

Mündliche Anhörung



Gemeinsamer
Bundesausschuss

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

hier: Wirkstoff Nivolumab (D-717)

Videokonferenz des Gemeinsamen Bundesausschusses
am 6. Dezember 2021
von 13:30 Uhr bis 14:13 Uhr

– Stenografisches Wortprotokoll –

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

Frau Dr. Kupas

Frau Ehrhart

Frau Dr. Amann

Frau Rupp

Angemeldete Teilnehmende der Firma **Seagen Germany GmbH:**

Herr Koch

Frau Dr. Rancea

Angemeldete Teilnehmende der Firma **Merck Serono GmbH:**

Frau Reinke

Frau Dr. Knaup

Angemeldete Teilnehmende der Firma **Eisai GmbH:**

Frau Dahlems

Frau Dr. Schneider

Angemeldete Teilnehmende der Firma **Amgen GmbH:**

Frau Camdere

Frau Stein

Angemeldete Teilnehmende der Firma **MSD Sharp & Dohme GmbH:**

Frau Dr. Örter

Frau Dr. Steck

Angemeldeter Teilnehmender für die **Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten e. V. (DGVS):**

Herr Dr. Pox

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

Herr Prof. Dr. Wörmann

Angemeldeter Teilnehmender für den **Verband Forschender Arzneimittelhersteller e. V. (vfa):**

Herr Dr. Rasch



Herr Prof. Hecken (Vorsitzender): Herzlich willkommen im Unterausschuss Arzneimittel des G-BA! Wir behandeln Nivolumab in Kombination mit Ipi zur Behandlung des metastasierten Kolorektalkarzinoms mit Mismatch-Reparatur-Defizienz oder hoher Mikrosatelliteninstabilität bei Erwachsenen. Wir haben eine Dossierbewertung des IQWiG vom 28. Oktober 2021, die die Basis des Stellungnahmeverfahrens für das neue Anwendungsgebiet für Nivolumab darstellt. Stellung genommen haben der pharmazeutische Unternehmer BMS, als weitere pharmazeutische Unternehmer Amgen GmbH, Eisai GmbH, Merck Sorono GmbH und Seagen Germany GmbH. Von den Fachgesellschaften haben eine gemeinsame Stellungnahme abgegeben die Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie, die Deutsche Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten e. V. und die Arbeitsgemeinschaft Internistische Onkologie der Deutschen Krebsgesellschaft. Außerdem liegt eine Stellungnahme des Verbandes Forschender Arzneimittelhersteller vor.

Ich muss die Anwesenheit überprüfen, weil wir heute wieder Wortprotokoll führen und das dort entsprechend vermerkt werden muss. Für den pharmazeutischen Unternehmer zugeschaltet sind Frau Dr. Kupas, Frau Ehrhart, Frau Dr. Amann und Frau Rupp, von der Deutschen Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten Herr Dr. Pox, für die DGHO Herr Professor Dr. Wörmann, für Seagen Germany GmbH Herr Koch und Frau Dr. Rancea, für Merck Serono Frau Reinke und Frau Dr. Knaup, für Eisai Frau Dahlems und Frau Dr. Schneider, für MSD Frau Dr. Örter und Frau Dr. Steck, für Amgen Frau Camdere und Frau Stein sowie für den vfa Herr Dr. Rasch.

Dann würde ich zunächst dem pharmazeutischen Unternehmer die Gelegenheit geben, aus seiner Sicht die wesentlichen Punkte, bezogen auf den Wirkstoff in diesem Anwendungsgebiet und auf die Dossierbewertung des IQWiG, zu adressieren. Wer macht das für den pharmazeutischen Unternehmer? – Frau Kupas, bitte schön, Sie haben das Wort.

Frau Dr. Kupas (BMS): Danke schön. – Sehr geehrter Herr Professor Hecken! Sehr geehrte Damen und Herren! Danke für die freundliche Begrüßung und die Möglichkeit, ein paar einleitende Worte zu sagen. Bevor ich auf unsere umfangreiche Stellungnahme eingehe, möchte ich das Team vorstellen, das heute hier ist und Ihre Fragen beantworten wird. Frau Ehrhart ist Biostatistikerin und wird alle Fragen zu den methodischen Aspekten beantworten. Frau Dr. Amann wird die Fragen zu den medizinischen Aspekten beantworten. Frau Rupp beantwortet alle weiteren Fragen zum Dossier. Ich bin Katrin Kupas; ich leite bei BMS Deutschland den Bereich Biostatistik.

Bevor ich zu den wichtigsten Punkten des indirekten Vergleichs komme, möchte ich ein paar Worte zur Indikation sagen. Das Anwendungsgebiet von Nivolumab in der Kombination mit Ipilimumab umfasst Erwachsene mit metastasiertem Kolorektalkarzinom mit einer Mismatch-Reparatur-Defizienz oder hochfrequenter Mikrosatelliteninstabilität – ich werde im Folgenden immer MSI-high sagen –, die bereits eine fluoropyrimidinbasierte Kombinationschemotherapie erhalten haben. Insbesondere diese Patientengruppe unterliegt einer sehr schlechten Prognose und einem mangelhaften Ansprechen gegenüber den etablierten Wirkstoffen. Es besteht ein hoher therapeutischer Bedarf nach wirksameren Therapievarianten, die idealerweise die Resistenz der MSI-high-Tumore umgehen können. Die duale Immuncheckpointinhibition mit Nivolumab plus Ipilimumab bietet gerade deshalb eine sehr wirksame Therapie für diese Patienten im Anwendungsgebiet.

