Sie keine Konsistenz, das werten Sie dann nicht. Aber beides kann schlicht und einfach Zufall sein.

Sie können, weil Sie mit dieser Studie eine ganz andere Fragestellung beurteilen, nicht beurteilen, ob das konsistent oder nicht konsistent an dieser Stelle ist.

Zum Thema Effektmodifikation. Herr Hecken, Sie hatten es mit diffiziler Beurteilung eingeleitet. Völlig richtig. Wir haben das, was der Hersteller sowohl in der Studie als auch im Dossier gemacht hat, sehr stark eingedampft, weil das in der Umfänglichkeit, wie es vorher war, sowohl in der Studie angelegt als auch im Dossier, praktisch nicht mehr interpretierbar war. Wir sind auch hier, glaube ich, an der Grenze der Interpretierbarkeit. Das war eine ziemlich große Herausforderung. Wir haben nur Dinge interpretiert, die über mehrere Endpunkte konsistent in eine Richtung eine Effektmodifikation gezeigt haben. Wir haben keinerlei Effektmodifikation in das Gesamtergebnis übernommen, wo wir einmal nur ein Ergebnis in einem Endpunkt hatten. Das heißt, wir haben durchaus Konsistenz, nämlich innerhalb der Studien über Endpunkte. Darüber muss man sich bewusst sein. Das ist erst einmal eine Darstellung der Ergebnisse. Was der G-BA dann in seinem Beschluss macht, ist eine zweite Fragestellung. Aber es ist zunächst einmal der Ausgangspunkt.

An dieser Stelle ist ganz wesentlich und ganz führend natürlich die Effektmodifikation in dem Mutationsstatus. Diese Effektmodifikation findet übrigens nicht nur beim Overall Survival statt, sondern sie findet auch genau in der gleichen Richtung im PFS statt. Das heißt, in Ihrer Stellungnahme sagen Sie als Hersteller: Na ja, wenn man jetzt jedes Mal die Gesamtstudie anschaut, dann haben wir einen guten Overall Survival, und eigentlich muss man sich das insgesamt auch bei der Mutation anschauen und mit dem Overall Survival vorsichtig sein, und eigentlich haben wir einen nicht so quantifizierbaren Zusatznutzen in diesem Bereich. Sie haben auch beim PFS eine starke Effektmodifikation, und auch dort zuungunsten des PFS, der kleinen Gruppe mit den seltenen Mutationen. Das heißt, man hat auch hier konsistente Ergebnisse, die zum Nachteil dieser Gruppe sind. Man muss sich bei der Behandlung der Patienten – darum geht es ja, das ist das Primäre – darüber bewusst sein: Diese Studie zeigt ein schlechteres Ergebnis als die verfügbare Therapie für Patienten mit diesen Mutationen. Man muss sich, glaube ich, sehr bewusst sein, ob man dann sagt: Für diese Patienten hat diese Therapie einen Zusatznutzen.

Herr Hecken (Vorsitzender): Danke schön, Herr Kaiser. – Als nächster Herr Mayer.

Herr Dr. Mayer: Ich habe zwei Fragen. Die erste Frage betrifft die Aussage von Herrn Pfannkuche, dass die LUX-Lung-6-Studie unterstützend herangezogen werden soll, um praktisch Effekte der LUX-Lung-3-Studie zu bestätigen. Gerade zwischen LUX-Lung-3 und LUX-Lung-6 ist im Nebenwirkungsprofil, zum Beispiel bei Diarrhö, ein deutlicher Unterschied herausgekommen, den manche vielleicht auf die Ethnie zurückführen. Ich glaube, wir können es nicht beweisen, weil es zwei verschiedene Studien sind, aber letztlich haben wir bei der LUX-Lung-6-Studie 100 Prozent Asiaten oder Chinesen und in der LUX-Lung-3-Studie eben nur 75 Prozent, aber einen deutlichen Unterschied in der Diarrhö-Häufigkeit. Von daher frage ich mich, ob man wirklich generell Nebenwirkungen vergleichbar machen kann, wenn man schon isoliert in einem typischen Nebenwirkungsprofil der TKI einen solchen Unterschied in den beiden Populationen hat.

Eine zweite Anmerkung bezieht sich auf die Äußerung von Herrn Sebastian. Ich denke, grundsätzlich ist die Dosierung schon relevant für eine Vergleichstherapie. Auch bei Afatinib gibt es eine Anfangsdosierung, und es kann herunterdosiert werden. Es ist bei ganz vielen Chemotherapien bei anderen

Indikationen so, dass man natürlich irgendwo mit einer Standarddosierung anfängt, einige vertragen sie, bekommen die optimale Dosierung, wirksamkeitsbezogen, und dann herunterdosieren muss. Es ist in Fachinformationen immer so festgehalten. Aber Standard sollte sein, die empfohlene Startdosis zu nehmen, auch wenn man dann zügig herunterdosieren muss.

Herr Hecken (Vorsitzender): Herzlichen Dank, Herr Mayer. – Herr Wörmann, bitte schön.

Herr Prof. Dr. Wörmann (DGHO): Ich widerspreche Herrn Sebastian nicht, dass Gemcitabin grundsätzlich eine Alternative ist. Aber in den Leitlinien ist Pemetrexed an erster Stelle festgelegt worden. Insofern gibt es auch einen Grund, dass man nicht Cisplatin und irgendetwas in der in Europa durchgeführten Studie hat und man es sich aussuchen kann. Vielmehr ist Pemetrexed festgesetzt worden.

Ich will nicht zu dogmatisch darin sein, aber es gibt schon einen Grund, diese Dosis erst einmal so zu wählen. Es geht, glaube ich, überhaupt nicht darum, zu sagen, dass Afatinib in der LUX-Lung-Studie, die in Asien durchgeführt wurde, nicht wirkt. Es ist eine schöne Bestätigung, dass ein großer und dramatischer Effekt dabei herauskommt. Aber wenn es jetzt darum geht, um 0,1 oder 0,2 Grenzen zu berechnen, dann macht eine potentiell etwas schlechtere Vergleichstherapie einen potentiell großen Unterschied aus. Es wäre ganz schwierig, das allein als Grundlage zu nehmen. Das trifft, glaube ich, nicht nur diese, sondern auch andere Bewertungsverfahren.

Jetzt beantworte ich eine Frage, die gleich auch Herr Pfannkuche beantworten muss. Es kann sein, dass in anderen Ethnien die Diarrhö anders ausgeprägt ist. Ich glaube eher, dass es wirklich daran liegt, dass in der Zulassungsstudie extrem viele Zentren dabei waren, die nicht ganz so versiert waren, Diarrhö zu managen. Wenn wir denken, dass es ein Substanzklasseneffekt ist, den wir auch von Gefinitib und Erlotinib kennen, dann liegt die hier erreichte Diarrhö-Rate unter Afatinib schon deutlich höher. Ich würde eher denken, das ist eine Sache der mangelnden Erfahrung mit der Substanz. Wenn man es frühzeitig managt, wird eine Diarrhö Grad 3 und 4 nicht mehr erreicht.

Da die asiatische Studie, die LUX-Lung-6-Studie, nur in Asien mit viel weniger Zentren durchgeführt wurde, ist da viel mehr Erfahrung im einzelnen Zentrum mit dieser Nebenwirkung vorhanden gewesen. Ich kann nur spekulieren, dass es so ist. Aber es fällt schon heraus, dass beim Substanzklasseneffekt die Diarrhö in der europäischen Gruppe so hoch gewesen ist; dies ist eigentlich ungewöhnlich. Herr Sebastian darf mich korrigieren. Aber eigentlich ist das nicht das, was Sie aus der Routine erkennen.

Herr Hecken (Vorsitzender): Herr Sebastian direkt dazu.

Herr Dr. Sebastian (AIO): Direkt zur Diarrhö. Das ist sicherlich ein Problem, und da stimme ich Ihnen ganz zu. Aber ich glaube, das ist zentrenspezifisch. Zentren, die schon längere Erfahrung in der Therapie mit den TKIs haben, gerade bei Patienten mit einer aktivierenden Mutation, haben eine stärkere Diarrhö-Intensität. Aber man muss sich auch die Therapieabbruchrate anschauen. Das sind Patienten, die im Mittel über ein Jahr diese Therapie bekommen haben, und das werden Sie mit einer Diarrhö Grad 3 nie durchhalten. Diarrhö Grad 3 sind mehr als sieben Stuhlgänge pro Tag. Die klinische Realität ist so, dass wir ein Flarephänomen haben, dass wir in den ersten Wochen schon eine verstärkte Diarrhö haben. Die kann man relativ schnell kupieren, wenn man erfahren ist, indem man die Patienten richtig therapiert. Aber auch später ist ohne eine entsprechende Therapie die Diarrhö-Häufigkeit sehr gering. Man darf sich das nicht so vorstellen, dass sich die Diarrhö über 13 Monate hinzieht, sondern das ist eine sehr kurze Sache. Das wird in den Daten einfach nicht abgebildet, weil wir nicht wissen, wie lange die Diarrhö Grad 3 angedauert haben. Aus der klinischen Realität ist es so, dass sie relativ kurz ist, also wenige Tage bis Wochen.

