

# Mündliche Anhörung



**Gemeinsamer  
Bundesausschuss**

gemäß § 35 a Abs. 3 Satz 2 SGB V

**hier: Wirkstoff Indacaterolacetat/Glycopyrroniumbro-  
mid/Mometasonfuroat (D-571)**

Videokonferenz des Gemeinsamen Bundesausschusses in Berlin  
am 21. Dezember 2020  
von 11:00 Uhr bis 12:02 Uhr

– Stenografisches Wortprotokoll –

Angemeldete Teilnehmer der Firma **Novartis Pharma GmbH:**

Frau Dr. Eichinger  
Herr Dr. Diwischek  
Herr Dr. Wasmuth  
Herr Dr. Berschneider

Angemeldete Teilnehmer der Firma **Chiesi GmbH:**

Frau Dr. Reimann  
Herr Dr. Rellin

Angemeldete Teilnehmer der Firma **GlaxoSmithKline GmbH & Co. KG:**

Frau Wolf  
Frau Hermsen

Angemeldeter Teilnehmer der **Deutschen Gesellschaft für Pneumologie und Beatmungsmedizin (DGP):**

Herr Prof. Taube

Angemeldeter Teilnehmer des **Pneumologisch/onkologisch/internistischen Studienzentrums POIS Leipzig GbR:**

Herr PD Dr. Geßner

Angemeldeter Teilnehmer der **Universitätsmedizin Rostock:**

Herr Prof. Dr. Lommatzsch

Angemeldeter Teilnehmer des **Verbandes forschender Arzneimittelhersteller e. V. (vfa):**

Herr Dr. Rasch

Beginn der Anhörung: 11:00 Uhr

**Herr Prof. Hecken (Vorsitzender):** Meine sehr verehrten Damen und Herren! Herzlich willkommen im Unterausschuss Arzneimittel des Gemeinsamen Bundesausschusses. Wir haben heute Anhörungen, aktuell das Dossier 571, zu dem Stellung genommen haben zum einen Novartis als pharmazeutischer Unternehmer, die Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin, klinische Sachverständige, Herr Privatdozent Dr. Geßner vom POIS Leipzig, Herr Professor Lommatzsch von der Uni-medizin Rostock, Chiesi GmbH, GlaxoSmithKline GmbH & Co. KG und der vfa. Basis der heutigen Anhörung ist die Dossierbewertung des IQWiG vom 12. November dieses Jahres.

Ich muss zunächst, da wir wieder ein Wortprotokoll führen, die Anwesenheit feststellen. Für Novartis Pharma GmbH müssten Frau Dr. Eichinger, Herr Dr. Diwischek, Herr Dr. Wasmuth und Herr Dr. Berschneider da sein, dann müsste Herr Professor Taube von der DGP zugeschaltet sein, Herr Dr. Geßner vom POIS Leipzig. Herr Professor Lommatzsch fehlt wahrscheinlich, ich frage aber noch mal: Herr Lommatzsch? Nein, er ist nicht da, er hatte auch keine Rückmeldung gegeben. Dann müssten da sein Frau Dr. Reimann und Herr Dr. Rellin von Chiesi, Frau Wolf und Frau Hermsen von GlaxoSmithKline und Herr Dr. Rasch vom vfa. – Ist sonst noch jemand da, der nicht aufgerufen wurde? – Ich sehe keinen.

Dann begrüße ich Sie nochmals alle ganz herzlich. – Ich würde zunächst dem pharmazeutischen Unternehmer die Möglichkeit geben, aus seiner Sicht die wesentlichen Punkte zum Wirkstoff oder zur Kombination und insbesondere zur Dossierbewertung des IQWiG vom 12. November dieses Jahres darzustellen. Danach würden wir in die übliche Frage-und-Antwort-Runde eintreten. Wer macht das für den pharmazeutischen Unternehmer?

**Herr Dr. Wasmuth (Novartis):** Das mache ich, Herr Hecken.

**Herr Prof. Hecken (Vorsitzender):** Bitte, Herr Wasmuth, Sie haben das Wort.

**Herr Dr. Wasmuth (Novartis):** Vielen Dank. – Sehr geehrter Herr Professor Hecken! Sehr geehrte Damen und Herren! Vielen Dank für die Einladung zur heutigen Anhörung zur Dreifachkombination aus Indacaterol, Glycopyrronium und Mometason. Zu Beginn möchte ich zunächst die heutigen Teilnehmer für Novartis vorstellen. Für die Statistik ist heute Birte Eichinger dabei, Herr Korbinian Berschneider vertritt die Medizin, Florian Diwischek die Fragen rund um die Epidemiologie und Gesundheitsökonomie, und mein Name ist Timo Wasmuth; ich bin mit meinem Team verantwortlich für die Dossiers bei Novartis.

Der G-BA bewertet heute das Arzneimittel Enerzair, eine inhalative Dreifachkombination aus Indacaterol, Glycopyrronium und Mometason für Patienten mit schwerem Asthma. Wir alle sind sicher mit der Indikation Asthma vertraut. Ich möchte nur zwei Dinge zu Beginn hervorheben, auf der einen Seite die Patienten im Anwendungsgebiet. Das sind Patienten, die hochdosierte Standardmedikation nehmen. Sie haben deutliche Einschränkungen in ihrer Lebensqualität, und sie haben alle eine Vorgeschichte mit Exazerbationen. Auf der anderen Seite die Studie, über die wir heute reden müssen: die ARGON-Studie. Die ARGON-Studie hatte die Lebensqualität als primären Endpunkt, und sie wurde in kontinuierlicher Beratung mit dem Gemeinsamen Bundesausschuss als AMNOG-Studie aufgesetzt, also zusätzlich zu den Zulassungsstudien. Lassen Sie mich Bezug auf den Bericht des IQWiG nehmen. Ich möchte hier auf drei Punkte eingehen; zunächst als ersten Punkt die klinischen Relevanzschwellen und damit die Berücksichtigung der Verbesserung der Lebensqualität der Patienten; als zweiten Punkt den Evidenztransfer, den wir angewandt haben, und als dritten Punkt die Verbesserung der Symptomatik der Patienten.

Lassen Sie mich zum ersten Punkt kommen, den Ergebnissen zur Verbesserung der Lebensqualität und den damit verbundenen klinischen Relevanzschwellen: Wir haben hier sicher einen besonderen Fall. Die Lebensqualität wurde nämlich sogar als primärer Endpunkt über zwei Fragebögen erhoben, über den AQLQ und den SGRQ. Die Lebensqualitätsendpunkte waren auch Gegenstand der Beratung mit dem G-BA. Die Frage, die jetzt aufgeworfen ist: Ab welchem Schwellenwert zeigt sich nun die Relevanz für die Patienten? Uns sind hier drei Aspekte wichtig. Zum einen sind die im Dossier verwendeten Schwellenwerte in der Literatur beschrieben und auch validiert. In anderen Verfahren zur frühen Nutzenbewertung hat der G-BA diese Endpunkte und die damit verbundenen Schwellen bereits akzeptiert. Und: Für die Auswertung in diesem Dossier waren die entsprechenden Schwellenwerte präspezifiziert. Im Ergebnis zeigt sich, dass beim AQLQ signifikant mehr Patienten unter der Gabe von Enerzair eine Verbesserung in ihrer Lebensqualität erfuhren; beim SGRQ ist der Effekt sogar noch deutlicher.

Ich komme zum zweiten Punkt, dem Evidenztransfer. Blickt man auf die gesamte Studie, sieht man zunächst einen signifikanten Vorteil für Patienten mit Enerzair. Die Zulassung erfolgte jedoch fokussierter. Deshalb haben wir auch die relevante Teilpopulation im Dossier dargestellt. Im Kern stellt sich nun die Frage, ob die Ergebnisse der Gesamtpopulation auch auf die Teilpopulation übertragen werden können. ... (akustisch unverständlich) hat gezeigt, dass dies möglich ist. Zunächst die Grundvoraussetzungen: Wir brauchen signifikante und relevante Vorteile in der Gesamtpopulation; diese liegen vor. Zudem: Die Teilpopulation und die Gesamtpopulation sind hinsichtlich der Patientenmerkmale vergleichbar. Hier gab es noch offene Fragen; die wir mit der Stellungnahme nachgereicht haben. Besonders wichtig: die Effektrichtung. Die Effektrichtung ist jeweils gleich, teils sind die Effekte sogar numerisch identisch. Zusammengefasst: Die Kriterien für einen Evidenztransfer sind in diesem Fall voll erfüllt.