Die EMA hat aufgrund der extrem guten Wirksamkeit der Kombination eine Zulassung auf Basis der Kohorte 2 der nicht randomisierten Studie CheckMate 142 ausgesprochen. Sie kam dabei zu dem Schluss, dass die extrem gute Wirksamkeit, die in der Studie mit einer Mindestbeobachtungszeit von vier Jahren gesehen wurde, im Vergleich zu historischen Daten

nicht durch potenzielle Unterschiede in der Prognose oder im Ansprechen auf die Behandlung, also durch Verzerrung infrage gestellt werden können.

Für die Nutzenbewertung hatten wir somit keine direkt vergleichende Evidenz gegenüber der zweckmäßigen Vergleichstherapie vorliegen. Es war auch kein indirekter Vergleich gegenüber Daten aus klinischen Studien möglich, da der MSI-high-Status in den klinischen Studien nicht erhoben wurde. Wir haben daher bei BMS die Studie CA209-6EP aufgesetzt, in der wir einen indirekten Vergleich der Kohorte 2 der CheckMate 142 mit einer gemäß den Ein- und Ausschlusskriterien passenden Kohorte der Flatiron-Health-Datenbank durchgeführt haben, für die uns patientenindividuelle Daten vorliegen.

Folgende Punkte sind uns besonders wichtig; die möchte ich noch einmal betonen. Der erste Punkt ist die Umsetzung der zVT. Der zweite wichtige Punkt ist die Diskussion potenzieller Confounder und die Strukturgleichheit der Populationen. Der dritte ist die Einordnung der Ergebnisse zum Gesamtüberleben in den Kontext der anderen patientenrelevanten Endpunkte.

Zum ersten Punkt, zur Umsetzung der zVT. Wir hatten in den für das Dossier präspezifizierten Analysen die zVT nicht exakt abgebildet, sondern in der Flatiron-Kohorte alle Therapien als patientenindividuelle Therapie zugelassen, auch solche, die nicht vom G-BA benannt waren. Wir haben für die Stellungnahme die Analysen nochmals korrigiert und nur die Therapieoptionen als patientenindividuelle Therapie zugelassen, die explizit benannt wurden. Ich möchte betonen, dass wir sonst an den geplanten Analysen nichts verändert haben. Es war wirklich nur eine Korrektur der zVT. Durch die nun richtig umgesetzte zVT haben wir insgesamt 119 Patienten in der Flatiron-Kohorte. Beim 1:1-Propensity-Score-Matching finden wir 39 passende Patientenpaare. Wir sehen einen extrem großen Vorteil von Nivolumab plus Ipilimumab beim Gesamtüberleben mit einem Hazard Ratio von 0,24. Die obere Grenze des 95-Prozent-Konfidenzintervalls ist 0,52. Wir sehen diesen Effekt robust über alle durchgeführten Analysemethoden hinweg.

Der zweite wichtige Punkt ist das Confounding. Bei nicht randomisierten Vergleichen kann Confounding die Resultate stark verzerren. Deshalb lag bei uns von Anfang an ein besonderes Augenmerk darauf, relevante Confounder zu identifizieren, um Strukturgleichheit der Population in der Analyse zu ermöglichen und das Gesamtüberleben möglichst unverzerrt schätzen zu können. Wir haben deshalb mit dem G-BA in einem Beratungsgespräch eine Reihe von relevanten Confoundern abgestimmt und haben alle in den Analysen berücksichtigt. Zusätzlich haben wir eine systematische Literatursuche durchgeführt und haben für das Dossier solche als relevante Confounder klassifiziert, die in mehr als drei Literaturstellen berichtet wurden. In der Stellungnahme haben wir – auch auf Anregung des IQWiG hin – alle in der Literatursuche gefundenen potenziellen Confounder in Bezug auf ihre Relevanz und den möglichen Einfluss auf Therapieentscheidungen und Gesamtüberleben diskutiert. Wir sind dabei zu dem Schluss gekommen, dass, um komplett unverzerrt schätzen zu können, in unseren Analysen nur zwei Confounder fehlen. Zehn Confounder inklusive aller der vom G-BA genannten Confounder konnten wir berücksichtigen. Die zwei nicht direkt berücksichtigten Confounder sind mit den zehn berücksichtigten Confoundern assoziiert und somit teilweise mitberücksichtigt. Wir haben das in der Stellungnahme ausführlich dargelegt. Insbesondere in der Analyse des 1:1-Propensity-Score-Matching, wo wir passende Paare aus der CheckMate 142 und der Flatiron-Kohorte bilden, erreichen wir eine sehr gute Strukturgleichheit. Der Vorteil beim Gesamtüberleben, den wir hier unter Berücksichtigung fast aller relevanten Confounder sehen – ich wiederhole: zehn von zwölf Confoundern –, ist so groß, dass er nicht alleine durch Verzerrung, durch unberücksichtigtes Confounding oder ungemessenes Confounding zustande kommen kann. Das haben wir durch weitere Sensitivitätsanalysen in der Stellungnahme gezeigt.

Zu meinem dritten Punkt, zur Einordnung der Ergebnisse in die patientenrelevanten Endpunkte. Der indirekte Vergleich war nur für das Gesamtüberleben möglich. Um jedoch eine

Abwägung der Ergebnisse hinsichtlich der unerwünschten Ereignisse machen zu können, haben wir für die Stellungnahme die in den Fachinformationen und Primärpublikationen berichteten unerwünschten Ereignisse von allen den Therapien, die in der gematchten Kohorte der 39 Patienten im Rahmen der patientenindividuellen Therapie gegeben wurden, den in der Studie CheckMate 142 aufgetretenen unerwünschten Ereignissen gegenübergestellt. Wir konnten keinen indirekten Vergleich mit Propensity-Score-Matching berechnen, da wir keine Daten in der Flatiron-Kohorte zu unerwünschten Ereignissen vorliegen haben.