Herr Hecken (Vorsitzender): Herzlichen Dank, Herr Sebastian. – Jetzt Herr Dr. Pfannkuche, um die eben an Sie gerichtete Frage zu beantworten. Vielleicht können Sie auch noch etwas zu der Frage sagen, wie Sie die Patientengruppe mit ECOG 2 hinsichtlich des möglichen Zusatznutzens beurteilen, ja oder nein, weil das eine Fragestellung ist, die wir auf alle Fälle noch diskutieren müssen. Bitte schön, Herrn Pfannkuche.

Herr Dr. Pfannkuche (Boehringer): Ich habe mir ein paar Fragen aufgeschrieben, die jetzt offen waren. Von Herrn Kaiser war es die Effektmodifikation. Ich habe herausgehört, dass auch andere Mutationsarten ein Thema wären. Herr Mayer sprach den Unterschied bei der Diarrhö zwischen den beiden Studien an.

Zum Thema Effektmodifikation wird Herr Voß gleich noch etwas sagen. Zu seltenen und anderen Mutationen kann Herr Lechner etwas sagen, genau wie zum Thema Diarrhö. Zu dem Thema, das Sie gerade angesprochen haben, ECOG-Performance-Status 2, können wir sagen, dass wir die LUX-Lung-1-Studie haben. Das ist eine Studie, die wir mit der Stellungnahme noch eingereicht haben. Wir hatten sie im ursprünglichen Dossier nicht drin, weil sie nicht genau dem Label entspricht. Hier wurde keine systematische Testung vorgenommen, und es fand auch eine Vortherapie mit anderen Thyrosinkinase-Inhibitoren statt. Deswegen passt sie nicht genau auf die Situation. Nichtsdestotrotz haben wir in diese Studie insgesamt 585 Patienten eingeschlossen, von denen 45 einen ECOG-Performance-Status von 2 aufwiesen. Wenn wir uns jetzt die Hazard Ratios für PFS getrennt nach ECOG-Performance-Status 0, 1 und 2 anschauen, dann sehen wir, dass alle Patienten vergleichbare Effekte aufweisen. Daraus schließen wir, wie auch die Literatur es nahelegt, dass Patienten mit ECOG-Performance-Status 2 ähnlich reagieren wie Patienten mit ECOG-Performance-Status 0 oder 1.

Herr Dr. Lechner (Boehringer): Ich möchte anfangen mit der leichter zu beantwortenden Frage, das ist die Diarrhö. Herr Wörmann und Herr Sebastian haben das meiner Meinung nach so gut beantwortet, so gut kann ich es gar nicht mehr beantworten. Mit adäquatem Management ist die Diarrhö eine Nebenwirkung, die durchaus handhabbar ist. Ich möchte auch darauf hinweisen, dass in der LUX-Lung-3-Studie die Abbruchrate aufgrund von Diarrhö nur 1 Prozent war, in der LUX-Lung-6-Studie haben wir keine Abbrüche gehabt.

Bezüglich der Unterschiede der Ethnien ist für mich auch Folgendes interessant. Es gibt Hinweise, die sind vom Body-Mass-Index kleiner, die würden dann wahrscheinlich stärkere Nebenwirkungen haben als eine robuste kaukasische Population. Das ist eben nicht der Fall. Auch das ist für mich ein Argument, das man anführen könnte. Es gibt hier keine ethnischen Unterschiede.

Auf das Diarrhö-Management möchte ich nicht weiter eingehen. Wir sind der Meinung, mit adäquatem Management ist es in der klinischen Praxis durchaus zu handhaben, und das wurde von den Praktikern, die hier sitzen, bestätigt.

Bezüglich der Uncommon Mutations, also der seltenen Mutationen: Das ist ein guter Punkt. In beiden Studien waren zum ersten Mal diese seltenen Mutationen, die nur 10 Prozent in der Gesamtbevölkerung ausmachen, einbezogen. Das hat natürlich dazu geführt, dass das eine sehr heterogene und kleine Gruppe ist, wo innerhalb von einzelnen Mutationen auch sehr große Abweichungen bezüglich der Response Rates und auch des progressionsfreien Überlebens zu beobachten waren.

Die Aussagekraft ist hier natürlich beschränkt, aber man kann durchaus sagen, dass in bestimmten dieser seltenen Mutationen Afatinib einen Vorteil, eine Effektivität oder eine Wirkung zeigt. Wir haben dahin gehend auch eine Analyse unternommen, uns nur die seltenen Mutationen aus der LUX-Lung-3-, LUX-Lung-6- und LUX-Lung-2-Studie anzusehen – das war natürlich eine rein exploratorische Analyse –, ohne den Vergleichsarm hinzuzuziehen. Wir haben gesehen, dass durchaus in Clustern

der seltenen Mutationen ein Effekt besteht. In manchen ist der Effekt natürlich nicht sehr ausgeprägt. Insofern ist – ich lasse Herrn Voß auf die Statistik eingehen – die Aussagekraft natürlich begrenzt.

Abschließend zum ECOG 2. Mein Kollege Herr Pfannkuche hat das, glaube ich, sehr eloquent schon beantwortet. Ich bitte auch zu berücksichtigen, dass in unsere Studie LUX-Lung 3 ECOG-2-Patienten nicht eingeschlossen wurden, da wir ermöglichen wollten, dass die den vollen Benefit auch vom Vergleichsarm bekommen. Die Chemodoublette wäre für ECOG-2-Patienten unter Umständen nicht tolerabel gewesen. – Damit würde ich, wenn ich darf, an meinen Kollegen Herrn Voß übergeben.

Herr Hecken (Vorsitzender): Herr Voß, bitte.

Herr Dr. Voß (Boehringer): Es ist die Frage, ob man die LUX-Lung-6-Studie für die Effektmodifikation verwenden kann. Herr Wörmann hat gesagt, bei Unterdosierung ist der Effekt ein bisschen kleiner, ein bisschen größer, aber für die Effektmodifikation sind die Effekte in den Subgruppen homogen. Da kommt es darauf, dass der Effekt etwas größer oder etwas kleiner ist, weniger an. Wir schauen, ob die Effekte homogen sind. Das sind sie zum Beispiel bezüglich Alter oder eben auch bezüglich des Mutationsstatus in der LUX-Lung-6-Studie.

Bezüglich des PFS – ich weiß nicht, ob Herr Kaiser auf die L858R-Patienten angespielt hat –, da hatten wir Baseline-Imbalancen bezüglich prognostischer Faktoren. Wir haben auch eine Analyse dafür gemacht und dafür adjustiert. So etwas kann bei der Randomisierung einfach vorkommen. Dann sieht man, dass dieser Unterschied bei PFS doch bedeutend geringer ist.

Herr Hecken (Vorsitzender): Danke schön. – Jetzt habe ich direkt dazu Herrn Kaiser.

Herr Dr. Kaiser: Ich bleibe einfach bei den seltenen Mutationen. Die Effektmodifikation ist lange diskutiert worden.

Sie haben eine stratifizierte Randomisierung nach den Mutationen vorgenommen. Sie haben auch danach ausgewertet. Das ist ihre primäre Analyse. Da zeigt sich ein negativer Effekt für die seltenen Mutationen. Sie haben das durchaus schon als offensichtliche Hypothese gehabt, zu überlegen, wie das Ergebnis bei den seltenen Mutationen ist. Sie haben jetzt geschrieben, dass das, wenn Sie bestimmte Adjustierungen vornehmen, nicht mehr ganz so stark ausgeprägt ist. „Der negative Effekt nicht mehr ganz so stark ausgeprägt“, ist sehr zurückhaltend formuliert. Sie propagieren aber einen Zusatznutzen für diese Gruppe. Man muss sich fragen, auf welcher Basis der Daten Sie einen Zusatznutzen propagieren.

Herr Lechner, wenn Sie sagen, sie haben sogar innerhalb dieser Gruppe eine Unterscheidung zwischen verschiedenen Clustern gemacht und sehen bei manchen dieser Cluster bei diesen sehr wenigen Patienten sogar einen positiven Trend – oder wie auch immer es ist – zugunsten von Afatinib, bedeutet das im Umkehrschluss, Sie haben einen noch deletäreren Effekt für die anderen Patienten. Dann ist der Effekt eben nicht so stark ausgeprägt. Wir befinden uns hier in einer anderen Situation. Wir befinden uns nicht in der Situation, dass bei diesen Patienten der Effekt von Afatinib weniger stark oder gar nicht vorhanden ist, sondern er ist negativ für die Patienten. Wenn Sie von einem Teil dieser Patienten sagen, ein paar profitieren aber, bedeutet das in der Gesamtschau doch, dass es für die anderen Patienten noch schlechter ist. Ich bin verwundert darüber, wie mit diesen Patienten umgegangen wird in der Darstellung und den Ergebnissen zu diesen Patienten und der Behauptung, für diese Patienten gibt es einen Vorteil bei Einnahme dieses Medikaments.