Ein dritter Punkt ist uns noch wichtig. Wir haben von Klinikern gelernt, dass für Patienten mit schwerem Asthma jeder Tag ohne Beschwerden ein gewonnener Tag ist. In der ARGON-Studie führten die Patienten systematisch Tagebücher zur Dokumentation ihrer Symptome, zum Beispiel zu Hustenanfällen oder zur beschwerlichen Kurzatmigkeit. Im Ergebnis zeigt sich: Die Anzahl der symptomfreien Tage, also der gewonnenen Tage, war bei Patienten mit Enerzair signifikant höher als bei Patienten mit der Standardtherapie. Wir haben weitere Informationen zu den Auswertungen nachgereicht.

Lassen Sie mich zusammenfassen: Drei Endpunkte zeigen den Zusatznutzen von Enerzair: der AQLQ, der SGRQ – beides für Vorteile in der Lebensqualität – und die gewonnenen Tage ohne Beschwerden. Die Voraussetzungen für einen Evidenztransfer sind bei Enerzair erfüllt. Die Relevanzschwellen bei der Lebensqualität sind adäquat, damit also auch der Zusatznutzen für die Patienten mit stark symptomatischem unkontrolliertem Asthma. – Vielen Dank.

**Herr Prof. Hecken (Vorsitzender):** Ganz herzlichen Dank, Herr Dr. Wasmuth, für diese Einführung. – Meine erste Frage, bevor ich das Fragerecht an die Bänke und die PatV gebe, geht an die beiden Kliniker. Herr Wasmuth hat gerade über die Relevanz der Schwellenwerte zur Lebensqualität ausgeführt und gesagt, aus Sicht des pharmazeutischen Unternehmers seien die hier gezeigten Veränderungen für den Patienten relevant. Ich habe eine ganz banale Frage an Sie beide: Könnten Sie uns aus klinischer Sicht einmal verdeutlichen, was eine Verbesserung um 0,5 Punkte beim AQLQ – der einen Wertebereich von 1 bis 7 hat – und um 4 Punkte beim SGRQ – der einen Wertebereich von 0 bis 100 hat –, ganz konkret für den Patienten bedeutet? Herr Wasmuth hat noch auf die Patiententagebücher rekurriert, wo es um die Anzahl beschwerdefreier Tage geht. Aber mir fehlt jetzt ein wenig aus der Praxis das Vorstellungsvermögen, was 0,5 beim AQLQ und 4 Punkte beim SGRQ bedeuten. Vielleicht fangen wir mit Herrn Taube an und danach Herrn Geßner. – Herr Professor Taube, könnten Sie dazu etwas sagen?

**Herr Prof. Taube (DGP):** Ja, dazu kann ich etwas sagen. – Das Problem dabei ist, dass das vom Ausgangswert abhängt. Das sind Fragebögen, die bei Patienten mit obstruktiven Atemwegserkrankungen, also Verengung der Atemwege, eingesetzt werden. Es ist so, dass dort Veränderungen bei Patienten mit sehr schwerer Erkrankung, aber auch bei Patienten mit leichter Erkrankung nachweisbar sind. Dann können 4 Punkte etwas ganz anderes bedeuten. Ich glaube, der wichtige Punkt dabei ist, dass das wirklich Instrumente sind, die lange schon in der Pneumologie etabliert und auch sehr gut validiert sind. Die Validierungsstudien haben 2000 stattgefunden. Ganz besonders für den SGRQ ist das ein Instrument, das sehr häufig in pharmakologischen Studien, aber auch in nicht pharmakologischen Studien eingesetzt wird und wo es wirklich eine gute Menge an Literatur gibt, die ganz klar sagt: Die minimal clinically important difference, also der Unterschied, den der Patient noch merkt, liegt bei 4 Punkten. Das ist etwas, was international in den Fachgesellschaften und auch von anderen Behörden, wie zum Beispiel der FDA, ganz klar als Endpunkt anerkannt ist.

Deshalb war ein Kritikpunkt, den wir als Fachgesellschaft gemacht haben, dass sozusagen jetzt auf eigenen Untersuchungen des IQWiG, die sich aber nicht spezifisch mit diesen Fragebögen beschäftigen, prinzipiell gesagt wird: Okay, der Unterschied muss bei 15 Prozent von der Spannweite liegen. Wie gesagt, insbesondere für den SGRQ gibt es sehr viel Literatur, die ganz klar belegt: 4 Punkte sind ein Unterschied, den der Patient merkt. Das ist die minimal clinically important difference. Das ist nicht der maximale Unterschied, sondern das, was der Patient sozusagen gerade noch merkt. Aber es ist – jetzt in die Klinik übertragen – ein Unterschied, ob Sie jemanden haben, der sauerstoffpflichtig ist und dann mit einem Rollator herumläuft, oder jemanden, der noch einigermaßen gut körperlich aktiv ist und im Leben steht.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Herr Professor Taube. – Vielleicht ergänzend Herr Dr. Geßner. Dann habe ich Herrn Berschneider vom pU dazu und Frau Wieseler. – Herr Geßner.

**Herr Dr. Geßner (POIS Leipzig):** Vielen Dank, Herr Professor Hecken. – Ich kann das, was Herr Professor Taube gesagt hat, eigentlich nur unterstreichen. Es ist wirklich sehr patientenindividuell und es ist sicher – das müssen wir einmal kritisch in dem Raum sagen –, dass wir als Pneumologen oder als Ärzte sagen: Wenn diese 4 nicht erreicht wird, sondern wir nur einen Unterschied von 2 oder 3 sehen, dass wir dann selbstkritisch sagen: Okay, das ist ein Signal, aber nicht das Signal, das wir uns vorstellen. Was heißt das für den Patienten im Einzelnen? Das heißt im Endeffekt, dass er bei einer täglichen Belastung, wo er eine Limitation hat, wo er sein Notfallspray braucht, es plötzlich nicht mehr braucht oder den Alltag besser meistern kann. Vor allem ist wichtig, dass man nachts nicht mehr aufwacht und irgendwelche Probleme und Schwierigkeiten hat, was für den Gesamtverlauf sicher auch wichtig ist. Das sind sehr patientenindividuelle Unterschiede, und die kann man nicht auf eine allgemeine Formel bringen. Es ist aber, denke ich, ganz wichtig, dass wir es geschafft haben, vor allem mit diesen Fragebögen – ich finde die Validierungen, das ist das, was Herr Professor Taube auch sagte, ganz wichtig –, dass wir in der Lage waren, das für unser Fachgebiet auf eine wissenschaftliche Basis zu stellen und dass wir dort wirklich einen Unterschied sehen. Das sind also bessere Alltagsbewältigung, weniger Krankheitstage, besser Zurechtkommen, bessere Leistungsfähigkeit, mehr ein normales Leben. Wenn dort diese minimale important difference von 4 erreicht ist, war das für uns eigentlich schon ein deutliches Signal an der Stelle.

**Herr Prof. Hecken (Vorsitzender):** Ganz herzlichen Dank, Herr Geßner. – Jetzt habe ich noch Herrn Berschneider für den pU, dann Frau Wieseler vom IQWiG und Frau Wenzel-Seifert für die Kassenärztliche Bundesvereinigung. – Herr Berschneider, bitte schön.

**Herr Dr. Berschneider (Novartis):** Vielen Dank, Herr Hecken. – Ich möchte kurz ergänzen: Auch für uns ist die patientenrelevante Responderschwelle besonders wichtig, natürlich muss der Unterschied spürbar sein. Wie Herr Taube und Herr Geßner schon gesagt haben, ist das in der Literatur aus der evidenzbasierten Validierung für diese Fragebögen gegeben. Die Schwellen sind hier aus Patientensicht, aus Klinikersicht und auch ... (akustisch unverständlich) spürbar validiert.