Wir sehen bei den schweren UEs, insbesondere bei denen mit vergleichbaren Raten im Vergleich zur zVT, die Raten der SUEs sind etwas höher im Vergleich zur zVT. Das deckt sich mit dem, was wir schon in anderen Indikationen bei Nivolumab plus Ipilimumab in diesem Therapieregime gesehen haben. Das Nebenwirkungsprofil ist zudem sehr gut bekannt. Wir haben auch in dieser Studie bei einer medianen Behandlungsdauer von 25 Monaten keine neuen Signale entdeckt. Auch sehen wir in dieser Studie bei der Morbidität und Lebensqualität, wenn wir die Daten über den Zeitverlauf anschauen, dass wir eine stabile Lebensqualität zeigen können.

Wir sehen deshalb zusammenfassend aufgrund des extrem großen Effekts beim Gesamtüberleben mit einem Hazard Ratio von 0,24, der stabilen Morbidität und Lebensqualität und dem bekannten konsistenten Sicherheitsprofil einen nicht quantifizierbaren Zusatznutzen als gerechtfertigt an.

Sehr geehrter Herr Professor Hecken, sehr geehrte Damen und Herren, wir freuen uns nun sehr auf die Diskussion und insbesondere darauf, mögliche noch offene Punkte auszuräumen. – Herzlichen Dank.

Herr Prof. Hecken (Vorsitzender): Herzlichen Dank, Frau Dr. Kupas, für diese Einführung. Sie hatten eben schon ausgeführt, wie sich die in der schriftlichen Stellungnahme vorgelegte Neuberechnung von dem im Dossier vorgelegten Daten unterscheidet. Die Frage an Sie ist damit erledigt. Ich wende mich daher an die Kliniker. Wir haben in der Erstlinie auch eine Zulassung von Pembrolizumab, und deshalb stellt sich mir die Frage: Welchen Stellenwert hat die Kombination aus Nivo plus Ipi in der Therapie des Kolorektalkarzinoms mit Mismatch-Reparatur-Defizienz oder hoher Mikrosatelliteninstabilität, auch vor dem Hintergrund der Zulassung von Pembro in der Erstlinientherapie? Frau Kupas hat eben gesagt, das Nebenwirkungsprofil ist bekannt, das Nebenwirkungsprofil ist beherrschbar. Aber wir haben gesehen, dass wir Therapieabbruchraten von 20 Prozent haben. Insbesondere mit Blick auf die aus meiner Sicht hohen Therapieabbruchraten könnten Sie, Herr Dr. Pox, beginnen.

Herr Dr. Pox (DGVS): Herr Hecken, das ist eine sehr gute Frage. Die Daten von Pembrolizumab sind recht neu. Für die Zukunft ist es so, für die neuen Patienten, bei denen die MSI-Testung in der Leitlinie quasi vorgeschrieben ist, also empfohlen wird, wird wahrscheinlich der erste Standard Pembrolizumab sein. Es wird interessant sein, wie diese Kombination, die hier vorgestellt wird, in der Firstline wirkt, ob sie noch besser ist als Pembrolizumab. Insofern habe ich mir genau dieselbe Frage gestellt. Es geht um Patienten, die schon länger behandelt werden, bei denen nicht primär die MSI-Testung durchgeführt wurde, wie viele das sind. Bei den Patienten, die wir betreuen, ist das verschwindend gering, weil das seit einiger Zeit Standard ist.

Herr Prof. Hecken (Vorsitzender): Danke schön, Herr Pox. – Herr Professor Wörmann.

Herr Prof. Dr. Wörmann (DGHO): Ich kann dazu nicht viel ergänzen. Ein kritischer Punkt ist – das muss man ganz deutlich sagen –: Die MSI-high-Testung ist inzwischen Standard für diese Tumore. Ich glaube, für uns ist ganz gut, dass Pembrolizumab als Erstes bei Ihnen gelandet war. Wir hätten hier sonst ziemlich kritisch gegessen und diskutiert, ob das als Konzept überhaupt funktioniert. Aber dadurch, dass das Pembro-Verfahren in der Erstlinientherapie mit einer randomisierten Studie zuerst gelaufen ist, können wir uns sicher sein, dass die MSI-high, also die Mikrosatelliteninstabilität, ein prädiktiver Faktor für das Ansprechen auf Im-

muncheckpointinhibitoren ist. Ich sage das mit dem Prädiktiv so deutlich, weil ich mitbekommen habe, dass im Anschluss an die Diskussion zu demselben Marker beim Endometriumkarzinom eine Diskussion zwischen prognostischem und prädiktivem Marker aufkam. Es geht hier explizit um prädiktive Marker. Das heißt nicht automatisch, dass es eine schlechte Prognose hat. Wir sind interessiert, eine nicht so günstige Prognose, selbst wenn sie überdurchschnittlich ist, noch besser zu machen.

Der erste Punkt ist: Es ist ein stabiler Marker. Deswegen ist er mit einer gewissen Plausibilität auf die Secondline übertragbar. Der Punkt, den Herr Pox anspricht, ist genau der richtige: Im Moment sind die Patienten, die für diese Therapie qualifiziert sind, Nivo plus Ipi, die, die in der Erstlinientherapie anders behandelt wurden, zum Teil nach zwei oder drei Jahren Rezidive bekommen haben, nach anderer Vorbehandlung, oder einen Progress erhalten. Das sind die, die dafür qualifiziert sind. Ob Nivo plus Ipi nach Pembro-Vorbehandlung noch wirksam ist, ist völlig offen. Dieselbe Diskussion haben wir beim nichtkleinzelligen Lungenkarzinom. Das heißt, im Moment würden wir uns auf die fokussieren, die so behandelt wurden wie in der Zulassungsstudie, das heißt, die eine Nichtimmuncheckpointinhibitortherapie in der Erstlinie bekamen.

