Podiumsdiskussion auf dem 9. MVF-Fachkongress „Real-World-Evidenz“

Behring: „ABDE ist nicht für alle Arzneimittel sinnvoll“

Bei der vormittäglichen Podiumsdiskussion stellten sich den Fragen von Moderator Prof. Dr. Dr. Alfred Holzgreve die Vortragenden der ersten Runde: Thomas Müller, Leiter der Abteilung 1 des BMG, Dr. Antje Behring, kommissarische Leiterin der Abteilung Arzneimittel des Gemeinsamen Bundesausschusses (G-BA), und Prof. Dr. Jürgen Windeler, Leiter des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG).



>> **Holzgreve:** Der Begriff „Real-World-Data“ polarisiert. Für die einen ist es ein überflüssiger Quatsch, für die anderen die Lösung vieler Probleme. Nun stellt sich die Frage: Wo liegt die Wirklichkeit? Wer wird sich durchsetzen und was können wir mit diesen Daten machen? Und vor allem auch, mit welchem Ansatz: mit dem einer allgemeinen Datensammlung à la Big Data – es wird alles gesammelt und dann wird gesehen, was damit möglich ist. Oder sollte man sich gerade bei der begleitenden Datenerhebung nicht schon vorher genau überlegen, welche Ziele man hat und welche Daten in welcher Qualität dafür benötigt werden? Oder wäre es gar besser, gleich von allen neu zugelassenen Medikamenten zu fordern, in der Realversorgung Daten zu erheben, um in fünf oder auch zehn Jahren einen Großteil der Fragen, die ganz sicher aufkommen werden, beantworten zu können?



Behring: Ich bin nicht der Meinung, dass man jedes Arzneimittel mit einer Anwendungsbegleitenden Datenerhebung, ABDE, belegen muss. Der Grund dafür ist, dass es nicht für alle Arzneimittel sinnvoll ist, nach der Markteinführung weitere Daten zu erheben, wenn bereits durch die Zulassung in sehr guten RCTs entsprechende Daten vorliegen, die wenig Zweifel daran lassen, was ein Arzneimittel bringt oder eben nicht bringt. Das neue Instrument der Anwendungsbegleitenden Datenerhebung, das dem G-BA an die Hand gegeben worden ist, hat auch eine gewisse Schärfe. Wichtig wird sein, dass vorab ganz genau definiert, welche Evidenzlücken es möglicherweise gibt. Optimal wäre es natürlich, wenn man alle diese Fragen schon vor der Zulassung beantworten könnte.

Holzgreve: Herr Müller, welche „Real-World-Data“ kommen denn für Sie überhaupt infrage? Mit allen möglichen Daten kann man die irrwitzigsten Korrelationen bilden, angefangen dabei, ob die Mortalität steigt, wenn das OP-Personal blaue oder rote OP-Wäsche trägt.



Müller: In der Zukunft, wenn wir einmal eine elektronische Patientenakte in Deutschland haben werden, wäre das tatsächlich ein Instrument, um zum Beispiel Fragen zu Mortalität und Nebenwirkungen, auch Fragen der Nutzenbewertung erfassen und beantworten zu können. Deshalb liegt ein Schwerpunkt der Arbeit bei uns im Ministerium darin, erst einmal die Voraussetzung dafür zu schaffen. Sicher stimmt es, dass, wenn man genauer hinschaut, die Datenqualität oft nicht so toll ist. Das sehen wir bei den medizinischen Tumorregistern, bei denen wir ganz große Mühen haben, die Basisdaten zu erfassen. Da genügt es nicht, dass lediglich erfasst wird, dass ein Patient gestorben ist, sondern man muss wissen, in welchem Stadium, unter welchen Therapien. Da sind wir in Deutschland tatsächlich noch nicht so weit. Gerade darum haben wir bei der ABDE die Anwendung begrenzt und wollen, dass der Bundesausschuss damit auch verantwortungsvoll umgeht. Es ist aus Sicht des Ministeriums nicht so ohne, wenn wir die Therapiefreiheit des Arztes einschränken, weil er ein neues Medikament nicht mehr verordnen darf, wenn er nicht an der Datenerhebung teilnimmt. Wir halten das für gerechtfertigt und haben das auch durch die Ressorts gebracht, auch weil wir glauben, dass es ein sehr scharfes Instrument ist. Gleiches gilt für den Punkt Randomisation oder nicht. Ich verstehe ja, dass das aus der Perspektive der Methodiker etwas anders aussieht, weil natürlich der Vergleich in der Randomisierung qualitativ besser ist. Aber im Ministerium müssen wir praktisch überlegen, wie ein solches Instrument eingesetzt werden kann und ebenso müssen wir schauen, das im Verhältnis steht. Es kann aus Sicht der politischen Linie nicht so sein, dass, wenn ein neues Arzneimittel zur Verfügung steht, ein Patient gefragt wird, ob er sich in eine RCT einschließen lassen will und je nachdem, wie der Würfel fällt, ist er dann in der Gruppe mit dem neuen Arzneimittel oder eben nicht.