Ich möchte noch einen Punkt hervorheben: Dieser spürbare Unterschied wird auch dann besonders deutlich, wenn wir uns Szenarien für Patienten mit entsprechenden Unterschieden anschauen. Das geht besonders gut beim St. George's Respiratory Questionnaire, der aus vielen dichotomen Fragen besteht, die man ganz präzise mit Ja oder Nein beantworten und Szenarien bilden kann: Was ist ein Unterschied, wenn man hier Ja oder Nein angibt? Auch der Entwickler des St. George's Fragebogens, Paul Jones, hat für die repräsentative Verbesserung von 4 Punkten Szenarien aufgezeigt. Es kann zum Beispiel für einen Patienten bedeuten, dass in einem Szenario der Schlaf nicht mehr gestört ist, der Patient Tennis spielen kann und sich nicht mehr so viel in der Öffentlichkeit für Husten- und Atemgeräusche schämen muss. Auch diese Szenarien zeigen, dass es spürbare Effekte sind. Deshalb haben wir hier durchaus eine angemessene und auch anerkannte Responderschwelle.

**Herr Prof. Hecken (Vorsitzender):** Ganz herzlichen Dank für diese ergänzende Erläuterung, Herr Berschneider. – Jetzt Frau Wieseler vom IQWiG, dann Frau Wenzel-Seifert von der Kassenärztlichen Bundesvereinigung. – Frau Dr. Wieseler, bitte.

**Frau Dr. Wieseler:** Vielen Dank, Herr Hecken. – Ich glaube, was wir hier diskutieren, ist, wie wir mit einer Änderung im Stand der wissenschaftlichen Erkenntnisse umgehen, was die Datenerhebung von patientenberichteten Endpunkten angeht und insbesondere, wie wir damit umgehen, dass wir diese sogenannten minimal important differences in MIDs bestimmen wollen, um die Daten so auswerten zu können, dass wir sie auch interpretieren können. Ihre Frage an die Kliniker hat sehr schön gezeigt, dass es in der Tat schwierig ist, sich vorzustellen, was es für einen Patienten bedeutet, wenn er auf einer solchen Skala eine Änderung von 0,5 Punkten erfährt.

In den letzten fünf bis zehn Jahren haben diese patientenberichteten Endpunkte an Bedeutung gewonnen, sicherlich auch, weil sie in Verfahren wie diesem verstärkt eingesetzt werden und weil sie uns helfen, besser zu verstehen, was diese Behandlungen tatsächlich für die Patienten bringen. Ich sage mal: Seit etwa 2015 diskutiert die wissenschaftliche Community diese Frage der MIDs zunehmend kritisch. Wir haben das auch in diesem Gremium schon wiederholt adressiert. Das Problem, das wir mit diesem Verfahren sehen, ist, dass diese MIDs – auch das haben die Kliniker bestätigt – von vielen Faktoren abhängen. Sie können davon abhängen, wie schwer die Erkrankung ist, sie können von Ausgangswerten abhängen, sie können davon abhängen, ob ich eine Verbesserung oder eine Verschlechterung erfahre. Das heißt, so ganz trivial ist das nicht.

Es gibt schon seit einer ganzen Weile eine Reihe von Gruppen, die sich intensiv mit diesem Thema auseinandersetzen. In der letzten Woche gab es eine Publikation im *Journal of Clinical Epidemiology* dazu, in der diese Gruppen ihre Arbeit zusammengefasst haben. Die haben über die letzten Jahre systematisch Studien identifiziert, mit denen MIDs erhoben werden und haben sich die MIDs, die dabei herausgekommen sind, angeschaut. Diese Arbeit sagt uns, dass die Wissenschaftler 526 verschiedene Instrumente identifiziert haben, und für diese etwa 500 Instrumente haben sie 5 324 MIDs identifiziert. Also: Es gibt hier im Mittel etwa zehn verschiedene MIDs pro Instrument. Das spiegelt die methodische Schwierigkeit wider, die wir hier haben.

Wenn wir in so einer Situation sind, dass wir für ein einzelnes Instrument verschiedene MIDs haben, stellt sich die Frage: Welche davon ist die richtige? Da gehen wir in der evidenzbasierten Medizin so

vor, dass wir Kriterien beschreiben, mit denen wir die Studien bewerten, die diese MID's erhoben haben, und dann die Studien herausuchen, die uns die aussagekräftigsten MID's liefern, weil wir dann mit denen arbeiten können. Jetzt sind hier vom Hersteller und auch von den Klinikern diese Arbeiten zum Beispiel für den SGRQ benannt worden, die der pU in seiner Stellungnahme zitiert. Da sind wir genau in so einer Situation, dass das nur ein Ausschnitt der Arbeiten ist, die es zu dem SGRQ gibt.

Es gibt verschiedene weitere Arbeiten zum SGRQ, die mit anderen MID's als mit diesen vier Punkten herauskommen. Es gibt auch systematische Übersichten dazu, die das beschreiben, dass abhängig von den Patientencharakteristika sehr viel höhere Punktwerte notwendig sind, um für den Patienten einen spürbaren Effekt zu haben. Diese Situation beschreibt jetzt ganz exemplarisch das Problem, das wir haben. Da werden uns ausgewählte Arbeiten vorgelegt. Es gibt aber eine ganze Reihe von anderen Arbeiten, die zu anderen Ergebnissen kommen. Rein theoretisch müssten wir jetzt also für jede Bewertung erst einmal ein systematisches Review machen, um die vollständige Übersicht über die empirisch erhobenen MID's zu haben, um dann aus diesen die wissenschaftlich aussagekräftigsten zu ermitteln. Das halten wir in diesem Verfahren für schwierig. Wir alle gemeinsam machen mehr als hundert Bewertungen pro Jahr. Wir haben deshalb mit unserem Methodenpapier einen Vorschlag gemacht, der aus unserer Sicht sowohl wissenschaftlich tragfähig als auch einfach in diesem Verfahren durchzuführen ist.

Ich würde noch auf einen weiteren Punkt zurückkommen. Diese Arbeitsgruppen haben im Sommer einen Kriterienkatalog vorgestellt, mit dem man diese verschiedenen MID's bewerten soll oder die Studien, mit denen diese MID's erhoben worden sind. Dieser Kriterienkatalog ist aus unserer Sicht ausgesprochen nachvollziehbar und sinnvoll. Wenn wir diesen Kriterienkatalog aber auf die Studien anwenden, die aktuell verfügbar sind, sehen wir, dass diese Studien diese Kriterien in einem hohen Maße nicht erfüllen. Auch die Arbeit, die letzte Woche veröffentlicht wurde, hat gezeigt, dass zum Beispiel ein ganz wesentliches Kriterium – das ist die Korrelation zwischen dem Anker und dem Instrument, das ich bewerten möchte – in der überwiegenden Zahl der Studien überhaupt nicht berichtet wird. Auch weitere Kriterien, die diese Autoren aufstellen, werden in den aktuellen Studien nicht berichtet. Das heißt, wir würden einen hohen Aufwand betreiben, ein systematisches Review machen, damit wir wissen, welche MID's es gibt. Wir würden diese Studien bewerten und zu dem Schluss kommen: Es fehlt uns ganz viel Information, um zu sagen, diese MID aus den zehn, die ich zur Auswahl habe, ist diejenige, die aussagekräftig ist.

In dieser Situation haben wir einen Vorschlag gemacht, der, wie gesagt, aus unserer Sicht sicherstellt, dass wir mit Responsekriterien arbeiten, die hinreichend sicher einen spürbaren Effekt für den Patienten abbilden und ein Verfahren, das gleichzeitig hier in der Dossierbewertung machbar ist. Dieses Verfahren haben wir in unserem Methodenpapier beschrieben. Wir haben in der Erörterung des Methodenpapiers die Stellungnahmen dazu diskutiert. Das Ganze ist auch auf unserer Web-Site dokumentiert. Wir denken, dass ein Verbleiben auf dem Status, den wir kennen, also die alten MID's, unverändert weiter zu verwenden, den Stand der wissenschaftlichen Erkenntnis nicht abbildet. Wir haben in der Diskussion um unser Methodenpapier keine Alternativvorschläge aus akademischen Gruppen oder von der Industrie erhalten. Unser Vorschlag ist jetzt, mit dem von uns vorgeschlagenen Verfahren weiterzugehen und die wissenschaftliche Diskussion weiter zu verfolgen. Sobald wir andere Möglichkeiten haben, würden wir das wieder aufgreifen wollen.