Die anderen Daten, Herr Hecken – ich glaube, Sie brauchen uns bald nicht mehr –, hatten Sie genau so zitiert. Auch wir sehen, dass die Abbruchrate höher ist als unter der Monotherapie mit einem Immuncheckpointinhibitor. Mit 20 Prozent liegt sie aber in dem Bereich, den wir von anderen Nivo/Ipi-Studien auch kennen.

Herr Prof. Hecken (Vorsitzender): Herzlichen Dank, Herr Wörmann und Herr Pox. – Herr Kaiser vom IQWiG, bitte schön.

Herr Dr. Kaiser: Vielen Dank, Herr Hecken. – Ich möchte auf die nachgereichten Daten respektive auf die Einleitung eingehen. Man kann zwei Bereiche trennen. Das eine ist die zweckmäßige Vergleichstherapie, die Sie, Frau Kupas, erwähnt haben. Der andere Bereich ist das Thema Confounder. Ich werde auf den Bereich zweckmäßige Vergleichstherapie wenig eingehen. Es ist interessant, zu sehen, was das zum Teil mit der Population gemacht hat: weniger Patienten, aber deutlich mehr in dem Propensity Score Matching. Wie genau das passiert ist, ist mangels einer Darstellung im Ursprungsdossier zu der Propensity-Population nicht ganz beurteilbar. Das ist insofern gar nicht so relevant, weil das ganze Confounder-Problem überhaupt nicht gelöst ist. Sie haben eben gesagt, auf Anregung des IQWiG hin hätten Sie die Dinge diskutiert. Wir haben nicht angeregt, dass Sie das diskutieren, sondern dass die Confounder entsprechend berücksichtigt werden. Was Sie jetzt zum Teil gemacht haben, ist: Sie haben in Ihrer Selektion geschaut, ob Sie Beobachtungsstudien, also weitere Publikationen finden, die weitere Confounder in diesem Indikationsgebiet beschreiben. Sie haben vier gefunden. Das Problem ist nur, dass Ihre ursprüngliche Recherche gar nicht darauf angelegt war, Beobachtungsstudien zu identifizieren. Es hilft also nichts, in der schon ungeeigneten Recherche nachzuselektieren und vier Studien zu finden. Das muss man vollständig machen. Sie haben überhaupt nicht die neueren Publikationen aus der großen SEER-Datenbank in den USA identifiziert. Dort gibt es viele epidemiologische Daten, die verschiedene Zusammenhänge aufzeigen.

Ich würde gerne zwei Nachfragen stellen, weil das mit Blick auf den Umgang mit solchen Daten allgemein relevant ist. Sie haben eine Menge von Confoundern – Sie haben das eben so beschrieben – diskutiert, und dann blieben nicht mehr viele übrig, weil die einen mit den anderen zusammenhängen. So versuche ich das allgemeinverständlich zu beschreiben. Sie haben umfangreiche Abbildungen mit Kausalgraph gemacht. Dahinter steckt die Aussage: Weil wir den Confounder A berücksichtigt haben und der mit B zusammenhängt, braucht man B nicht mehr. Dazu habe ich zwei Fragen. Die erste Frage ist – ich habe keine Literatur gesehen, die Sie zitiert hätten –: Wie haben Sie die Zusammenhänge identifiziert und beschrieben? Die zweite Frage ist ganz konkret. Man kann an einem Beispiel ganz gut illustrieren, dass das so nicht geht. Sie haben zum Beispiel gesagt, den Body-Mass-Index haben Sie

nicht berücksichtigt – das ist übrigens in epidemiologischen Studien ein relevanter Confounder für die Indikation in diesem Stadium –, weil der durch den ECOG-Status schon abgedeckt ist. Der Body-Mass-Index ist ein stetiges Merkmal. Der ECOG-Status hat zwei Ausprägungen, 0 und 1. Meine zweite Frage ist: Wie genau, denken Sie, ist der Body-Mass-Index durch die beiden hier vorhandenen ECOG-Stadien 0 und 1 abgebildet, ohne dass man ihn separat berücksichtigen müsste? Denn Sie sehen eine erhebliche Imbalance in Ihrer Propensity-Kohorte von über 0,5 Standardabweichungen, was den Body-Mass-Index angeht.

Herr Prof. Hecken (Vorsitzender): Danke schön, Herr Kaiser. – Wer möchte für den pU antworten? Oder wer muss? – Frau Kupas.

Frau Dr. Kupas (BMS): Ich starte einmal. Es waren relativ viele Punkte. Ich fange bei der zVT an. Was hat sich geändert, und warum finden wir mehr Paare in einer gematchten Kohorte, als wir vorher gefunden haben? Wir haben Patienten, die vielleicht vorher eine Medikation bekommen haben, die nicht der zVT entsprochen hat, aber in einer späteren Linie eine zVT-konforme Medikation bekommen haben, in unsere Kohorte einschließen können, aber in einer etwas späteren Linie. Wir sehen in der 142, dass die Patienten tendenziell in einer etwas späteren Linie sind als in der Flatiron-Kohorte, und finden dadurch mehr passende Paare. Dadurch, dass die Linien jetzt etwas besser zusammenpassen, haben wir 39 statt 29 Paare gefunden und sehen die robusten Ergebnisse, die wir zu allen anderen Methoden auch sehen, mit dem großen OS-Effekt, weil die Patienten jetzt einfach besser zueinander passen. Das ist das, was sich durch die zVT geändert hat. Wir haben in der Stellungnahme die Baseline-Charakteristika dieser Populationen eingereicht, sodass man sehen kann, was es für eine Population ist, die wir hier beschreiben.