Holzgreve: Das sprach mir wirklich aus dem Herzen, weil ich auch schon zum älteren Semester gehöre. Herr Windeler, Sie haben in Ihrem Vortrag die alte Kategorie der Anwendungsbeobachtung erwähnt, die früher eine gute Zusatzeinnahme war. Doch wir alle wussten eigentlich, dass das ein Mäntelchen für den Verkauf war und waren eigentlich froh, dass die AWBs endlich abgeschafft worden sind. Sie haben ebenfalls angedeutet, dass sie es eigentlich für ausgeschlossen halten, dass Real-World-Data etwas an den Ergebnissen randomisierter Studien, wenn sie ordnungsgemäß durchgeführt worden sind, ändern.



Windeler: Meine Hoffnung stirbt sowieso nicht, dass der G-BA unter dem Begriff Anwendungsbeobachtungen etwas anderes und sinnvolles gemacht bekommt als bisher. Den Begriff an der Stelle zu verwenden, war sicherlich für dieses Ziel nicht sehr zuträglich. Ich meine, dass die Industrie genau weiß, was Anwendungsbeobachtungen sind, was man da macht und was man da alles angeblich nicht machen darf. Wenn der G-BA sich an diese Begriffswelt meint gebunden zu fühlen, dann sehe ich da nicht sehr viel Perspektive. Aber vielleicht kommen wir gemeinsam zu guten neuen Ideen, was vergleichende Anwendungsbeobachtungen angeht. Als zweiten Punkt eine kurze Bemerkung zu Real-World: Das ist eines der Buzzwords, die überhaupt nichts erklären. Das ist ein Schlagwort, das ungefähr zehn verschiedene Definitionen hat. Der entscheidende Punkt ist der, dass der Begriff Real-World praktisch bedeutet, dass Zulassungsstudien nicht in der realen Welt entstanden sind. Die Patientenvertreter



„Je kleiner die zu erwartenden Unterschiede der Therapieeffekte sind, desto wichtiger ist der faire Vergleich."
Prof. Dr. Jürgen Windeler, Leiter des IQWiG

müssten auf die Barrikaden steigen und sagen: Was macht Ihr da eigentlich, uns mit Laborstudien Zulassungen zuzumuten, die in unserer Versorgung ja eingesetzt werden sollen? Das ist ein Begriff, der einem überhaupt nicht weiterhilft. Jetzt zu den randomisierten Studien: Ich habe nicht gesagt und werde das auch nicht sagen, dass es ausgeschlossen ist, mit Real-World-Data für randomisierte Studien zusätzliche ergänzende Erkenntnisse zu bekommen. Wenn ich schon eine Studie habe, dann kann das, was mit Real-World hinzukommt, immer nur Ergänzung zu dem sein, was ich habe. Real-World-Data meint Daten aus der realen Versorgung, was über das Studiendesign noch gar nichts aussagt. Meine entscheidende Anforderung an diese Daten im Zuge der Nutzenaussage ist ein aussagefähiger, fairer Vergleich. Da gibt es statistische Verfahren, die das an randomisierte Studien annähern können. Jeder Methodiker und Theoretiker weiß, erreichen werden wir sie nie, aber das ist praktisch vielleicht gar nicht so schlimm. Wenn Sie mit Propensity Scores, Adjustierung und sonstigen Dingen arbeiten wollen, bedeutet das die Anforderung an die Datenqualität dieser sogenannten Real-World-Daten, die Sie mit Routinedaten schon mal komplett vergessen können und für die Sie sonst, und deswegen weise ich die Industrie mit Nachdruck darauf hin, einen großen Aufwand betreiben müssen.