Ich würde noch einen Punkt unterstreichen wollen. Der Vorschlag, den wir machen, dass wir jede prä-spezifizierte MID heranziehen, die mindestens 15 Prozent ist oder aber eine Post-hoc-Analyse, wenn sie genau 15 Prozent umfasst, ermöglicht es, jede Studie, die nicht unter diesen Vorgaben designt wurde, auch zu verwerten. Diese Post-hoc-Analyse ermöglicht es, für jedes Dossier eine Responderauswertung zu machen, die akzeptiert werden kann. – Danke.

**Herr Prof. Hecken (Vorsitzender):** Herzlichen Dank, Frau Wieseler, für diese bewertende und einordnende Klarstellung. Dazu hat sich vom pharmazeutischen Unternehmer Herr Diwischek gemeldet, dann Herr Rasch vom vfa. Dann würde ich die beiden Kliniker noch einmal befragen, ob sie zu den Anmerkungen von Frau Wieseler zu der jetzt zugrunde liegenden Methodik beim IQWiG etwas sagen möchten, und dann würde ich Frau Wenzel-Seifert das Wort geben, weil sie auch eine Frage gestellt hatte. – Fangen wir mit Herrn Diwischek an, dann Herr Rasch. – Herr Diwischek, bitte.

**Herr Dr. Diwischek (Novartis):** Danke, Herr Professor Hecken. – Die wichtige Frage ist doch in dem Zusammenhang: Was ist für den Patienten spürbar? Was ist eine relevante Schwelle? Wir haben uns angeschaut, wie der Stand der Wissenschaft hier ist. Es gibt validierte Schwellen für eine spürbare Veränderung in der Lebensqualität beim AQLQ und SGRQ. Das IQWiG hat via Methodenpapier einen generischen Vorschlag gemacht, auch wenn es nach Einreichung unseres Dossiers in Kraft getreten ist. Das würde für den SGRQ zum Beispiel 15 Punkte als Schwelle bedeuten. Wir haben uns an die asthmaspezifischen Schwellen gehalten; die sind in der Fachwelt und auch vom G-BA anerkannt. Das bedeutet für den SGRQ 4 Punkte. Das heißt, es wäre mit dem generischen Vorschlag des IQWiG fast eine Vervielfachung der Schwelle. Wir haben nach den asthmaspezifischen ausgewertet und so signifikante und bedeutsame Unterschiede für die Patienten in der Lebensqualität gezeigt. – Danke.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Herr Diwischek. – Herr Rasch, ich nehme an zum Inkrafttreten der neuen IQWiG-Methode. – Herr Rasch, bitte.

**Herr Dr. Rasch (vfa):** Ganz genau. Vielen Dank, Herr Hecken. – Ich möchte jetzt nicht die IQWiG-Methodik detailliert diskutieren. Mir wäre es wichtig, anzumerken: Frau Wieseler hat gesagt, es sei aus den Ausführungen von Herrn Professor Geßner und Taube unklar, was 4 Punkte für die Patienten bedeuten. Ich sah das ganz anders. Ich habe das so verstanden, dass die vier Punkte als eindeutig relevant angesehen werden. Darüber hinaus muss man anmerken, dass dieser Fragebogen auch vom IQWiG in der Vergangenheit akzeptiert wurde, ohne dass je methodische Kritik daran geäußert wurde.

Aber ganz allgemein oder das einmal hier anzumerken: Wir sehen das als vfa auch sehr kritisch. Wir stehen der angewandten neuen Methodik sehr kritisch gegenüber. Das haben wir in unserer Stellungnahme zum Methodenpapier kundgetan, das haben wir aber auch unserer jetzigen Stellungnahme zu diesem Verfahren beigelegt. Die ausführliche Position liegt der Stellungnahme bei; deshalb will ich sie nicht im Detail wiederholen. Wichtig ist tatsächlich, um das gut zusammenzufassen, dass dieser neue Ansatz ohne Patientenbeteiligung ermittelt wurde. Er folgt auch nicht – ganz eindeutig nicht – dem Entwicklungsansatz der internationalen Wissenschaft, der auf gängigen und bislang vom IQWiG akzeptierten Qualitätskriterien basiert. Dieser Ansatz geht in eine ganz andere Richtung. Der Ansatz der internationalen Wissenschaft geht in eine ganz andere Richtung als das, was das IQWiG als Einheitsmaß hier vorgeschlagen hat, und dieses Einheitsmaß sehen wir auch deshalb kritisch, weil es für alle Situationen gelten soll, die aber doch unterschiedlich sind und auch die Unterschiede in der Krankensituation abbilden. Er würde – das haben wir auch unserer Stellungnahme beigelegt – dazu führen, dass der absolute Großteil – ich meine, das sind über 90 Prozent – der bisher akzeptierten MIDs damit wegfallen würde, neben diesen akzeptierten, über die wir hier sprechen, auch im Bereich der Onkologie, QLQ-C30, also ganz gängige Fragebögen.

Wir würden, ohne das methodisch zu vertiefen, nur anregen, dass der G-BA in solchen Fällen stets eine Einzelfallprüfung vornimmt und dieses generische Richtmaß ohne Weiteres nicht anwendet, vorsichtig vorgeht. Wir würden auf jeden Fall dafür plädieren, dass man hier einen allgemein akzeptierten Katalog an Beratungskriterien anstrebt, der diese Bestrebungen der wissenschaftlichen Gemeinschaft



in den letzten Jahren und Jahrzehnten unterstützt und nicht komplett negiert. Ich glaube, die Methodik sollte erst dann verändert werden, wenn es dazu irgendeine Art Konsens bei diesen Diskussionen gibt. Darüber hinaus – das hat, glaube ich, die betroffene Firma schon angesprochen –, ich glaube, solange das so strittig ist, sollte man auch hier auf die Verfahrenskonsistenz achten, und zumindest sollten die Fragebögen oder die MIDs, besser gesagt, die bislang akzeptiert wurden, auch weiterhin akzeptiert werden, also die gängigen und validierten und bislang akzeptierten Fragebögen. – Vielen Dank.

**Herr Prof. Hecken (Vorsitzender):** Herzlichen Dank, Herr Rasch. – Jetzt habe ich Ergänzung dazu von Frau Wieseler, dann würde ich noch mal Herrn Taube und Herrn Geßner fragen, danach kommen wir zu Frau Wenzel-Seifert, damit wir in der Fragerunde weiterkommen. – Frau Wieseler.

**Frau Dr. Wieseler:** Ich möchte nur kurz darauf Bezug nehmen, dass Herr Rasch sagt, da gab es keine Patientenbeteiligung. Ich meine, die generische Schwelle, die wir vorschlagen, haben wir nicht aus der Luft gegriffen. Wir haben systematisch nach systematischen Reviews zu MIDs in verschiedenen Indikationen gesucht und diese Schwelle an den Ergebnissen dieser systematischen Zusammenstellung orientiert. Das ist im Detail in der Würdigung der Stellungnahme zu unserem Methodenpapier ausgeführt. In dieser ganzen Diskussion lässt mich ein wenig ratlos, dass wir die Befunde, die wir ganz offensichtlich haben, also den Stand der wissenschaftlichen Erkenntnis, hier überhaupt nicht aufgreifen. Ich höre keinen Vorschlag und, wie gesagt, wir haben das auch in der langen Diskussion zu unserem Methodenpapier nicht gehört, wie man mit den Dingen, die auf dem Tisch liegen, alternativ umgehen soll. Wenn Herr Rasch darauf hinweist, dass ein Katalog von Bewertungskriterien erarbeitet werden soll, dann haben wir diesen Katalog. Dieser Katalog ist von der wissenschaftlichen Community erarbeitet worden. Diese Gruppen arbeiten seit mehr als 30 Jahren in diesem Bereich. Ich glaube nicht, dass wir darauf warten müssen, dass wir hier jetzt einen weiteren Katalog erarbeiten und zur Verfügung stellen.