Zu Ihrem zweiten Punkt, zur Literatursuche. Wir haben eine systematische Suche gemacht und haben Literaturstellen mit einem Ausschlussgrund ausgeschlossen, der sich auf Beobachtungsstudien bezog, und haben im Zuge der Stellungnahme alle diese Literaturstellen noch einmal durchgeschaut und haben Confounder gefunden. Es ist aber kein weiterer Confounder aufgetaucht, der nicht schon vorher in der alten Literatur so vorhanden gewesen ist. Wir haben jetzt in der Stellungnahme alle Confounder, die in allen diesen Literaturquellen aufgetaucht sind, die in dieser Literatursuche erschienen sind, auf die Relevanz hin diskutiert. Dazu haben wir medizinische Experten zu Rate gezogen und auch unsere Medizinabteilung gefragt: Wie sind die Zusammenhänge zwischen den einzelnen Faktoren? Was ist relevant für das Gesamtüberleben? Was ist relevant für die Therapieentscheidung? Wir sind so zu dem kausalen Graphen gekommen.

Zu Ihrer Frage nach dem Zusammenhang zwischen BMI und ECOG. Was wir mitgenommen haben, ist, dass das, was beim BMI ausschlaggebend ist, die Komorbiditäten sind, die darüber entstehen können, und die hängen mit dem ECOG zusammen. So ist diese Verbindung hergestellt. Wir sehen in den Baseline-Charakteristika aber auch, dass der BMI zwischen den beiden Kohorten nicht 100-prozentig zusammenpasst, aber die zVT-Optionen werden häufig gewichtsbasiert oder körperoberflächenbasiert gegeben. Das heißt, die Patienten sind entsprechend ihres Gewichts, ihres BMIs behandelt worden. Wir haben das deswegen nicht als so sehr relevant klassifiziert, dass es als separater Confounder in die Analysen hätte eingehen müssen. So sind wir vorgegangen. Wir haben uns viel medizinisches Expertenwissen geholt und haben nicht systematisch für jeden Confounder eine Literatursuche gemacht, wo wir den Einfluss untersucht haben.

Ich hoffe, ich habe Ihre Fragen beantworten können. Ich hoffe, ich habe nichts vergessen. Sie haben relativ viel aufgebracht. Ansonsten bitte gern noch einmal nachfragen.

Herr Prof. Hecken (Vorsitzender): Danke schön, Frau Kupas. – Herr Kaiser.

Herr Dr. Kaiser: Ich hatte ganz explizit zwei Fragen gestellt; dazu komme ich gleich noch. Zur Vergleichstherapie. Man weiß nicht, wie das im Vergleich zu der vorhergehenden Dossierkohorte war, weil Sie die Basisdaten nicht beschrieben hatten. Man kann nicht beurteilen, was

genau passiert ist. Es ist schon interessant, dass Sie spätere Therapielinien drin haben – das habe ich verstanden –, weil die späteren nach zVT die relevanteren waren, die Patienten eigentlich älter sein müssten. Das mittlere Alter ist in dieser Kohorte aber deutlich zurückgegangen.

Schauen wir uns den Punkt an, den Sie gerade beschrieben haben. Sie haben diese Studientypen angeschaut. Ich versuche das zu illustrieren, weil nicht ganz eindeutig war, was ich meine. Ich versuche zu illustrieren, was das Problem ist. Verstehen sie bitte nicht falsch, dass ich ein allgemeinverständliches Beispiel nehme. Mir geht es lediglich um die Verständlichkeit. Sie gehen in einen Supermarkt, in dem gibt es ausschließlich Reis, und Sie kaufen Reis. Dann sagen wir, das IQWiG: Eigentlich braucht man für dieses Gericht Reis und Nudeln. Sie gehen in den gleichen Supermarkt, der keine Nudeln führt, und sagen: Wir finden keine Nudeln bzw. zwei Nudeln unter irgendeinem Regal. Das ist nicht richtig. Sie müssten in einen Supermarkt gehen, der sowohl Reis als auch Nudeln führt. Das Problem Ihrer Recherchestrategie ist: Sie haben nicht nach Beobachtungsstudien gesucht. Deswegen macht es keinen Sinn, allein in den ausgeschlossenen Publikationen nach Beobachtungsstudien zu suchen. Ich erwähne noch einmal: Es gibt die große epidemiologische Datenbank in den USA, wo in den letzten fünf bis zehn Jahren ganz viele Publikationen erschienen sind. Die haben Sie alle nicht drin.

Vielleicht noch einmal zum BMI. Zum BMI gibt es gute Publikationen, die wie in anderen Indikationen auch eine U-Kurve des Zusammenhangs von BMI und Mortalitätsrisiko zeigen. Das können Sie mit dem ECOG-Status gar nicht abbilden. Das ist das, was Sie in der ergänzenden Tabelle beschreiben: Dass der BMI nicht abgebildet werden muss, weil er bereits durch den ECOG-Status abgebildet ist. Sie werden diesen Zusammenhang so nicht finden. Im Gegensatz zu dem, was Sie gerade gesagt haben, haben Sie einen drastischen Unterschied beim BMI zwischen der Flatiron-Kohorte und der Nivolumab-Kohorte von über einer halben Standardabweichung und das deutlich höhere Risiko in der Flatiron-Kohorte ausweislich des Mittelwerts und des Medians. Es gibt noch drei, vier andere unbalancierte Dinge. Diese Situation hatten wir auch bei Dostarlimab. Da hat man noch ein paar Sachen adjustiert. Aber viele Dinge sind unbalanciert geblieben. Das ist hier auch der Fall. Deswegen löst das, was Sie gemacht haben, das Problem der Kohorten nicht.

Herr Prof. Hecken (Vorsitzender): Danke schön, Herr Kaiser. – Replik des pharmazeutischen Unternehmers? – Frau Kupas.