Herr Hecken (Vorsitzender): Herr Voß direkt dazu.

Herr Dr. Voß (Boehringer): Ich kann kurz etwas dazu sagen. Meine Aussage vorhin hat sich auf die L858R-Patienten bezogen und nicht auf die seltenen Mutationen. Zu den seltenen Mutationen hat Herr Lechner schon ausgeführt, dass das eine sehr heterogene Gruppe ist. Del 19 und L858R sind homogene Gruppen, und bei den seltenen Mutationen werden ganz viele unterschiedliche Sachen zusammengefasst. Deshalb ist da eine allgemeine Aussage schwierig. Ich weiß nicht, ob Herr Lechner noch etwas hinzufügen will.

(Zuruf: 10 Prozent!)

Herr Hecken (Vorsitzender): 10 Prozent. – Herr Wörmann dazu.

Herr Prof. Dr. Wörmann (DGHO): Zu den Subgruppen. Ich mache jetzt leider einen kurzen Exkurs in die akuten Leukämien. Wir waren vorher der Überzeugung, dass akute myeloische Leukämien bei älteren Patienten nicht behandelbar und bei jüngeren mit hohen Heilungsraten gut behandelbar waren. Dann haben wir gelernt, dass altersabhängig die Biologie bei diesen Patienten völlig anders ist. Inzwischen wissen wir dasselbe vom Mammakarzinom. Ältere Mammakarzinompatienten haben biologisch ein anderes Mammakarzinom als jüngere Patienten. Was wir schwierig für Leitlinien finden, ist, aus Studien, bei denen nicht primär diese Subgruppen als Frage angelegt worden sind, dann Ableitungen zu finden. Wenn wir Subgruppen aus großen Studien machen, würden wir in einer Leitlinie nie eine starke Empfehlung aus einer retrospektiven Subgruppenanalyse machen, selbst wenn vorher stratifiziert worden ist, weil die Power der Studie dafür nicht angelegt worden ist. Das heißt, wir würden dann – das haben wir bei den Leukämien in Deutschland exemplarisch durchgeführt – nur mit älteren Patienten, mit einer bestimmten genetischen Veränderung, eine Studie machen und zeigen, ob man die beeinflussen kann oder nicht.

Wir finden schwierig – das ist eine Grundproblematik des IQWiG –, dass wir es in Leitlinien ablehnen, Subgruppenanalysen für Empfehlungen oder starke Empfehlungen zu nutzen, aber in der Nutzenbewertung anders herangegangen wird. Wir würden sagen, Subgruppen sind wichtig, aber eigentlich nur als Hypothese für die nächste Studie. Ich glaube, an diesen Subgruppen ist etwas dran, aber es wird eine Interaktion zum Beispiel mit Alter und anderen Faktoren geben. Da muss man eine ausreichend gepowerte Studie anlegen. Ich glaube, dass wir das brauchen, gerade auch beim Lungenkarzinom. Es läuft ja auch schon so. Wir wissen, das läuft unter anderem auch in Deutschland. Aber zum jetzigen Zeitpunkt nachträglich würden wir aufgrund einer solchen Analyse in den Leitlinien keine anderen Empfehlungen als Schlüsse ziehen, auch um nicht Personen eine Therapie vorzuenthalten, die möglicherweise davon profitieren. Wir finden es schwierig, eine unterschiedliche Methodik jetzt als Basis eigentlich evidenzbasierter Medizin zu verwenden.

Herr Hecken (Vorsitzender): Herzlichen Dank, Herr Wörmann. – Herr Sebastian dazu.

Herr Dr. Sebastian (AIO): Wir müssen – das ist mir eine Herzensangelegenheit – bei den mutierten Patienten zwei Gruppen unterscheiden. Das sind die Patienten mit einer aktivierenden Mutation, dazu gehören auf jeden Fall die Deletion 19 und die Punktmutation 21, die in zwei Gruppen untersucht wird. Es ist ganz interessant, wie das gemacht wurde. Das ist allerdings – da schließe ich mich Ihnen an – in keiner Leitlinie und in keiner klinischen Realität bislang Wahrheit geworden. Die große Problematik sind die 10 Prozent, bei denen wir sagen, das sind Patienten mit seltenen Mutationen, und die können wir nicht einfach so über einen Kamm scheren, wie es gemacht wird. Da sind nämlich Patienten darunter – da gebe ich Ihnen Recht –, die haben ein fantastisches Ansprechen. Das sind Punktmutationen im Exon 19, aber es sind dann auch Patienten darin, die, wie wir gesehen haben, überhaupt kein Ansprechen haben, weil es keine aktivierende Mutation ist. Das ist möglicherweise ein In-

nocent Bystander, ein zufälliger Befund; diese Tumoren sind gar nicht von der Mutation abhängig. Oder es sind, wie die T790M oder andere, Mutationen, die einfach eine wesentlich höhere Konzentration des Wirkstoffs brauchen, bei denen wir mit den bisherigen Thyrosinkinase-Inhibitoren keinen Fuß in die Tür bekommen. Diese kleine Gruppe von 10 Prozent der Patienten müssen wir von der klinischen Sicht eigentlich wesentlich differenzierter betrachten, weil es da Patienten gibt, die in der Tat keinen Nutzen haben, aber es gibt auch Patienten, die sicherlich einen erheblichen Nutzen haben. Ich habe da ein Problem, dass es nicht der klinischen Realität entspricht, wie wir die jetzt auseinanderdividieren.

Herr Hecken (Vorsitzender): Danke schön. – Frau Teupen.

Frau Teupen: Es passt jetzt vielleicht nicht ganz so. Ich wollte etwas zu Patientenpräferenzen fragen. Sie hatten es gerade ausgeführt, und es ist aus unserer Sicht etwas euphemistisch, dass das als akzeptiertes Instrument schon etabliert wäre. Uns ist wichtig, was für Patienten befragt wurden. Ich wollte fragen, ob die sich an der Studie beteiligt haben, auch im asiatischen Raum. Waren das Patienten aus den Studien selber, waren das andere Studien? Die Unterlagen hierzu liegen uns nicht vor.

Ich habe auch noch die Frage, wo Sie einen Unterschied zu den Morbiditäts- und den Lebensqualitätsfragen sehen. Der EQ-5D wurde vom IQWiG nicht ausgewertet. Im QLQ-C30 und QLQ-LC13 sind mit Sicherheit auch die relevanten Fragen abgebildet worden.

Wir hätten eine Frage noch an Sie. Es sind viele Therapieabbrüche zum ersten Datenschnitt, von 70 bis ungefähr 45 Prozent. Vielleicht könnten Sie darauf noch kurz eingehen.

Herr Hecken (Vorsitzender): Wer macht das? – Bitte schön, Frau Drechsler.

Frau Dr. Drechsler (Boehringer): Frau Teupen, ich möchte als Erstes auf die Patientenpräferenzstudie eingehen. Ihre Frage war, an welchen Patienten sie durchgeführt worden ist. Die Patienten waren alles deutsche Patienten, das heißt, wir haben bei den 211 Patienten ein deutsches Patientenkollektiv. Den Abschlussbericht haben wir im Dossier angehängt. Da können Sie, wenn es Sie interessiert, weiter nachlesen.

Ich möchte auf Ihre zweite Frage eingehen, wo der Unterschied zwischen den Symptom- und Lebensqualitätserhebungen bei LC13 und C30 ist. Es ist eine andere Methode. Bei den Symptom- und Lebensqualitätsdaten nach LC13 und C30 geht es darum, zu erfassen, wie die Symptome bei dem Patienten sich verbessern oder verschlechtern. Bei der Patientenpräferenzstudie hat man eine andere Fragestellung. Im ersten Schritt identifiziert man mit einer systematischen Literaturrecherche mögliche patientenrelevante Endpunkte, dann wird ein Fragebogen entworfen, und in qualitativen Prätests, das heißt Eins-zu-eins-Interviews mit Patienten, wird geschaut, ob das die relevanten Eigenschaften sind, wie verständlich der Fragebogen ist. Bei der Durchführung haben wir dann verschiedene Alternativen. Der Patient bewertet immer Therapie A versus Therapie B. Bei Therapie A sind verschiedene Eigenschaften enthalten. In unserer Studie war das PFS, die Nebenwirkungen, die Symptome. In Therapie A war zum Beispiel ein geringeres PFS, dafür eine höhere Symptombelastung durch beispielsweise Diarrhö, verglichen mit einer zweiten Alternative, Therapie B, wo wir ein höheres PFS hatten, dafür eine stärkere oder geringere Symptombelastung. Das heißt, bei der Patientenpräferenzstudie gibt es immer einen Trade-off zwischen verschiedenen Therapieeigenschaften.

Herr Hecken (Vorsitzender): Eine Nachfrage von Frau Teupen.

Frau Teupen: Vielleicht können Sie sagen, wie weit das einen Nutzen im Sinne von AMNOG generieren kann.