Da noch mal mein Hinweis: Die Anwendung dieses Katalogs zu diesem Zeitpunkt würde in der Mehrzahl der Fälle dazu führen, dass wir die Qualität dieser MIDs nicht bewerten können, weil die Informationen nicht vorliegen. Unser Vorschlag ermöglicht uns, in dieser Situation trotzdem weiterzugehen, jede Studie, die durchgeführt würde, diesbezüglich in das Verfahren einzuschleusen. Natürlich werden wir weiterschauen, wie sich das Feld entwickelt. Aber die Befunde, die auf dem Tisch liegen, zu negieren und zu sagen, nein, wir bleiben einfach bei den alten MIDs und hoffen das Beste, ist, glaube ich, nicht der richtige Ansatz.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Frau Wieseler. – Jetzt noch mal Frage an Herrn Taube, Herrn Geßner. Wir diskutieren jetzt über Relevanzschwelle, also Methodenpapier, Relevanzschwelle, MIDs 15 Prozent. Was oder wenn überhaupt sagen Sie dazu? Wir haben beim IQWiG dann eine generische Festsetzung, und der Kontrapunkt ist aufgemacht worden: Wir haben mit den beiden Instrumenten AQLQ und SGRQ keine generischen Fragebögen, sondern indikationsbezogene Fragebögen, wo man die 15 Prozent nicht sieht. Da hat aber Frau Wieseler gesagt, na ja, die haben sich jetzt auf vier Domänen und vier Endpunkte kapriziert, bei denen man dann diese Effekte sieht. Es gibt aber in der Literatur, ich habe die Zahl jetzt nicht mehr im Kopf, aber eine Vielzahl von Publikationen, bei denen noch auf andere Punkte geschaut wird, sodass es die große Frage ist, ob die 0,5 oder die 4 Prozent, die Sie hier zeigen, am Ende des Tages auch im Lichte anderer Publikationen Bestand haben könnten. Das ist, glaube ich, zusammengefasst die Diskussion, die wir im Augenblick führen. Was können Sie, möchten Sie dazu sagen, bevor wir Frau Wenzel-Seifert das Wort geben? – Fangen wir wieder mit Herrn Taube an, dann Herr Geßner.

**Herr Prof. Taube (DGP):** Vielen Dank. – Es ist so, dass gerade für den St. George's die Schwelle von 4, denke ich, wissenschaftlich gut belegt ist. Es gibt viele Studien, die das machen und, wie gesagt, auch viele Zulassungsbehörden, die gerade diesen Weg wählen, dass wir sagen: Okay, 4 Punkte ist der Cut-off. Wir machen eine Responderanalyse, wie viele Patienten es über die 4 Punkte schaffen. Wie viele schaffen es nicht? Um daran, an den Vergleichen dann eine doch, denke ich, relevante Aussage machen zu können, wie viele Patienten wirklich von dieser Medikation profitieren. Ich habe mir das Methodenpapier des IQWiG noch mal angeschaut. Da ist es so – ich kann das auch nachvollziehen bei der Vielzahl der unterschiedlichen Fragebögen –, dass man den St. George's nicht noch mal adäquat neu bewertet oder sich noch mal die Studien angeschaut oder neue Berechnungen durchgeführt hat – das geht wahrscheinlich auch gar nicht von der Menge her –, sondern dass diese generische Zahl von 15 Prozent aus einer Mischung von unterschiedlichen Fragebögen, wahrscheinlich Orthopädie, Schmerzfreiheit, andere Lebensqualitätssachen entstanden ist. Ich glaube nicht, dass das für ein so gut qualifiziertes Instrument, das wirklich akzeptiert ist und sowohl von allen internationalen Journalen, wie *New England Journal of Medicine* usw., als auch von anderen Zulassungsbehörden wie FDA als Instrument ganz klar anerkannt ist – – Das Problem zum jetzigen Zeitpunkt ist, denke ich, dass man dort eine Analyse gemacht hat, die nach meinem Dafürhalten State of the Art ist, und die anhand dieser neuen Kriterien, die jetzt aufgestellt wurden, als ungültig oder sozusagen nicht zu bewerten klassifiziert wird.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Herr Taube. – Herr Geßner, wahrscheinlich gleiche Anmerkung.

**Herr Dr. Geßner (POIS Leipzig):** Ich möchte nicht wiederholen, was Herr Taube gesagt hat. Ich denke sicher, dass man den spezifischen Blick auf die Indikation, in dem Fall das Asthma, haben muss, dass vielleicht bei bestimmten Dingen, die in der Orthopädie oder woanders zu sehen sind, diese gleichen Levels nicht im Asthma erreicht werden. Ich halte die 15 Prozent, wenn ich den Alltag sehe, für deutlich zu hoch. Ich sehe eher den Patienten vor mir und das, was ich sehe, frage und erlebe, da muss man mit Blick auf die ARGON-Studie sagen, wenn man auf andere Dinge schaut, dann sehen wir in der ARGON-Studie, dass wir in der Gruppe mit entsprechender Hochdosissteroidtherapie auch eine Verbesserung der Lungenfunktion um 100 ml gesehen haben. Das Ganze ist doch in völliger Übereinstimmung zu diesen Daten, die wir zur Lebensqualität haben. Das unterstützt das Ganze in dem Sinne auch, dass ich weiß – und das weiß ich aus meinem klinischen Alltag –, dass ein Patient, der 100 ml mehr Lungenvolumen hat und der, sagen wir mal, eine schwere Erkrankung hat, dort plötzlich seinen Alltag klinisch deutlich besser meistern kann, besser zurechtkommt, das sind soundso viele hundert Meter mehr gelaufen und Ähnliches. Das ist das, was wir im Alltag sehen. Deshalb sind diese Daten und vor allem auch diese Grenze von 4 für mich als Kliniker sehr glaubwürdig, sehr gut und helfen mir eigentlich im Alltag. Ich verstehe völlig die methodischen Dinge, gebe aber zu, ich bin zu wenig Statistiker, um mir anmaßen zu wollen, dort eine Methodenkritik durchführen und sagen zu können: Okay, das ist der Aspekt oder der. Nur, ich denke, der indikationsspezifische Blick ist ein ganz wichtiger. – Danke.

**Herr Prof. Hecken (Vorsitzender):** Herzlichen Dank, Herr Dr. Geßner. – Jetzt habe ich Frau Wenzel-Seifert, dann noch mal Herr Rasch. – Frau Wenzel, bitte.

**Frau Dr. Wenzel-Seifert:** Hören Sie mich?

**Herr Prof. Hecken (Vorsitzender):** Ja, wunderbar.

**Frau Dr. Wenzel-Seifert:** Ich möchte noch einmal auf dieses Thema zurückkommen. Frau Wieseler hat vorhin erwähnt, dass die Betrachtung einer MID auch vom Schweregrad abhängt. Nun bezieht sich diese Schwelle von 15 Prozent, die wir jetzt haben – vorher waren es 4 Prozent, die alte MID von der gesamten Spannweite – auf die gesamte Spannweite, die in diesem Test erfasst werden kann. Ist es bei den Patienten, um die es hier geht – die sind in der Stufe IV –, überhaupt realistisch, anzunehmen, dass sich ihre Werte tatsächlich auf diese gesamte Spannweite von 0 oder 1, ich weiß nicht genau, wo es losgeht, bis 100 bezieht? Oder ist das nicht eigentlich ein viel geringerer Bereich, den man da betrachten muss, sodass sich relativiert, worauf sich die 15 Prozent eigentlich beziehen müssten; auf die gesamte Spannweite, die vielleicht in diesem Schweregrad nicht von allen erreichbar ist, sondern möglicherweise bewegen die sich in einem ganz anderen, viel kleineren Bereich von – ich weiß nicht – 30 Punkten. Das ist die Frage an die Kliniker: Ist es realistisch für Patienten mit diesem Schweregrad, dass die tatsächlich Punkte zwischen 0 und 100 erreichen können, oder bewegen die sich vielmehr in einem geringeren Segment, wo sie therapeutisch beeinflussbar sind?

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Frau Wenzel-Seifert. – Herr Taube vielleicht oder Herr Geßner? – Herr Taube, bitte.

**Herr Prof. Taube (DGP):** Das ist sicherlich ein sehr guter Punkt. Ich glaube, der Mittelwert bei dem St. George's in der ARGON-Studie – das wird Herr Geßner bitte noch ergänzen –, war um die 40. Ja. Wenn wir Patienten mit schwersten Erkrankungen wie zum Beispiel ausgebrannter COPD oder sehr schwerem Asthma haben, gehen wir in die Richtung 50 oder höher. Es ist so, dass die Spannbreite insgesamt in der Studie dann deutlich geringer ist als diese 9 Punkte, die, glaube ich, vom IQWiG mit angegeben wurden.