Frau Dr. Kupas (BMS): Danke für das schöne Beispiel mit dem Supermarkt. Ich glaube, das war sehr illustrativ. Genau so war es aber nicht. Dieses Beispiel kann man hier nicht heranziehen, um das abzudecken, was wir gemacht haben. Wir haben mit einem Beratungsgespräch angefangen und haben mit dem G-BA besprochen, welche Confounder für dieses Anwendungsgebiet relevant sind. Für alle diese haben wir gematcht. Der Supermarkt hat alle diese Confounder geführt, und wir konnten das Matching machen. Wir hatten Reis und Nudeln, wir hatten vielleicht keine Vollkornnudeln. Bleiben wir bei diesem Beispiel. Wir haben mit Hilfe der Experten zwei Confounder identifiziert, die noch relevant gewesen wären, weil sie auf Therapieentscheidungen und auf Outcome Einfluss nehmen, haben für zehn Confounder Strukturgleichheit hergestellt und haben einen extrem großen Effekt, wo wir über Analysen gezeigt haben, dass er sich durch die Verzerrung der fehlenden Confounder nicht erklären lässt. Für den BMI haben wir nicht gematcht; das ist richtig. Wir sehen einen BMI von 24 vs. 26 in der gematchten Kohorte, wenn ich es richtig im Kopf habe; ich habe die Zahlen nicht 100-prozentig vor mir. Der Unterschied ist da. Er ist vielleicht nicht so riesig. Wir haben für zehn Confounder justiert. Wir haben gesagt: Der BMI ist dann relevant, wenn es Komorbiditäten gibt, und die sind über den ECOG abgedeckt. Wir sehen diesen wirklich sehr großen Effekt. Für zehn Confounder haben wir Strukturgleichheit. Zusammen mit dem sehr großen Effekt lässt uns das zu dem Schluss kommen, dass dieser Effekt sich nicht alleine durch Verzerrung erklären lässt, auch wenn wir ein ungemessenes Confounding drin haben. Es wird sowieso keine RCT; ich glaube, darin sind wir beide uns einig. Auch wenn wir ein un-

gemessenes Confounding drin haben, wird dieser Effekt so groß sein und sich nicht alleine dadurch erklären.

Über den BMI kann man sich streiten. Den hätten wir in die Analyse aufnehmen können; wir haben ja die Daten. Wir haben hergeleitet, dass der Einfluss nicht so groß ist und dass er zusätzlich zu den zehn Confoundern nicht mit hinein muss. Ich glaube, dass unsere Suche schon vollständig war. Wir haben sehr viele Confounder diskutiert. Wir haben den Einfluss diskutiert. Wir haben uns Expertenwissen dazugeholt. Ich schaue gerne noch in die Literatur, die Sie genannt haben, ob da etwas enthalten ist, was wir in unserer riesigen Liste noch nicht erfasst haben. Es waren extrem viele Confounder, die wir in der Stellungnahme diskutiert haben. Der erste Punkt ist, dass wir vorher mit dem G-BA gesprochen haben, welche Confounder wichtig sind. Für die haben wir gematcht. Diesen Punkt sollten wir voranstellen. Ich glaube, dass die Analyse robust genug ist, dass wir einen nicht quantifizierbaren Zusatznutzen ableiten können, natürlich mit der Unsicherheit, die wir immer noch haben, weil wir niemals für alle Confounder matchen können. Es kann an dieser Stelle auch unbekannte Confounder geben.

Herr Prof. Hecken (Vorsitzender): Danke schön, Frau Kupas. – Frau Müller, Kassenärztliche Bundesvereinigung.

Frau Dr. Müller: Danke, Herr Professor Hecken. – Ich habe eine kurze Nachfrage dazu an Sie, Frau Kupas. Ich konnte der Argumentation mit dem BMI, mit dem es anscheinend eine deutliche Imbalance gibt, nicht ganz folgen. Sie sagten, das wäre etwas, was durch den ECOG wegen der Begleiterkrankung abgedeckt ist. Ich denke, nicht jede Begleiterkrankung führt zu einem verminderten ECOG, sondern erst in einem etwas fortgeschrittenerem Stadium, kann aber trotzdem Risikofaktor sein. Zum anderen hat Herr Kaiser darauf hingewiesen, dass es eine U-förmige Kurve ist, dass sozusagen sowohl ein sehr niedriger als auch ein sehr hoher BMI kritisch sind. Der niedrige BMI geht regelhaft nicht unbedingt mit Begleiterkrankungen einher, schon gar nicht mit solchen, die negative Auswirkungen prima vista auf den Allgemeinzustand haben. Vielleicht könnten Sie das noch einmal erläutern.

Herr Prof. Hecken (Vorsitzender): Danke schön, Frau Müller. – Frau Kupas.

Frau Dr. Kupas (BMS): Ich versuche es noch einmal. Das mit der U-Kurve ist richtig. Wir sehen in den Baseline-Charakteristika, dass die Patienten im Median einen BMI von 24 und 26 haben. Die ganz kleinen BMI und der Einfluss durch Untergewicht ist an dieser Stelle nicht relevant. Unsere Argumentation war, dass, wenn es durch einen zu hohen BMI Komorbiditäten gibt, sie sich im Gesundheitszustand widerspiegeln, der durch den ECOG abgedeckt ist. Für diesen haben wir die Kohorten angeglichen. Wir haben auch die Baseline-Charakteristika für BMI und für Körperoberfläche in den beiden Kohorten dargestellt. Es geht nicht so weit auseinander. Auch werden die Therapien, die in der Flatiron-Kohorte gegeben werden, häufig gewichtsbasiert oder körperoberflächenbasiert gegeben. Das heißt, die Patienten sind adäquat therapiert. Wir haben keinen Einfluss, der durch eine Dosis der zVT zustande kommen kann. Wir haben deswegen gesagt, dass das Matching über den ECOG ausreicht, um den Komorbiditätenaspekt abzudecken, den wir sehen. Der Gewichtsaspekt, der durch die Dosierung zustande kommt, ist durch die körperrgewichtsbasierte oder körperoberflächenbasierte Therapie abgedeckt. Das war die Argumentation. Ich hoffe, es ist klarer geworden, was ich meinte.