Frau Dr. Drechsler (Boehringer): Nun, man kann die Patientenpräferenzstudie einmal für die Auswahl patientenrelevanter Endpunkte nutzen, die man zum Beispiel auch in der Studie erhoben hat, und danach, wie der Patient diese patientenrelevanten Endpunkte gewichtet. Deswegen – wie ich am Anfang gesagt habe – kann es hilfreich sein, wenn man den Zusatznutzen über verschiedene Endpunkte saldiert, solche Evidenz einzubeziehen.

Herr Hecken (Vorsitzender): Frau Dr. Faber, eine Nachfrage dazu?

Frau Dr. Faber: Ich hatte jetzt verstanden, dass Sie 211 deutsche Patientinnen und Patienten befragt haben. Da aber der Anteil asiatischer Patienten so hoch war, wäre meine Zusatzfrage: Warum haben Sie das da nicht gemacht?

Herr Hecken (Vorsitzender): Frau Drechsler.

Frau Dr. Drechsler (Boehringer): Wir gehen grundsätzlich nicht davon aus, dass asiatische Patienten oder deutscher Patienten Unterschiede in den Präferenzen haben. Wir wollten wirklich für den deutschen Markt eine Patientenpräferenzstudie durchführen.

Herr Hecken (Vorsitzender): Herr Pfannkuche, dazu eine Frage?

Herr Dr. Pfannkuche (Boehringer): Ganz kurz dazu. Ich will noch einmal verdeutlichen, was der Sinn der Präferenzstudie war und warum wir sie durchgeführt haben. Wir müssen uns vorstellen, man hat drei Endpunkte, die einen Zusatznutzen zeigen, drei Endpunkte, die keinen Zusatznutzen zeigen. Dann ist halt die Frage: Wie gleichwertig gehen die ein? Ist es wie im Fußball ein 3 : 3 oder ein Unentschieden, oder kann ich das eine als Bonustor mit Sternchen sehen, sodass es höher gewichtet eingeht? Ich denke, das war das Ziel, weswegen wir die Patientenpräferenzstudie durchgeführt haben.

Das Zweite, was angesprochen wurde, waren die Therapieabbrüche in der Studie. Ich denke, diese Frage war noch offen. Dazu würde Herr Lechner kurz etwas sagen, um auch das zu klären.

Herr Hecken (Vorsitzender): Herr Lechner, bitte.

Herr Dr. Lechner (Boehringer): Ich antworte auf die Frage von Frau Teupen zu den Therapieabbrüchen. Im Afatinibarm der LUX-Lung-3-Studie waren es 14 Prozent. Die Abbrüche wegen der am häufigsten vorkommenden Nebenwirkungen Diarrhö und Haut waren 1 Prozent und 0 Prozent.

Herr Hecken (Vorsitzender): Herr Kaiser.

Herr Dr. Kaiser: Zur Patientenpräferenzstudie. Sie hatten es im Dossier noch nicht vorgelegt, sondern es ist jetzt in den Stellungnahmen. Ich habe es kurz überflogen. Ist es richtig, dass in diesen sieben Attributen das Gesamtüberleben nicht vorhanden ist? Das heißt, das Gesamtüberleben spielt keine Rolle?

Die zweite Frage. Ist als siebtes Attribut die Frage, ob es eine Tablette oder eine Infusion ist? Ich frage dies, ob ich es richtig verstanden habe, dass das das Set ist, das dann beurteilt wurde, und das Gesamtüberleben, wie gesagt, keine Rolle spielte.

Frau Dr. Drechsler (Boehringer): Wir haben die Patientenpräferenzstudie im Dossier dargestellt, in gewissen Teilen, haben sie auch angehängt, im Anhang ist der Abschlussbericht von Herrn Professor Mühlbacher drin.

Ihre zweite Frage, Herr Kaiser, zum OS. Uns ist sehr wohl bekannt, dass OS ein sehr wichtiger patientenrelevanter Endpunkt ist, das sehen wir genauso. Bei uns war, als wir die Studie aufgesetzt haben, aber wichtig: Welche Endpunkte sind für Patienten über Overall Survival weiterhin interessant, weil wir interessiert daran waren, eine Gewichtung bezüglich Symptomdaten und möglichen Nebenwirkungen zu bekommen. Deswegen haben wir die Fragestellung auf diese eingeschlossenen Attribute fokussiert.

Herr Hecken (Vorsitzender): Verstehe ich das jetzt richtig: Haben Sie OS vor die Klammer gezogen und dann nicht abgefragt, oder ist es mit abgefragt worden und dann hinten runtergefallen?

Frau Dr. Drechsler (Boehringer): Wir haben es als Attribut, als Therapieeigenschaft, nicht mit in die Patientenpräferenzstudie eingeführt. Ich habe letzte Woche mit Professor Mühlbacher gesprochen. Er hat beim NSCLC eine weitere Patientenpräferenzstudie laufen. Da hat er OS mit drin, und da ist OS an erster Stelle, gefolgt von PFS und den Symptomen an zweiter Stelle.

Herr Hecken (Vorsitzender): Frau Müller.

Frau Dr. Chr. Müller: Es hat jetzt doch etwas länger gedauert. Herr Kaiser hat die stratifizierte Randomisierung nach Mutationstyp thematisiert. Das war auch bei mir auf der Agenda. Vielleicht könnten Sie noch einmal darauf eingehen. Sie haben bei der LUX-Lung-3-Studie nicht nur nach Mutationstyp stratifiziert, sondern auch nach Ethnie, nach asiatisch und nicht asiatisch. Sie werden sich etwas dabei gedacht haben.

Die zweite Frage, die ich habe, wurde von Dr. Sebastian schon kurz angerissen – das ist schon eine Weile her –, und sie geht eigentlich an alle Stellungnehmer: Wie sehen Sie den Stellenwert von Afatinib? Wie sehen Sie das im Vergleich zu den TKIs der ersten Generation? Manchmal klang ein Vergleich an, der aber kein direkter Vergleich ist. Eine Frage an den pU: Planen Sie dahin gehend etwas für die Zukunft, möglicherweise direkt vergleichende Studien?

Herr Hecken (Vorsitzender): Herr Pfannkuche, bitte.

Herr Dr. Pfannkuche (Boehringer): Ich gehe auf die letzte Frage zuerst ein, was den Vergleich der Thyrosinkinase-Inhibitoren anbelangt. Da läuft eine Studie, die LUX-Lung-7-Studie, die diese Frage untersuchen wird. Zu den genauen Zeiten und dem genauen Studiendesign kann Herr Lechner Weiteres sagen. Auf die beiden anderen Punkte, stratifizierte Randomisierung und Ethnie, wird Herr Voß noch etwas detaillierter eingehen.

Herr Hecken (Vorsitzender): Die Ethnie hatten wir partiell schon angesprochen. – Herr Lechner.

Herr Dr. Lechner (Boehringer): Noch kurz zu dem Vergleich mit den ersten Generationen an reversiblen Thyrosinkinase-Inhibitoren. Es gibt im Moment, wie schon erwähnt, keine Daten aus einem direkten Vergleich, es gibt nur sogenannte Cross-trial Comparisons, die natürlich von der wissenschaftlichen Validität zu hinterfragen sind. Die LUX-Lung-7 Studie läuft in einer Patientenpopulation, die vergleichbar der in der LUX-Lung-3-Studie ist, also Erstlinie EGFR-Mutation-positiv, nicht behandelt, und vergleicht Gefitinib mit Afatinib. Wir erwarten uns die Daten im Jahre 2015.

Herr Hecken (Vorsitzender): Herr Voß noch einmal zur Ethnie.

Herr Dr. Voß (Boehringer): Ich war in die Diskussion nicht involviert, warum Ethnie als Stratifizierungsvariable aufgenommen wurde. Aber es ist die erste Studie, die beide Ethnien genommen hat. Deshalb ist es eine wichtige Variable. Wir sehen mit dem Ergebnis allerdings, dass die Ethnie anscheinend keine Rolle spielt. Damit haben wir sozusagen diese Frage beantwortet. Vielleicht kann Herr Lechner noch etwas zu den Gründen sagen.

Herr Hecken (Vorsitzender): Das hatten Sie schon dargestellt. – Herr Professor Wöhrmann, Sie wollten darauf antworten? – Wenn Sie schon das Wort haben: Eben ist, als es um die ECOG-2-Patienten ging, so salopp gesagt worden, da kann man schon davon ausgehen, mit dem Blick auf allerlei Umstände, dass hier eine Wirksamkeit belegt ist. Das ist dann so im Raume stehen geblieben. Vielleicht können Sie dazu noch einen Satz sagen.