**Herr Prof. Hecken (Vorsitzender):** Danke schön. – Herr Geßner, vielleicht Ergänzung.

**Herr Dr. Geßner (POIS Leipzig):** Ich kann das nur unterstützen. Ich hätte nichts Substanzielles weiter.

**Herr Prof. Hecken (Vorsitzender):** Danke schön.

**Herr Prof. Taube (DGP):** Wenn ich noch eine kleine Ergänzung machen darf?

**Herr Prof. Hecken (Vorsitzender):** Ja, bitte, Herr Taube.

**Herr Prof. Taube (DGP):** Tut mir leid, aber das ist vielleicht noch ein ganz wichtiger Punkt. Es ist so, dass wir in der Vergangenheit auch andere Studien bei Asthma mit anderen Medikamenten durchgeführt haben; die sind auch hier im G-BA besprochen worden. Das ist die Antikörperbehandlung, und da haben wir im St. George's, der dort ausgewertet wurde, Veränderungen von 8 oder 9 Punkten gesehen, die aber mit einer erheblichsten – erheblichsten! – klinischen Verbesserung einhergegangen sind. Das sind teilweise Behandlungen – das ist auch dort thematisiert worden –, die wirklich zu einer Veränderung des Lebens bei diesen Patienten führen, die ganz massiv ist. Wie gesagt, wenn man das jetzt sozusagen als minimal clinical important difference zugrunde legt, dann greift das, glaube ich, diesem Instrument zu kurz.

**Herr Prof. Hecken (Vorsitzender):** Herzlichen Dank, Herr Professor Taube. – Frau Wenzel-Seifert, beantwortet das Ihre Frage?

**Frau Dr. Wenzel-Seifert:** Noch nicht ganz. Ganz konkret gefragt: Ist es denkbar bei diesen Patienten, von denen Sie gerade gesagt haben, die hatten in der ARGON-Studie einen Mittelwert von 40, dass

die runtergehen auf 0 bzw. auf 100? Ist das möglich, dass sie überhaupt auf dieser Spannweite laufen können?

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Frau Wenzel-Seifert. – Ich frage noch mal Herrn Professor Taube, dann Herrn Geßner.

**Herr Prof. Taube (DGP):** Nein, nein, es ist nicht zu erwarten, dass die auf 0 runtergehen oder dann auf maximal 100 hochgehen; das sind zu dramatische Veränderungen.

**Herr Prof. Hecken (Vorsitzender):** Danke. – Herr Dr. Geßner.

**Herr Dr. Geßner (POIS Leipzig):** Das entspricht nicht der klinischen Realität, das kann ich nur unterstreichen.

**Herr Prof. Hecken (Vorsitzender):** Okay. Danke. – Frau Wenzel-Seifert, jetzt beantwortet? – Ich dachte, Sie hätten noch gefragt, was in der klinischen Realität die ideale Veränderung nach unten ist.

**Frau Dr. Wenzel-Seifert:** So viel verlange ich nicht.

**Herr Prof. Hecken (Vorsitzender):** Aber das kann man wahrscheinlich nicht beschreiben. – Jetzt haben wir Herrn Rasch und dann Frau Wieseler. – Herr Rasch, bitte.

**Herr. Dr. Rasch (vfa):** Vielen Dank. – Also einmal kurz zurück zu dem, was Frau Wieseler gesagt hat: Es ist nicht so, dass die Methodik zwingend an dieser Stelle völlig neu erfunden werden muss. Wir sprechen hier über die üblichen Qualitätskriterien, die Frau Wieseler selbst in den vergangenen Anhörungen vor zwei, drei Jahren hier eingebracht hat, die in der Literatur den Forschungsstand der letzten 20, 30 Jahre abbilden. Die 15 Prozent, die das IQWiG hier vorgeschlagen hat, sind nur eine simple Mittelung aus den beobachteten Werten zwischen 10 und 20 aus irgendeiner Literaturlauswahl. Wir haben uns das beim vfa genauso angeschaut und uns alle MIDs angeschaut, die bislang im AMNOG verwendet und akzeptiert wurden. Das ist eine Spannweite, die ganz anders ist. Die liegt zwischen 2 und 40 Prozent. Das spiegelt auch die Spezifika und die Individualität der Fälle wider. Da würde man zu einem ganz anderen Mittelwert kommen. Aber die Frage ist in der Tat: Wir haben dieses völlig neue generische Richtmaß auf der einen Seite, was für mich offenkundig nicht die Individualität der Erkrankungssituation widerspiegelt, und wir haben auf der anderen Seite etablierte, validierte und auch international akzeptierte Schwellenwerte, die diese krankheitsspezifischen Spezifika erfüllen; und es ist einfach die Frage, wonach man sich richtet. Wie gesagt, unser Appell an den G-BA wäre, dass man das sehr genau und kritisch prüft.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Herr Rasch, für diese Anmerkung und diesen erneuten Hinweis. – Frau Wieseler, IQWiG, bitte.

**Frau Dr. Wieseler:** Ich möchte nur noch mal darauf hinweisen, dass insbesondere diese 4 Punkte für den SGRQ, den wir hier sehr intensiv diskutieren, mitnichten durchgängig akzeptiert sind. Es gibt zum Beispiel ein systematisches Review aus dem Jahre 2018, das ganz klar zu der Aussage kommt, dass diese 4 Punkte auf Basis anderer Studien zu niedrig sind. Auch was Herr Rasch beschreibt, dass die Kriterien, die wir bisher akzeptiert haben, zwischen 2 und 40 Prozent des Scalenranges liegen, zeigt uns – Ich meine, es ist aus meiner Sicht nicht vorstellbar, dass 2 richtig ist und 40 auch nicht als MID. Das zeigt nur noch mal die Probleme, die wir da haben. Was die Kriterien angeht, wie gesagt, jetzt gibt es einen erweiterten Kriteriensatz für die Bewertung der Qualität dieser MIDs, weil die Probleme

beschrieben worden sind und man erkannt hat, das, was wir bisher gemacht haben, ist nicht ausreichend. Ich meine, der AQLQ ist eine Arbeit aus den 90er-Jahren. Das war eine der ersten Arbeiten, die überhaupt diesen Ansatz angewendet hat; sicherlich eine hoch verdienstvolle Arbeit, weil sie den Patienten in der Form erst mal eine Stimme gegeben hat. Diese MID beruht auf etwa zehn Patienten. Da können wir nicht sagen, das ist etabliert, das haben wir immer so gemacht, das muss so bleiben. Das ist einfach nicht adäquat.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Frau Wieseler. – Anmerkungen dazu? – Keine. Weitere Fragen? – Frau Biester, IQWiG, anderes Thema.

**Frau Biester:** Genau. – Ich hätte noch einen Punkt zu dem Asthmatagebuch. Es ist schon mehrfach darauf hingewiesen worden, dass man in der Studie ARGON in der Asthmasymptomatik Vorteile sehen würde. Der Hersteller hat einige Informationen nachgeliefert, die uns zur Operationalisierung gefehlt haben, ob die Fragen prädefiniert waren, die Auswertung prädefiniert war oder nicht. Erst mal vielen Dank für die Informationen, die Sie nachgereicht haben. Wir sind daraus schlauer geworden, und zwar insofern, dass die Operationalisierung, die einzige Operationalisierung, aus der Sie einen Vorteil ableiten, eine Post-hoc-Analyse ist. Aus der Stellungnahme ist mir die Rationale nicht klar geworden, das so zu machen.