Herr Prof. Hecken (Vorsitzender): Danke schön, Frau Kupas. – Herr Professor Wörmann.

Herr Prof. Dr. Wörmann (DGHO): Ganz kurz ein Punkt, damit sich das nicht beim BMI festbeißt. BMI ist über viele Jahre ein interessanter Punkt gewesen, weil er relativ leicht messbar war. Klinisch geht es bei diesen Patienten tendenziell um die Kachexie aufgrund der fortgeschrittenen Erkrankung. Was bei uns dabei im Vordergrund steht, ist die Sarkopenie, also der Verlust an Muskelmasse. Der korreliert eher mit einem ungünstigen klinischen Verlauf. Dafür gibt es bessere Liquidmarker. Ich will nicht sagen, dass BMI nicht greift. Für die methodische

Diskussion, die Herr Kaiser angefangen hat, passt es. Ich möchte nur nicht, dass es zu sehr eine Anhörung über einen Parameter gibt, den wir in Zukunft nicht mehr verwenden werden.

Herr Prof. Hecken (Vorsitzender): Und das noch im Zusammenhang mit Nudeln, Herr Wörmann. Da schließt sich der Kreis.

Herr Prof. Dr. Wörmann (DGHO): Es ist der kritische Punkt, ob wir bei den Preisen nicht eher über KaDeWe als über einen Supermarkt reden.

Herr Prof. Hecken (Vorsitzender): Das nehmen wir so hin, ohne darüber zu diskutieren. – Herr Kaiser, IQWiG.

Herr Dr. Kaiser: In die Delikatessenabteilung von Karstadt und KaDeWe schauen wir jetzt nicht hinein, um zu sehen, was da jeweils zu haben ist. – Ich habe den BMI gar nicht angesprochen, um zu beschreiben, wenn der BMI berücksichtigt wäre, wäre alles gut. Ich habe das vielmehr zur Illustration genommen, um Ihnen aufzuzeigen, dass das, was Sie gemacht haben, erstens einen nicht literaturgestützten kausalen Graphen, um die Dinge wegzureden, zweitens zu beschreiben, der BMI hängt mit dem ECOG zusammen, weswegen man ihn nicht berücksichtigen muss, kein adäquates Vorgehen ist. Darum geht es mir. Das kann man mit den anderen Confoundern auch durchdeklinieren. Sie sagen zum Beispiel, Sie haben nicht nach Regionen adjustiert, weil die Flatiron-Datenbank nur USA-Daten hat. Das ist kein Argument. Dann muss man aus der Nivolumab-Studie nur die USA-Daten nehmen. Wenn Sie sich die USA-Daten aus der Nivolumab-Studie anschauen, sehen Sie, dass Sie eine um 10 Prozentpunkte höhere Mortalität in den USA als in Europa haben. Aber jetzt bringen Sie Europadaten mit den Vergleichsdaten aus den USA zusammen. Es sind also viele verschiedene Dinge.

Ich habe noch eine Frage zum grundsätzlichen Vorgehen. Sie haben beschrieben, Sie haben die Einschlusskriterien der Nivolumab-Studie auf die Flatiron-Kohorte angewandt. Das ist so nicht richtig. Sie haben einzelne Einschlusskriterien angewandt. Dazu gehören Ein- und Ausschlusskriterien. Sie haben zum Beispiel in der Nivolumab-Studie, so wie das ganz oft bei klinischen Studien ist, als Ausschlusskriterium genommen, dass eine Nebenwirkung einer vorherigen Therapie vollständig zurückgegangen sein muss, mit ein paar Ausnahmen, Neuropathie, und Sie haben eine große Latte von Laborparametern, die im Normbereich sein müssen. Das sind die Neutrophilen, das sind Nierenwerte, das sind Leberwerte etc. Warum haben Sie das nicht in der Flatiron-Datenbank gemacht? Denn dadurch haben Sie eine Positivselektion in der Nivolumab-Kohorte, dass diejenigen, die aus der vorhergehenden Therapie irgendwelche Probleme mitgenommen haben, nicht in die Studie hineinkommen. Ein Indiz dafür sind die Daten zur Zeit der Beendigung der letzten Vortherapie bis zur Gabe von Nivolumab bzw. Vergleichstherapie. 100 Prozent in der Flatiron-Kohorte haben diesen Wechsel innerhalb von drei Monaten gehabt. In Ihrer Kohorte sind das nur 50 Prozent. Bis zu über sechs Monate war hier die Latenzzeit zwischen vorhergehender Therapie und Nivolumab-Therapie, was vermutlich an den für die klinischen Studien völlig nachvollziehbaren Kriterien liegt, dass die Nebenwirkungen aufgelöst sein müssen, dass Sie Normwerte in den Laborparametern haben. Aber warum haben Sie das nicht bei der Flatiron-Kohorte angewandt? Es ist ein ganz wesentlicher Selektionsunterschied, den Sie hier machen.

Herr Prof. Hecken (Vorsitzender): Danke schön, Herr Kaiser. – Wer möchte antworten?