Herr Prof. Dr. Wörmann (DGHO): Zunächst die Antwort auf die Frage, ob die drei jetzt vorhandenen Tyrosinkinase-Inhibitoren völlig unterschiedlich zu werten sind. Im Moment würden wir erst einmal von Gleichwertigkeit ausgehen. Wir gehen von einer Substanzklasse aus. Es kann durchaus etwas anderes herauskommen, wenn man eine große vergleichende Studie macht. Es gibt inzwischen eine wirklich gut publizierte erste Metaanalyse, die etwa auf gleiche Wirksamkeit hinausläuft. Wir haben einmal eine Tabelle gemacht, aus der man ersehen kann, dass ihre Remissionsraten relativ im gleichen Fenster liegen. Es gibt potentielle leichte Unterschiede, was die Nebenwirkungen angeht. Das ist im Grunde das, was wir vorhin diskutiert haben. Auch da muss man das im direkten Vergleich sehen.

Der zweite Punkt: ECOG 2. Das Problem mit ECOG 2 ist – deswegen haben wir uns auch erlaubt, die Vergleichstherapie zu kritisieren –: Das war früher eine Grundsatzdiskussion. Fängt man bei jemandem mit einem sehr schlechten Allgemeinzustand überhaupt mit einer Chemotherapie an? Wir haben es hier nicht nur mit der Krankheit selbst, sondern mit Komorbidität und schlechtem Allgemeinzustand zu tun. Deswegen finden wir es schwierig, jetzt eine Chemotherapie als Vergleich anzusetzen. Die persönliche Diskussion mit dem Patienten läuft anders. Die Frage ist: Gibt es eine Sinnhaftigkeit für die Therapie? Wenn es eine Sinnhaftigkeit gibt, dann trauen wir uns heute, mit einer Tablettentherapie anzufangen, wenn diese Mutationen da sind, und würden das in der Realität auch geben. So haben wir, bevor es Afatinib gab, gesagt, dann sollte man die TKI als Erstlinie ansetzen. Wir würden gar nicht auf die Idee gekommen, Cisplatin und auch nicht das manchmal durchaus nebenwirkungsbeladene Gemcitabin einzusetzen. Deswegen können wir nicht sagen, es ist belegt. Aber in der Realität sind die sehr gut verträglich. Wir wissen auch schon nach ein oder zwei Wochen, ob die Patienten reagieren. Es ist deutlich schneller als mit jeder Chemotherapie. Das ist an der Realität vorbei.

Herr Hecken (Vorsitzender): Danke schön, Herr Wörmann. – Frau Müller, war Ihre Frage jetzt beantwortet?

Frau Dr. Chr. Müller: Ja, sie ist beantwortet.

Herr Hecken (Vorsitzender): Dann haben wir Frau Eckardt.

Frau Eckhardt: Ich habe noch zwei Verständnisfragen. Zum Ersten ist mir eines nicht ganz klar. Sie haben für das PFS eine Sensitivitätsanalyse und eine Analyse gemacht, um das Ungleichgewicht für die Imbalancen zu justieren. Warum haben Sie das nicht auch für Overall Survival gemacht, oder habe ich es vielleicht nur übersehen?

Die zweite Frage wäre an die Praktiker, inwieweit bei der Bestimmung des EGFR-Status die einzelnen Mutationen mitbestimmt werden.

Herr Hecken (Vorsitzender): Wer möchte? – Herr Sebastian, Sie hatten sich gemeldet.

Herr Dr. Sebastian (AIO): Wir haben noch keinen allgemeingültigen Standard, wie die EGFR-Rezeptor-Mutationen in Deutschland getestet werden. Es gibt verschiedene Ansätze. Es ist allerdings so, dass es zertifizierte Labors gibt, die mit einer sehr hohen Sensitivität die Mutationen testen. Das geht dann hoch bis zum Next Generation Sequencing, wo wir wirklich alles bekommen. Aber die Mutationen, die in der Studie in 90 Prozent der Fälle behandelt wurden, hat man in dem Ansatz drin. Das war Deletion im Exon 19 und die Punktmutation im Exon 21. Die meisten Institute testen jetzt auch die seltenen Mutationen.

Vielleicht eine kurze Anmerkung noch zum Performancestatus 2. Das ist genau die Patientenpopulation – da muss ich Sie dezidiert unterstützen –, die selbstverständlich von der Therapie mit einem Thyrosinkinase-Inhibitor wesentlich mehr profitiert. Das sind kranke Patienten. Da müssen wir die ja so mittelgut geliebten Surrogatparameter PFS und vor allem auch die Ansprechraten – sie haben eine Ansprechraten von 60, 70 Prozent; das ist in der Onkologie nach wie vor ganz selten und als Sensation zu werten – betrachten. Gerade für Patienten mit Performancestatus 2 oder 3, die vor allem durch diese Erkrankung diesen schlechten Performancestatus haben, ist das eigentlich die einzig realistische Therapieoption.

Herr Hecken (Vorsitzender): Danke schön, Herr Sebastian. – Herr Voß.

Herr Dr. Voß (Boehringer): Es war, glaube ich, noch die zweite Frage offen, warum wir diese Analyse gemacht haben. Mir liegt da nichts vor. Ich glaube, wir haben das bisher nicht gemacht und würden das gegebenenfalls machen, wenn das Gesamtüberleben mit mehr Events besser abschätzbar ist.

Herr Hecken (Vorsitzender): Ist Ihre Frage damit beantwortet, Frau Eckardt?

Frau Eckhardt: Es ist nicht gemacht worden. Warum, ist nicht klar. Weil Sie noch zu wenig Ereignisse haben?

Herr Dr. Voß (Boehringer): Diese Analyse ist aus dem Studienbericht. Dort war die Analyse des PFS die primäre Analyse, und für die haben wir das Ganze gemacht, weil wir sehen wollten: Hat das einen Effekt, oder hat es keinen? Wir haben es dann eben nicht für die sekundären Endpunkte gemacht.

Herr Hecken (Vorsitzender): Danke schön. – Herr Mayer.

Herr Dr. Mayer: Ich habe eine Frage zu den Mutationen. Ich habe da eine Schwierigkeit. Der G-BA muss jetzt einen Zusatznutzen für die Substanz Afatinib aussprechen. Jetzt haben das IQWiG und auch Sie diese drei Mutationen als Stratifizierungsmerkmal gewählt. Jetzt heißt es einerseits, wir haben erheblichen bzw. keinen Zusatznutzen für die beiden häufigen Mutationen und einen geringeren für die seltenen Mutationen. Wenn man die Gesamtpopulation betrachtet, haben wir keinen Unterschied im zweiten Datenschnitt in der Gesamtüberlebensrate. Ich tue mich jetzt ein bisschen schwer. Es ist alles ein biometrisches Problem. Wir haben die ersten beiden häufigen Mutationen – zumindest die Del 19 ist signifikant pro Afatinib –, die werten wir für den Zusatznutzen, und die seltenen, weil sie eben so inhomogen sind, aber eigentlich dazu beigetragen haben, dass die Gesamtmortalität nicht

signifikant unterschiedlich ist, sollen wir jetzt weglassen. Es fällt mir schwer, im Endeffekt eine Aussage zu treffen. Entweder müssen wir alle Mutationen in die Nutzenbewertung einbeziehen, oder wir machen keine Unterschiede und müssen dann alle über einen Kamm scheren, dann auch die 19er und 21er. Mir fällt es schwer, weil es rein biometrisch ist. Wir haben auch bei den seltenen Mutationen einen biometrisch-statistisch nachgewiesenen geringeren Zusatznutzen, genauso wie wir bei den anderen einen positiven Effekt haben. Ich tue mich schwer, zu sagen: Was lassen wir jetzt weg, oder was werten wir, oder werten wir überhaupt keine unterschiedlichen Mutationen?

Herr Hecken (Vorsitzender): Herr Voß dazu, bitte.

Herr Dr. Voß (Boehringer): Ich denke, das ist wieder eine Frage an mich. Wenn wir uns den Effekt der Del-19-Mutation anschauen und ihn mit dem der L858R vergleichen – das sind die zwei großen homogenen Gruppen –, dann sehen wir den Effekt beim Gesamtüberleben sowohl in der LUX-Lung-3-Studie als auch in der LUX-Lung-6-Studie. Biometrisch sieht man in beiden den gleichen Effekt. Wir haben biologisch, soviel ich weiß, keine Begründung, warum das so ist. Wir können das nicht begründen. Man sieht es eben in beiden Studien. Für alle anderen Endpunkte sieht man es nicht. Deshalb kann man das aus unserer Sicht höher gewichten.

Herr Hecken (Vorsitzender): Herr Mayer, eine Nachfrage.

Herr Dr. Mayer: Es fällt mir trotzdem schwer. Man hat danach stratifiziert und hat ein Gesamtergebnis über alle Patienten vom Overall Survival, das halt nicht Unterschiede zwischen den beiden Therapiearmen in der LUX-Lung-3-Studie zeigt. Ich habe Schwierigkeiten, rein statistisch manche Dinge als inplausibel zu werten und sie wegzulassen, und die anderen, die plausibel sind, werten wir dann als Hauptbeleg für den Nutzen. Vielleicht müsste ich auch das IQWiG fragen, was biometrisch überhaupt okay ist. Wenn man sagt, man vertraut den Mutationsstratifizierungen gar nicht, dann würde ich sagen: Okay, dann muss man eigentlich nur die Gesamtpopulation vergleichen dürfen.