Es sind in diesem Patiententagebuch sieben Fragen, es gibt einen sogenannten Mean total daily symptom score. Das ist eine Operationalisierung, die alle sieben Fragen umfasst, die im Modul IV zwar vorgelegt wurde, die der pU aber nicht heranzieht und auch nicht betrachtet, in keinster Weise. Stattdessen nimmt der pU eine Operationalisierung, die er auch in seinem Eingangsstatement beschrieben hat, Anteil der Tage ohne Asthmasymptome, die ebenfalls alle sieben Fragen umfasst; die ist eben nicht präspezifiziert, die ist Post hoc. In Ihrer Stellungnahme beschreiben Sie, die sei – und da zitiere ich jetzt – „offenkundig abgeleitet von der präspezifizierten Operationalisierung“. Sie beschreiben nicht, was daran offenkundig sein soll, möglicherweise dass es alle sieben Fragen sind. Ich muss sagen, es hat ein Geschmäcke, weil die einzige Operationalisierung, auf der etwas herauskommt, die Post-hoc-Operationalisierung ist. Und: „Offenkundig“ kann ich auch deshalb nicht nachvollziehen, weil aus dieser präspezifizierten Operationalisierung mit allen sieben Fragen kein Vorteil herauskommt. Deshalb stimmt dieses Argument meiner Meinung nach nicht ganz. Man sieht das auch alles in der Asthmasymptomatik; das ist nicht nachvollziehbar. Wenn der pU mag, kann er vielleicht zu dem „offenkundig“ noch etwas sagen, aber ich finde, aus den nachgereichten Unterlagen wird sehr klar, dass man diese Daten für die Nutzenbewertung nicht heranziehen kann.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Frau Biester. – Ich habe jetzt Herrn Berschneider zu dieser Fragestellung, dann Frau Müller zu den MIDs. – Herr Berschneider vom pharmazeutischen Unternehmer, bitte schön.

**Herr Dr. Berschneider (Novartis):** Danke, Herr Hecken. – Ich gehe gern noch mal auf unsere Tagebücheranalyse nach Tagen ohne Symptome ein. Ich möchte am Anfang klarstellen: Ein symptomfreier Tag ist ein gewonnener Tag für den Patienten. Die Auswertung war präspezifiziert für den Mean-Score nächtliche Symptome beim Aufwachen tagsüber. Relevant sind aber die 24 Stunden. Das ist ein relevantes Maß, und aus unserer Sicht ist ein ganzer Tag ohne Symptome ganz klar erfassbar und deutlich spürbar. Es gibt quasi aus unserer Sicht eine natürliche Ableitung, das kann ganz klar mit Ja oder Nein beantwortet werden. Waren es 24 Stunden ohne Symptome, ja/nein, kein Mean-Score, sondern eine ganz klare Ableitung. Da sind die Patienten signifikant im Vorteil. Die Patienten hatten 21 Tage mehr

ohne Symptome im Jahr. Aus unserer Sicht ist deshalb diese Auswertung verwertbar und zur Ableitung des Zusatznutzens auch anwendbar.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Herr Berschneider. – Jetzt habe ich noch mal Frau Müller zu diesen MIDs. – Frau Müller, bitte, Kassenärztliche Bundesvereinigung.

**Frau Dr. Müller:** Vielen Dank. – Tut mir leid, dass ich noch mal zurückspringe. Ich glaube, es ist nicht schnell genug angekommen. Ich hatte dazu erst danach eingeklickt. Noch mal ganz kurz zurück: Ich wollte noch zwei Aspekte ansprechen. Ich hätte zunächst eine Frage an den pU und wollte noch einen allgemeinen Aspekt ansprechen. Der allgemeine ist eigentlich nach dem, was Frau Wieseler eben zu der kritischeren Herangehensweise in der wissenschaftlichen Community zu den bis jetzt verwendeten MIDs ausgeführt hat. Das betrifft nicht nur die Lebensqualität, sondern auch die Morbidität, wenn Fragebögen eingesetzt werden. Meine Frage dazu ist: Was ist eigentlich zu unseren ganzen bisherigen Bewertungen – unseren und auch anderen –, die auf diesen MIDs fußen, ... (akustisch unverständlich)? Was heißt das eigentlich für unsere Bewertung bei den letzten zehn, 15, 20 Jahren, vielleicht sogar schon länger, ich will mal sagen, die bei uns eine besondere Relevanz bekommen haben und besonders gewichtet wurden? Das ist meine eine Frage.

Die andere ist eine sehr konkrete an den pU. Frau Wieseler hat eben auf den erweiterten Kriterienkatalog rekurriert, wobei neben der Frage, die schon länger klar ist, dass man eigentlich ankerbasierte und verteilungsbasierte Verfahren verwenden wollte, auch noch auf andere Punkte eingegangen wurde. Würden Sie es so sehen, dass das, was Sie hier bezüglich der Validierung der MIDs vorgelegt haben, auf diese Aspekte, die aktuell in der wissenschaftlichen Community diskutiert werden, eingeht? Die zweite Frage ist: Frau Wieseler ist darauf eingegangen, dass es eine nicht ganz vollständige Auswertung gewesen wäre, weil nur auf diese bisher akzeptierten MIDs eingegangen wurde, aber nicht auf andere. Wie würden Sie das sehen? Haben Sie da die entsprechende Literaturrecherche gemacht? Oder, um es zusammenzufassen: Würde das, was Sie hier vorgelegt haben, geeignet sein, sozusagen unter strengeren, jetzt gültigen wissenschaftlichen Kriterien dieser MIDs noch mal zu bewerten, unabhängig davon, ob das überhaupt mit vertretbarem Aufwand durchführbar ist?

**Herr Prof. Hecken (Vorsitzender):** Wer möchte dazu? – Herr Dr. Wasmuth.

**Herr Dr. Wasmuth (Novartis):** Wenn zunächst die Frage an den pU ist, kann ich das gerne machen. – Danke noch mal für die Nachfrage, Frau Müller. Wir schauen uns natürlich schon die aktuelle Evidenz an, und die Bewertung des IQWiG haben wir jetzt noch mal zum Anlass genommen, uns intensiver mit der Thematik auseinanderzusetzen. Ich glaube, die Diskussion mit den medizinischen Fachexperten zeigt, dass das eigentlich immer noch State of the Art ist, was hier passiert ist. Das heißt, wir denken, dass es immer noch eine sehr aussagekräftige Analyse ist, dass diese Relevanzschwellen immer noch die sind, die man für die Bewertung heranziehen muss. Gleichzeitig verstehe ich das Argument von Frau Wieseler, die sagt, wir müssen ganz viele bewerten, und es gibt Dinge, die sozusagen besser oder schlechter untersucht sind. Ich denke, hier haben wir einen Fall, der deutlich auf der einen Kategorie ist, nämlich auf den besser untersuchten, und ich denke, man muss das – das hat, glaube ich, auch einer der Fachexperten gesagt – immer genau anschauen. Hier haben wir entsprechende Evidenz, sie wurde auch in der Vergangenheit so herangezogen, hier ist das Feedback der Fachexperten so, und deshalb sollte man in diesem Fall auf die etablierten MIDs zurückgreifen.

**Herr Prof. Hecken (Vorsitzender):** Danke schön.

**Herr Dr. Wasmuth (Novartis):** Ich hoffe, ich habe die Fragen jetzt alle beantwortet, Frau Müller.

**Frau Dr. Müller:** Ja, ich verstehe, dass Sie das praktisch nach den bisherigen Kriterien vorgelegt haben und das einfach weiter heranziehen wollen. Die Frage ist beantwortet.

**Herr Dr. Wasmuth (Novartis):** Wenn ich das noch kurz ergänzen darf, Frau Müller: Das war auch der Stand des Methodenpapiers, als wir das Dossier eingereicht haben. Das muss man noch dazu sagen.

**Herr Prof. Hecken (Vorsitzender):** So sind Sie auch beraten worden, das muss man auch dazu sagen. Also, da ist der St. George's ausdrücklich genannt worden, und bei der alten Methodik war nicht die Rede von den 15 Prozent. Okay. – Jetzt habe ich noch mal Frau Wieseler.