Frau Dr. Kupas (BMS): Wir sind nicht sicher, ob wir die Laborwerte der Patienten in der Flatiron-Kohorte von Therapie zu Therapie haben. Ich glaube, wir sind uns einig, wir sind in dem indirekten Vergleich, nicht in einer randomisierten kontrollierten Studie, und wir werden diese Qualität nicht herstellen können. Wir haben nach den Ein- und Ausschlusskriterien, die unserer Ansicht nach relevant waren, selektiert, haben die Patienten gematcht. Wir haben 39 passende Paare gefunden. Wir haben zehn Confounder berücksichtigt. Darunter sind alle, die vom G-BA genannt wurden. Wir haben einen wirklich extrem großen Effekt. Wir haben Sensitivitätsanalysen gemacht und uns angeschaut: Welchen Einfluss haben Imbalancen zwi-

schen den Kohorten auf das Hazard Ratio, das wir am Ende sehen? Bzw.: Es war ein relatives Risiko. Wir haben gesehen, es müssen wirklich extrem große Imbalancen, gepaart mit einem extrem großen Effekt sein, damit dieser Effekt infrage gestellt werden kann. Wir sind der Meinung, dass dieser Effekt extrem groß ist. Wir haben Strukturgleichheit für viele wichtige Faktoren hergestellt. Wir haben nicht alle; das ist so. Wir denken aber, in diesem Kontext ist „nicht quantifizierbar“ gerechtfertigt.

Herr Prof. Hecken (Vorsitzender): Danke schön, Frau Kupas. – Herr Hastedt, GKV-SV.

Herr Dr. Hastedt: Guten Tag! Es wurde hinlänglich diskutiert. Das IQWiG sieht große Probleme bei dem indirekten Vergleich. Wann ist mit ersten Ergebnissen der im EPAR erwähnten RCT, dieser CA209-8HW, zu rechnen? Eine Frage an den pharmazeutischen Unternehmer: Ist diese RCT Ihrer Einschätzung nach auch für die Nutzenbewertung geeignet? Wenn ich es richtig verstanden habe, ist nur ein Teil der Studienpopulation in der zweiten Linie. Aber die EMA wollte die Studie in diesem Zusammenhang sehen. Was können wir wann erwarten?

Herr Prof. Hecken (Vorsitzender): Danke schön, Herr Hastedt. – Gibt es schon einen Zeithorizont? – Frau Amann.

Frau Dr. Amann (BMS): Die Studie rekrutiert aktuell noch. Es ist die CheckMate-8HW-Studie. Das Studienende wird mit 2025 erwartet. Wie Sie gerade schon erwähnt haben, rekrutieren in dieser Studie alle Linien – es ist quasi ab der ersten Linie –, fast, nicht vollständig, quasi die Patienten, die im hier besprochenen Anwendungsgebiet sind.

Herr Prof. Hecken (Vorsitzender): Danke schön. – Herr Hastedt, ist das okay, oder haben Sie eine Nachfrage?

Herr Dr. Hastedt: Sind Interimsanalysen geplant? Sie sagten: Studienende 2025.

Frau Dr. Amann (BMS): Es sind natürlich Interimsanalysen geplant. Zum jetzigen Zeitpunkt liegen uns allerdings noch keine Daten aus Interimsanalysen vor.

Herr Prof. Hecken (Vorsitzender): Danke schön. – Herr Hastedt, okay?

Herr Dr. Hastedt: Für wann sind die Interimsanalysen geplant?

Frau Dr. Amann (BMS): Wir planen, erste Daten Ende 2022/Anfang 2023 zu erhalten.

Herr Prof. Hecken (Vorsitzender): Eine weitere Frage, Herr Hastedt?

(Herr Dr. Hastedt: Nein, danke, das wäre es!)

Gibt es weitere Fragen, Anmerkungen, Hinweise? – Ich sehe keine mehr. Dann würde ich Ihnen, Frau Dr. Kupas, die Gelegenheit geben, sofern Sie das möchten, die wesentlichen Punkte aus den letzten 40 Minuten in den Fokus zu rücken. Bitte schön, Frau Kupas.

Frau Dr. Kupas (BMS): Danke, Herr Professor Hecken. – Ich werde mich kurzhalten. Ich habe vieles eben schon gesagt. Ein paar Punkte sind mir noch wichtig. Wir haben einen indirekten Vergleich gegenüber der Flatiron-Health-Kohorte gemacht, wo wir präspezifizierte Analysen und eine vorgeplante Liste an Confoundern berücksichtigt haben. Wir hatten für die Stellungnahme die zVT korrigiert. Wir haben eine gematchte Kohorte, wo zehn relevante Confounder eingehen, wo wir 39 passende Paare haben, wo wir eine sehr gute Strukturgleichheit sehen, wo wir riesige Effekte sehen. Ich wiederhole: Das Hazard Ratio liegt bei 0,24, die obere KI-Grenze ist 0,52. Wir sind der Meinung, dass dieser Effekt nicht alleine durch Verzerrung zustande kommen kann, insbesondere weil wir schon für zehn Confounder justiert haben. Natürlich ist der Effekt nicht unverzerrt. Das haben wir auch nicht behauptet. Das möchte ich noch einmal betonen. Wir denken, dass ein nicht quantifizierbarer Zusatznutzen gerechtfertigt ist.

Danke für die Diskussion. Wir hoffen, dass wir zumindest ein paar Unklarheiten ausräumen konnten. Wir wissen um die Limitation des Vergleichs natürlich auch. Wir denken aber, der riesige Effekt ist relevant.

Herr Prof. Hecken (Vorsitzender): Danke schön, Frau Dr. Kupas. – Herzlichen Dank allen, die hier Fragen gestellt oder Antworten gegeben haben. Wir werden das selbstverständlich in die Diskussion einbeziehen.

Vor diesem Hintergrund können wir diese Anhörung beenden. Einen schönen Tag an alle, die uns jetzt verlassen!

Schluss der Anhörung: 14:13 Uhr