Herr Hecken (Vorsitzender): Herr Kaiser.

Herr Dr. Kaiser: Die Frage ist natürlich völlig berechtigt. Herr Wörmann, wenn Sie sagen, Sie geben in den Leitlinien keine Empfehlung auf Basis der Subgruppenanalyse, dann ist das Ergebnis von Afatinib, dass es keinen Vorteil bezüglich Gesamtüberleben gibt, unabhängig davon, wie man PFS oder was auch immer bewertet. Diese Studie zeigt keinen Vorteil für das Gesamtüberleben – wenn man in die Gesamtpopulation schaut, wenn man sagt, Subgruppenanalysen spielen keine Rolle für Entscheidungen.

Schauen wir uns die Subgruppen an. Man kann immer darüber diskutieren, wenn die Subgruppenergebnisse sehr nahe beieinanderliegen und man nur einen Hinweis auf eine Effektmodifikation hat, was ist das dann für eine Folge? Hier haben wir eine sehr starke Effektmodifikation mit sehr stark divergierenden Ergebnissen. Ich glaube schon, dass es da berechtigt ist, daraus eine Schlussfolgerung zu ziehen; denn das hat unmittelbare Konsequenzen für die Behandlung der Patienten. Da spielt die Power überhaupt keine Rolle mehr, weil man in den Gruppen tatsächlich den statistisch signifikanten Unterschied sieht. Das Powerargument zieht hier überhaupt nicht. Man kann sagen, dass für die einen Patienten ein Vorteil bezüglich des Überlebens nachgewiesen ist, und für die anderen Patienten zeigt diese Studie, dass es sogar eine Verringerung des Überlebens gibt.

Jetzt kann man wieder die Diskussion führen, in den seltenen Mutationen gibt es noch diese oder jene Patienten. Das muss man dann aber entsprechend aufbereiten. Das ist hier jetzt nicht der Fall. Eben wurde angemerkt, man kann nicht das eine und das andere machen. Entweder man sagt, die

Subgruppenanalysen interessieren nicht, dann geht man auf das Gesamtstudienresultat, dann hat man keinen Vorteil für das Gesamtüberleben, oder man sagt, wir sehen hier einen so deutlichen Effekt in dieser Gruppe, der auch zu „erheblich“ in der Einstufung führt, dann muss man aber konsequent sein und sagen, dann muss ich natürlich die anderen Subgruppen entsprechend betrachten. Das eine geht nicht ohne das andere.

Herr Hecken (Vorsitzender): Danke schön, Herr Kaiser. – Jetzt habe ich Herrn Pfannkuche.

Herr Dr. Pfannkuche (Boehringer): Ich denke, unter dem Strich sehen wir, was die Gesamtpopulation anbelangt, einen Vorteil im Hinblick auf die Symptome, sowohl die Zeit bis zur Verschlechterung der Symptome als auch die Zeit bis zur Verbesserung der Symptome. Wir sehen bei der Gruppe einen Vorteil ganz klar auch im Hinblick auf die Lebensqualität. Wir sehen auch in der Gesamtpopulation, dass die klassenspezifischen Nebenwirkungen an dieser Stelle handhabbar sind, weswegen wir an dieser Stelle für die Gesamtpopulation von einem beträchtlichen Zusatznutzen ausgehen.

Was die Del-19-Mutation anbelangt: Ja, es ist eine starke Effektmodifikation da, die den Überlebensvorteil in dieser Population zeigt. Die kann man nicht wegdiskutieren. Das ist sicherlich eine schwierige Frage für die Entscheidung, wobei ich Ihnen hier jetzt aber leider keine Antwort geben kann, was falsch oder richtig ist. Ich denke, das ist eine ganz diffizile Frage, weil man hier einfach die eine Gruppe hat, die anscheinend einen sehr großen Benefit von der Therapie hat.

Herr Hecken (Vorsitzender): Herr Sebastian.

Herr Dr. Sebastian (AIO): Ich glaube, um es klar zu sagen, wir sind auf dem Holzweg, wenn wir die Studie mit dem Gesamtüberleben interpretieren. Das ist ein Patientenkollektiv, das von der Inhibition der mutierten Thyrosinkinase abhängt. Es ist so, dass sicherlich über zwei Drittel der Patienten im Verlauf ihrer Erkrankung einen Thyrosinkinase-Inhibitor bekommen haben. Ob das nun ein Erstgenerations- oder Zweitgenerationsinhibitor ist, das Gesamtüberleben wird dadurch, dass Patienten im Chemotherapiearm später noch einmal einen Thyrosinkinase-Inhibitor bekommen haben, verwässert. Wir können immer nur wieder sagen, die Thyrosinkinase-Inhibitoren in dieser Erkrankung waren eine Revolution. Wir können, wenn wir es mit historischen Kollektiven vergleichen, zeigen, dass wir das Gesamtüberleben für diese Patientengruppe nur durch die Gabe von TKI deutlich verlängern konnten. Das ist auch klinisch eindrucksvoll.

Jetzt haben wir verschiedenste Subgruppen, und wir haben die gar nicht für die weiteren Therapien kontrolliert. Ich habe ein großes Problem, dass wir bei diesem Schlüssel-Schloss-Prinzip wie in dieser Erkrankung das Gesamtüberleben hernehmen. Für mich sind das progressionsfreie Überleben und die Ansprechrate, gerade bei den seltenen Mutationen, wesentlich interessanter und auch zielführender. Haben die Patienten mit einer sehr frühen Mutation ein Ansprechen, dann ist das eine wirksame Therapie, möglicherweise nicht ganz so wirksam wie die klassische Mutation. Aber ich glaube nicht, dass wir das Patientenschicksal gut abbilden, wenn wir das Gesamtüberleben in dieser Patientengruppe betrachten. Das gilt auch für die Diskussion – das hatten wir bei Crizotinib genauso –, wir werden diese Diskussion in den nächsten Sitzungen immer wieder führen, wenn wir solche zielgerichteten Substanzen beim Nachweis einer Mutation prüfen. Ich halte es für ganz schwierig, dass wir da immer noch auf dem Gesamtüberleben herumreiten.

Herr Hecken (Vorsitzender): Herr Kaiser hat eben, wenn ich es richtig gesehen habe, auch nur auf den Widerspruch hingewiesen, wenn man sagt, wenn man auf der einen Seite aus dem Gesamtüberleben sogar einen erheblichen Zusatznutzen ableiten will und in einem anderen Bereich aus welchen

Gründen auch immer dann eben sogar eher Nachteile bezogen auf diesen patientenrelevanten Endpunkt und eher unklare Aussagen hinsichtlich möglicher Vorteile da sind, dann muss man sich irgendwann einmal entscheiden, hopp oder top. Wir hatten vor etwa einer halben Stunde die Fragestellung: Welche Rolle hat das Gesamtüberleben bei den Patientenpräferenzen gespielt? Wir hatten herausgearbeitet, dass es vor die Klammer gezogen worden ist. Ich glaube, man muss schon mit einer gewissen Stringenz herangehen und kann nicht sagen: Okay, in der zahlenmäßig größten Gruppe betrachten wir wirklich zentriert das Gesamtüberleben, sagen, da haben wir den Vorteil von 10, 11 Monaten in der relativ kleinen Subgruppe, die eben hier ohne Zusatznutzen herausgegangen ist, stellen es aber hinten an, sagen, bei ein paar gibt es gigantische Effekte, bei anderen gibt es möglicherweise negative Effekte, und versuchen, das dann irgendwie zu kompensieren.

Man muss da schon ein Stück weit stringent in der Argumentation bleiben. Nicht mehr und nicht weniger wollte Herr Kaiser zum Ausdruck bringen. Wenn wir uns hier verständigen würden, okay, dann betrachten wir das Gesamtüberleben über die gesamte Studienpopulation, und dann versuchen wir, anhand anderer patientenrelevanter Endpunkte und des Gesamtüberlebens über die Studie eine Zusatznutzenbewertung anzustellen, dann wage ich einmal die Prognose, es wird wahrscheinlich unter dem Strich weniger gut sein können, als wenn wir hier von populations- und subgruppenbezogenen Betrachtungen ausgehen. Aber sei es drum, es ist nicht meine Rolle hier, Ihre Argumentation zu überprüfen oder zu hinterfragen. Ich will nur die Logik, die jedenfalls ich in den Ausführungen von Herrn Kaiser gesehen habe, aus meiner Sicht noch einmal unterstreichen. – Herr Wörmann dazu.