**Frau Dr. Wieseler:** Vielen Dank. – Ich möchte auf die erste Frage von Frau Müller zurückkommen, die fragt, wie diese methodische Weiterentwicklung in Bezug auf das einzuordnen ist, was wir bisher getan haben. Ich meine, wir haben, wie Sie ganz richtig sagen, bisher mit diesen anderen MIDs gearbeitet. Aber ich meine, das ist ein Vorgehen, das wir nicht nur mit dieser Fragestellung haben, sondern was immer wieder geschieht, dass es eine Weiterentwicklung in der wissenschaftlichen Erkenntnis gibt, die wir irgendwann im Verfahren abbilden. Wir sind in einer permanenten Methodenentwicklung, was die Anwendung der evidenzbasierten Medizin angeht. Wir haben sehr früh angefangen, das im Ausschuss zu adressieren. Es hatte auch eine Reihe von Stellen bereits Entscheidungen getroffen, gewisse MIDs nicht mehr heranzuziehen. Wir sind jetzt aber mit diesen Arbeiten, die in diesem Jahr veröffentlicht wurden und die die Evidenzbasis für diese Methodenanpassung noch einmal darlegen, in einer Situation, in der wir glauben, dass wir jetzt tatsächlich eine Änderung vornehmen müssen. Wie gesagt, eigentlich glaube ich auch, dass der methodische Ansatz, den wir jetzt fahren, so eine Art Übergangsansatz ist. Wir ignorieren nicht die ganzen Befunde, die es gibt, und haben einen Vorschlag entwickelt, der es erlaubt, jede einzelne Studie, die noch anders geplant wurde, trotzdem für dieses Verfahren zu nutzen.

Wir werden uns die Studien zur Entwicklung von MIDs weiter anschauen, und wenn die irgendwann die Informationen haben, die jetzt als Qualitätskriterien vorgeschlagen werden, dann können wir wieder damit arbeiten. Aber es ist jetzt absehbar, dass wir – – Wie gesagt, wenn ich jetzt auf diese Literatur hier schaue, sind in über 90 Prozent der Studien die Informationen für diese Kriterien nicht enthalten. Ich finde es auch für das vorliegende Verfahren nicht sinnvoll, zu sagen: Ja, gut, das ist der neue Vorschlag. Den setzen wir jetzt um, und wir wissen jetzt schon, wir werden dann in über 90 Prozent der Fälle sagen: Ja, wir haben aber die Information nicht. Dann scheint mir der Vorschlag, der jetzt auf dem Tisch liegt, mit dem wir in dieser Übergangsphase gegebenenfalls arbeiten können, für alle Beteiligten sinnvoller zu sein. Es wäre ein erheblicher Aufwand in der Dossiererstellung, wenn ich für jede MID, die ich einsetzen möchte, erst mal ein systematisches Review machen muss; und nach diesen Befunden müsste ich das.

**Frau Dr. Müller:** Vielen Dank.

**Herr Prof. Hecken (Vorsitzender):** Danke schön, Frau Wieseler. – Also, ich sage einmal, obwohl wir hier in der Anhörung sind, ist für uns die entscheidende Frage – und das hat Herr Rasch eben auf den Punkt gebracht – gar nicht die Frage: Ist die in die Zukunft weisende Übergangsmethodik, der Übergangsansatz des IQWiG jetzt richtig, weniger richtig, falsch? Das ist überhaupt nicht der Punkt, sondern der Punkt ist: Wie müssen wir im konkreten Fall in Ansehung des Umstandes, dass wir beraten haben, in Ansehung des Umstandes, dass wir in der Vergangenheit insbesondere den St. George's anerkannt haben, allerdings bei größeren Veränderungen – – Eben ist es gesagt worden, wir hatten nach meiner Erinnerung eine Bewertung, wo wir eine Veränderung von 15 Punkten hatten, die dann als relativ

signifikant, also als beträchtlich signifikant angesehen wurde. Die Frage, die sich für uns jetzt im konkreten Fall stellt, neben der Frage: Ist diese Grundsatzmethodik des IQWiG jetzt in Stein gemeißelt mit Blick auf die Vorgeschichte, ist die Frage: Sind die 4 Punkte hier in der konkreten Therapiesituation so evidenziell abgesichert, dass man sagen kann: Okay, da kommt man nicht daran vorbei? Einfach, um das ein wenig auf den Punkt zu bringen. Das scheint mir fraglich zu sein.

Die klinischen Experten haben gesagt: 4 Punkte ist immer patientenindividuell relevant. Frau Wieseler hat eben aber ausgeführt, dass es eine Reihe von Publikationen gibt, in denen bezogen auf den St. George's gesagt wird: Die 4 Punkte als solche reichen nicht aus, um einen Effekt darzustellen, der am Ende einen patientenrelevanten Vorteil abbildet. Das ist auch ein Stück weit das, was Gegenstand Ihrer Frage war, Frau Müller. Wie gehen wir mit Blick auf die Spruchpraxis in der Vergangenheit und dann in Ansehung der 4 Punkte hier mit dem konkreten Bewertungsverfahren um, das jetzt in die Veränderung der Methodik des IQWiG fällt, die für die 1 Milliarde unterschiedlichen Messinstrumente, die wir hier haben, sicherlich in vielen Fällen ein sehr praktikabler Ansatz ist, weil er es ermöglicht, ohne in die Tiefe zu gehen und über eins, zwei Punkte zu diskutieren, auf so einem Mittelwert dann immer rekurren zu können. – Aber das jetzt nur, um die Diskussion ein wenig zu sortieren, die wir in der letzten halben Stunde hatten. Wir haben weniger über den Wirkstoff gesprochen, sondern über die Methodik. Aber das eine ist untrennbar mit dem anderen verbunden, weil wir eben nur bei diesen beiden Indikatoren einen Unterschied sehen.

Jetzt schaue ich mal in die Runde. Wen haben wir noch? – Keinen mehr. Dann ist das gut. Dann sind wir genau in der Zeit. Dann würde ich dem pharmazeutischen Unternehmer die Möglichkeit geben – Herr Wasmuth, ich nehme an, Sie machen das –, kurz zusammenzufassen, was wir in der letzten Stunde diskutiert haben, und dann könnten wir die Anhörung beenden. – Herr Wasmuth, oder auch ein anderer, wenn es gewünscht wird.

**Herr Dr. Wasmuth (Novartis):** Das übernehme ich wieder, Herr Professor Hecken. Vielen Dank. – Ich denke, wir haben in der letzten Stunde sehr ausführlich über die Relevanzschwellen diskutiert und dabei am Ende auch das Methodenpapier des IQWiG, das, wie schon gesagt, noch keine Gültigkeit hatte, als wir das Dossier eingereicht haben. Aber ich denke, wir konnten hoffentlich deutlich machen, auch zusammen mit den medizinischen Fachexperten, dass die Schwellenwerte, die wir in unserem Dossier verwendet haben, diejenigen sind, die sich in der Literatur finden, die im Alltag auch verwendet werden und – Sie haben es angesprochen – die tatsächlich vom G-BA in anderen Verfahren schon so akzeptiert wurden.

Worüber wir heute nicht gesprochen haben – dazu würde ich noch gerne einen Satz am Ende sagen –, ist das Thema Evidenztransfer. Hier wurde aus unserem Dossier und auch aus der Stellungnahme deutlich, dass wir in der Teilpopulation und in der Gesamtpopulation sehr vergleichbare Patientenkollektive haben. Wir haben eine Effektrichtung, die gleich, teilweise sogar identisch ist, sodass wir denken, dass man die positiven Effekte, die man in der Lebensqualität gesehen hat bei AQLQ und SGRQ unter Verwendung der heute viel diskutierten Schwellenwerte, übertragen kann und damit am Ende einen Zusatznutzen für alle Patienten mit stark symptomatischem unkontrolliertem Asthma, für die Enzair zugelassen ist, ableiten kann. – Vielen Dank dafür.

**Herr Prof. Hecken (Vorsitzender):** Ganz herzlichen Dank, Herr Wasmuth, für die Zusammenfassung. Herzlichen Dank an die klinischen Experten, an die Vertreter der Firma, an die Vertreter des IQWiG und die Bänke für die Teilnahme an der Diskussion. Wir werden das, was hier gesagt worden ist, selbstverständlich in unsere Bewertung einbeziehen müssen, wir werden es diskutieren müssen. Damit ist



diese Anhörung beendet. Ich wünsche Ihnen frohe Weihnachten, einen guten Rutsch, und es kann eigentlich im nächsten Jahr nur besser werden; hoffen wir es jedenfalls. Dann schauen wir, dass wir bis zur Impfung oder bis der Impfstoff wirkt, alle gesund bleiben. Alles Gute und herzlichen Dank, dass Sie heute da waren. Die Anhörung ist beendet.

Schluss der Anhörung: 12:02 Uhr