Herr Prof. Dr. Wörmann (DGHO): Direkt dazu. Ich würde deutlich dafür argumentieren, die gesamte Gruppe zu nehmen, weil die Studie so angelegt war. Ich glaube, dass wir hier ein typisches Beispiel für eine Sequenztherapie haben. Das sieht man an den Cross-overs. Die Cross-overs waren zum einen in dem Afatinibarm, aber auch in anderen TKIs, unter dann hatten noch fast die Hälfte der Patienten, die TKIs hatten, noch eine Chemotherapie bekommen, die in einem guten Allgemeinzustand waren, was der deutschen Situation entspricht, weil sie dafür zugelassen sind. Wir sehen auch, wenn ein Patient nicht mehr anspricht, er weiterhin in einem guten Zustand ist, bekommt er eine Chemotherapie. Das ist sehr real. Das heißt, wir reden über eine Sequenztherapie, mit der wir ganz große Probleme haben, Overall Survival am Ende noch zu bewerten.

Das, was Sie sagen, Herr Kaiser, kann ich als Entscheidung sehr gut nachvollziehen. Aber im Grunde reden wir dann auch darüber, ob eine frühe Afatinibtherapie einen Unterschied gegenüber einer späteren Afatinibtherapie macht. Da kann es eine Gruppe geben, die dann vielleicht mit einem Overall Survival herauskommt. Ich habe bewundert, mit welcher Akribie das IQWiG Overall-Survival-Daten analysiert hat. Ich finde das hoch anerkennenswert. Ich finde es aber auch Rosinenpicking, sich das herauszusuchen, bei dem ein großer Unterschied herauskommt.

Von der Konsequenz her ist es, glaube ich, richtiger, das als Gesamtgruppe zu sehen, denn die Untergruppen waren nicht primär so angelegt. Bei anderen Mutationen kommt dann heraus, dass eine Untergruppe der Kontrollgruppe, glaube ich, noch 11 Patienten hat. Es ist aber schwierig, daraus noch einen Vergleich zu berechnen, der aussagekräftig ist. Ich würde dabei bleiben, ich glaube, dass man die Gesamtgruppe werten muss. Wir haben gedacht, das progressionsfreie Überleben war so dramatisch gut, plus eindeutigen Patientenbenefit, das ist das, was zu Buche schlagen sollte. Wenn das dann weniger wert als das Overall Survival ist, dann ist das trotzdem das, was wir als Wert sehen würden, und es ist der Realität entsprechend. Wir können die Patienten nach der Studie in dieser Situation nicht weiter behandeln, wenn es eine Erstlinientherapie ist.

Herr Hecken (Vorsitzender): Danke für dieses klare Bekenntnis. – Herr Kaiser dazu.

Herr Dr. Kaiser: Zunächst einmal: Wir picken ja keine Rosinen. Für uns sind das alles Holzklötzchen ohne Wertigkeit, und wir schauen einfach, wie die Ergebnisse sind. Bei der Interpretation von Cross-over – es ist ein wesentlicher Punkt: spielt Cross-over eine große Rolle, oder spielt Cross-over keine große Rolle? – kann man sich zwei Dinge anschauen. Das eine ist: Wenn man die Beziehung zwischen PFS und Overall Survival an dieser Stelle herstellen möchte, dann sieht man, dass auch beim PFS die Ergebnisse identisch sind. Die gehen – vor Cross-over – für die Patienten mit seltenen Mutationen eben auch in die negative Richtung. Zusätzlich kann man sich die Kaplan-Meier-Analysen für das Overall Survival anschauen. Wenn Cross-over eine große Rolle spielt, dann würden Sie nicht schon früh ein Auseinanderdriften der Kurven haben. Sie haben das frühe Auseinanderdriften schon zuungunsten von Afatinib in dieser Gruppe. Das kann nicht durch Cross-over verursacht worden sein. Das heißt, ich kann verstehen, dass man Cross-over berücksichtigen muss, aber ich glaube, dass man nicht einfach formal dabei stehen bleiben sollte: Wir haben hier ein Cross-over und deswegen ein Problem mit dem Gesamtüberleben, sondern dass man sich die Gesamtheit der Informationen anschauen sollte. Wenn man sich die Gesamtheit der Informationen anschaut, denken wir, dass man das Gesamtüberleben interpretieren kann.

Herr Hecken (Vorsitzender): Danke schön. – Herr Lechner dazu.

Herr Dr. Lechner (Boehringer): Ich möchte abschließend darauf hinweisen, dass es uns schon darum geht, dass der Overall Survival Benefit für die größte Gruppe mit den am meisten auftretenden Mutationen – hier sieht man die Unterscheidung zwischen dem Del 19, bei dem der Overall Survival Benefit sehr ausgeprägt ist –, den Punktmutationen im Exon 21, nicht so ausgeprägt ist. Wir verstehen nicht warum. Das ist Tumorbilogie. Da verstehen wir noch zu wenig. Wir lernen ja auch aus diesen Studien. Für die Subgruppe der sehr seltenen Mutationen verstehen wir einfach zu wenig. Ich möchte Herrn Kaiser um Vorsicht bitten. Wir wissen es nicht. Sieht man etwas, oder sieht man es nicht? Es ist eine sehr kleine heterogene Gruppe. Aber von dieser Gruppe auf das Cross-over, auf die Gesamtpopulation, zurückzugehen, ist meiner Meinung nach auch nicht unbedingt adäquat. Ich möchte schon betonen: Wir sind uns bewusst, mit diesen seltenen Mutationen ist es schwer, die Daten zu interpretieren. Für uns ist allerdings die Lage für Overall Survival für die Gruppe der Deletion 19 sehr klar. Dies wollte ich hier noch einmal betonen, wenn ich darf.

Herr Hecken (Vorsitzender): Danke schön. – Herr Dintsios, jetzt haben Sie das Wort.

Herr Dr. Dintsios (vfa): Ich führe auf eine ganz andere Thematik zurück, weil ich das Wort da nicht bekommen hatte. Es sind vier kleine Punkte. Ich mache es auch sehr schnell. Es geht um die Präferenzstudie bei den Patienten. Einmal kam die Frage auf, ob das Relevanz im Rahmen der AMNOG-Bewertung hat. Natürlich hat es eine Relevanz, ganz einfach weil Sie bei der Gegenüberstellung positiver und negativer Effekte auch die Wertigkeit der entsprechenden Endpunkte betrachten, um überhaupt das – ich sage es in Anführungszeichen – „Saldieren“ betreiben zu können. Hier ist eine Aussage von denjenigen, die selber vom Krankheitsbild betroffen sind, sicherlich wertvoller und belastbarer als eine, die vielleicht Experten treffen, die vielleicht aus der Statistik oder Biometrie kommen, aber das Krankheitsbild per se nicht kennen. So gesehen würde ich die Antwort als gegeben erachten, dass das AMNOG-relevant ist.

Das Zweite bezieht sich auf die Aussage von Herrn Kaiser und die Nichteinbeziehung von Overall Survival im Rahmen eines Discrete-Choice-Experiments. Discrete-Choice-Experimente haben einen Riesenmethodenvorteil. Sie führen zu keinem Rank Reversal, wenn man die entsprechenden Attribute mit ihren Ausprägungen herausnimmt, hineinfiltert oder ändert. Wir haben andere multikriteriale Entscheidungsanalysen, die dieses Phänomen durchaus zeigen. Es ist methodisch bedingt, dass sich

die Reihenfolge der anderen Endpunkte untereinander mit oder ohne Overall Survival, wenn man sie gegenüberstellt, überhaupt nicht ändert. Das ist ein methodisches Phänomen. Das liegt an der theoretischen Verortung der Mathematik dahinter. Das ist eine Probit-Logit-Regression, und die ist relativ stabil beim Ein- und Ausführen von entsprechenden Endpunkten.

Der dritte Punkt. Ich wollte hier das IQWiG befragen. Wir haben es auch in unserer Stellungnahme angeführt, aber heute ist es irgendwie vollkommen irrelevant gewesen. Die Deutung des EQ-5D und die Nichtberücksichtigung, weil es sozusagen nicht psychometrisch interpretiert wurde. Ich sage es in eigenen Worten. Auf gut Deutsch: Es waren nicht die einzelnen Ergebnisse der entsprechenden fünf Dimensionen mit ihren drei Ausprägungen dargestellt, sondern der entsprechende Nutzwert, der hergeleitet wird, auf Neudeutsch Utility. Dann muss uns das IQWiG auch ein bisschen aufklären. Ich habe es auch in den letzten drei Anhörungen vorgetragen, weil ich als Verbandsvertreter meinen Verbandsmitgliedern sagen muss: Macht es überhaupt noch Sinn, so etwas zu erheben und in die Dossiers einzubringen, wenn das IQWiG damit nichts anfangen will oder anzufangen weiß?

Ich schließe mit einem Punkt ab, der wieder mit den Endpunkten zusammen zu sehen ist. Eine interessante Konstellation, in der Einzelaspekte, die herausgebrochen werden, um zum Beispiel die Morbidität aus einem Lebensqualitätsinstrument zu beurteilen, dazu führen, dass diese Endpunkte der Kategorie „nicht schwerwiegend“ zugeordnet werden, dafür das Instrument selber, die Lebensqualität als solche, aber in der Kategorie mit den schwerwiegenden Ereignissen oder der schwerwiegenden Morbidität gleichgestellt wird. Das führt mich dazu, die entsprechende Schlussfolgerung zu ziehen, dass das Instrument selber mehr wert ist als die Summe seiner einzelnen Bestandteile. Da würden wir gerne das IQWiG zurückfragen wollen, wieso so vorgegangen wird, dass bei einem Morbiditätsendpunkt, der aus einem Lebensqualitätsinstrument hergeleitet wird, dieser in einer spezifische Kategorie zugeordnet wird, die unter denjenigen liegt, in denen man das Ergebnis aus dem gesamten Lebensqualitätsinstrument eigentlich zuordnen würde. Das ist für uns auch methodisch relevant.

Herr Hecken (Vorsitzender): Dankeschön, Herr Dintsios. – Herr Kaiser, ganz kurz bitte.

Herr Dr. Kaiser: Ganz kurz, zu zwei Sachen. Ich fange mit dem Letzten an. Das ist hier keine Diskussion des Methodenpapiers des IQWiG. Der G-BA wird seine Einstufung der Ergebnisse vornehmen. Dafür ist das, was wir gemacht haben, an dieser Stelle irrelevant. Aber Sie können diesen Punkt gern bei der nächsten Methodendiskussion des IQWiG einbringen. Bei der letzten Methodendiskussion hat der vfa das nicht eingebracht.

Der andere Punkt, EQ-5D. Sie können den Verbandsvertretern sagen, dass es Sinn macht, die Fragen einzeln auszuwerten und darzustellen, was völlig üblich ist und was man machen könnte. Und Sie können denen dann auch noch sagen, dass sie Analysen vollständig vorlegen für dieses Lebensqualitätsinstrument, so wie es für andere Lebensqualitätsinstrumente, in diesem Fall den QLQ, auch gemacht wurde. Es macht also keinen Sinn, für EQ-5D weniger zu machen als für einen QLQ, also zum Beispiel keine Subgruppenanalysen vorzulegen. Also einfach eine ganz reguläre Auswertung, und dann schauen wir uns an, was man davon verwenden kann.

Herr Hecken (Vorsitzender): Herr Dintsios dazu.

Herr Dr. Dintsios (vfa): Das ist auch in unserer Stellungnahme eingebracht. Ich verweise Herrn Kaiser gerne auf sein anderes Ressort im gleichen Institut. Der EQ-5D ist kein psychometrisches Instrument. Sie können es nicht so auswerten wie andere psychometrisch fundierte Instrumente. Er kommt aus der Random Utility Theory und leitet einen Nutzwert her. Das einzige, was Sie mit diesen fünf Dimensionen mit jeweils drei Ausprägungen machen, ist, Sie ordnen die Zustände zu. Das sind 243

Krankheitszustände. Hinterlegt ist eine klassische Regression mit Time Trade-off. Somit ist die Forderung, die Herr Kaiser hier aufstellt – da kann man die Entwickler des EQ-5D gerne befragen, die eine solche Darstellung nur als deskriptive Darstellung vorschlagen –, methodisch nicht zielführend. Sie hat mit der Entwicklung und mit der Verwertung dieses Instruments rein gar nichts zu tun.

Herr Hecken (Vorsitzender): Ich bitte um Verständnis, Herr Dintsios, ich nehme Ihre Äußerungen jetzt zu Protokoll. Ich kann sie nicht bewerten, weil sie so tief in die Methodik hineingehen und es hier offensichtlich auch Schulenkriege innerhalb einzelner Institutionen und über Institutionen hinweg gibt, sodass Sie mich an einer Stelle sehen, wo mir einfach die Intelligenz oder die Kenntnis fehlt, um das jetzt zielgerichtet in eine Diskussion einmünden zu lassen. Wir sind im Übrigen heute nicht hier – ihre Stellungnahme haben wir ja –, den nächsten Rundbrief an die Mitglieder ihres Verbandes vorzubereiten. Herr Kaiser hat eine Antwort gegeben. Wenn sie nicht zufriedenstellend ist, dann, bitte schön, Methodendiskussion beim IQWiG findet regelmäßig statt.

Weitere Fragen? – Keine. Dann würde ich dem pU noch einmal die Gelegenheit geben, zusammenfassend Stellung zu nehmen und aus Ihrer Sicht es noch einmal darzustellen. Bitte schön, Herr Pfannkuche.

Herr Dr. Pfannkuche (Boehringer): Vielen Dank für die Möglichkeit, ein paar abschließende Worte zu sagen. Wir halten insgesamt fünf Punkte für sehr wichtig und hier noch einmal erwähnenswert. Zum einen bezogen auf die Ausgangsfrage: Welchen Stellenwert hat die LUX-Lung-6-Studie? Unserer Ansicht nach ist diese Studie zu berücksichtigen, da gerade zu einem solch frühen Zeitpunkt nach der Markteinführung die bestverfügbare Evidenz berücksichtigt werden sollte, um ein umfassendes Bild über ein neues Arzneimittel zu erhalten, und das in diesem Fall eben zur Einschätzung der Aussagesicherheit, der Effektivität und den Nebenwirkungen, und das vor allem, weil wir konsistente Effekte zwischen den beiden Studien sehen und es einfach die Möglichkeit bietet, Subgruppeneffekte unabhängig voneinander zu testen.

Das Zweite ist: Der Zusatznutzen ist unabhängig vom Alter. Das zeigt sich beim Vergleich der beiden Studien LUX-Lung-3 und LUX-Lung-6 und auch bei der Variation der Trennwerte. Aus therapeutischer Sicht scheint es hier eher geboten, nach Komorbiditäten und dem Allgemeinzustand zu schauen.

Der dritte Punkt: Wir sehen keinen Unterschied zwischen den Ethnien. Auch da sind die Studien sehr konsistent.

Der vierte Punkt ist: Die Patientenpräferenzstudie zeigt, dass spezifische Lungenkrebs-symptome eine besondere Bedeutung für die Patienten haben und deswegen bei der Saldierung des Zusatznutzens entsprechend berücksichtigt werden müssten.

Der fünfte Punkt: Die Patientenpräferenzstudie ist eine von mehreren Möglichkeiten, wenn man hier herangeht. Aber es ist eben auch eine Möglichkeit, die man nicht links liegen lassen kann, weil hier systematisch Patienten befragt worden sind.

Wenn ich es unter dem Strich zusammenfasse, worin wir den Zusatznutzen sehen, dann kann ich an dieser Stelle noch einmal festhalten, wir sehen einen beträchtlichen Zusatznutzen für die Gesamtpopulation, weil wir im Hinblick auf die Symptomkontrolle, Zeit bis zur Verbesserung, Zeit bis zur Verschlechterung, auf beiden Ebenen einen Vorteil zeigen konnten. Wir konnten einen Vorteil bei der Zeit bis zum Progress, beim progressionsfreien Überleben, zeigen. Ich denke, auch das ist an dieser Stelle sehr erwähnenswert. Wir können einen Vorteil im Hinblick auf die Lebensqualität zeigen, und wir haben keine Signale, dass es einen Schaden aufseiten der Nebenwirkungen gibt. Ich denke, auch

das ist bei solchen Therapien ganz wichtig. Zusätzlich haben wir noch den sehr stark ausgeprägten Effekt bei der großen Gruppe der Patienten mit Del-19-Mutation, was das Gesamtüberleben anbelangt.

Dies ist aus unserer Sicht zusammengefasst, was unsere Einschätzung des Zusatznutzens anbelangt. Wenn im Nachgang hierzu noch Fragen bestehen, sind wir gerne bereit, weitere Daten bei Bedarf nachzureichen. – Vielen Dank.

Herr Hecken (Vorsitzender): Ganz herzlichen Dank für diese sehr konstruktive Diskussion. Ich will noch ein persönliches Wort zu den Patientenpräferenzstudien sagen. Sie werden sicherlich von niemandem außer Acht gelassen oder links liegen gelassen. Ich habe nur immer ein sehr großes Problem damit, wenn vorher selektiv bestimmte Endpunkte in den Raum gestellt werden. Der zweite Punkt ist dann die Belastbarkeit solcher Ergebnisse. Nur damit nicht der Eindruck aufkommt, es würde niemanden interessieren, was von Patienten präferiert wird.

Ich danke für die Diskussion, die uns ein Stück weitergeholfen hat in dem Prozess, von dem Herr Kaiser eben mehrfach gesagt hat, dass ihn der G-BA anzustellen hat, nämlich der Bewertung Ihrer Stellungnahmen auf der Basis der vom IQWiG vorgelegten Dossierbewertung. Das werden wir in den nächsten Tagen und Wochen zu bewältigen haben. Es wird sicherlich nicht ganz einfach werden, aber wir werden dann selbstverständlich all das, was hier vorgetragen worden ist, in unsere Betrachtungen einbeziehen.

Danke, dass Sie da waren. Schönen Tag noch. Es hat sich mittlerweile eingetrübt, sollte sich aber nicht auf unser allgemeines Befinden auswirken. Bis zum nächsten Mal.

Schluss der Anhörung: 13.04 Uhr