Stegmaier: Welches Evidenzlevel oder welche Evidenzqualität reicht denn aus? Wie wird das IQWiG denn reagieren, das in der Vergangenheit schon soundso viele RCT-Studien ablehnt, weil sie die Mindestqualität nicht erreicht haben? Wie wird man dann mit Real-World- Data umgehen?

Windeler: Es wird darauf ankommen, ob wir aussagefähige Vergleiche haben. Ich sage an dieser Stelle ausdrücklich, dass diese aussagefähigen Vergleiche nicht immer aus RCTs stammen müssen, aber es müssen aussagefähige Vergleiche sein. Die Frage, was aussagefähige Vergleiche sind, bedeutet, welche Anforderungen an die Adjustierung von Störvarianten gestellt werden. Das hängt wiederum von der Größe der Effekte ab. Wenn Sie sehr große Effekte haben, brauchen Sie sich um die Adjustierung von vornherein so gut wie gar nicht zu kümmern. Da können Sie einfach sagen, passt schon. Wenn Sie mittlere Effekte haben, dann werden Sie sich sorgfältiger darum kümmern müssen. Und wenn Sie kleine Effekte haben, werden Sie sich ganz sorgfältig kümmern müssen. Ich gebe Ihnen eine persönliche Aussage: Ich kann niemandem empfehlen, für kleine Effekte keine randomisierten Studien zu machen. Wahrscheinlich ist es, dass sich die Gesetzgebungssituation und auch die Entscheidung des G-BA möglicher- und sinnvollerweise auf Situationen konzentrieren werden, wo es eben nicht um kleine Effekte geht. Weil eben da andere Designs möglicherweise auch durchaus Sinn machen.

Behring: Ein Kommentar dazu. Es liegt nicht daran, dass rando-

misierte kontrollierte Studien keine gute Qualität haben. Sie haben einfach nicht die Fragestellung beantwortet, die wir haben wollen. Es ist total wichtig, dass man sich genau überlegt, für welche Fragestellung man diese oder jene Studie machen will. Und dass man dann die richtige Datenquelle für die richtige Fragestellung wählt. Es geht nicht darum, dass die Sachen abgelehnt worden sind, weil die Datenqualität im RCT nicht passt, sondern weil sie irgendwie an der Zielstellung vorbeigegangen sind.

Holzgreve: Es ist ein tatsächlich erlebtes Phänomen, dass Patienten bei Patient-Reporting-Outcome-Studien Patienten irgendwann merken, es ist besser, wenn sie sagen, es geht mir gut, weil dann nämlich keiner kommt und sie belästigt. Das heißt, dass man unheimlich aufpassen muss, ob die gesammelten Daten valide sind.



Griesinger: Wie stellen Sie sich denn die zeitgleiche Randomisierung von Parallel- oder Vergleichsgruppen vor? Herr Müller hat gerade gesagt hat, dass das eigentlich bei bereits zugelassenen Medikamenten schwer vorstellbar ist. Nehmen wir ein Medikament, das eine 90-prozentige Ansprechrate hat, da wird man keine zeitgleiche Randomisierung in Real World oder in Registern fordern können. Oder gibt es nicht vielleicht doch die Möglichkeit, wenn es eine gute Datenqualität in einer Kohortenstudie gibt, in der Confounder definiert wurden und in der auch historische, gut klinisch beschriebene Patientenkollektive als entsprechende Vergleichsgruppen zu identifizieren sind.

Windeler: Der entscheidende Punkt ist ein aussagefähiger Vergleich. Wenn Sie in einer Situation sind, dass Sie im Zusammenhang mit Zulassungen und mit dem AMNOG auf der Basis von randomisierten Studien oder auch der Basis ganz anderer Daten zu der Überzeugung kommen, dass es ein überzeugendes Medikament mit einem beträchtlichen Zusatznutzen ist, stellt sich die Frage weiterer Studien gar nicht mehr. In dem Moment, wo man eine zuverlässige Bewertung eines Arzneimittels oder einer Intervention machen kann, stellt sich die Frage nach randomisierten Studien nicht mehr. Übrigens auch nicht die Frage nach anderen Studien. Anders hingegen bei Situationen, bei denen man bezüglich hoffentlich relevanter Fragen im Verhältnis der neuen Therapie zu dem bisherigen Standard unsicher ist. In dieser Situation grundsätzlich zu randomisieren ist grundsätzlich kein Problem, aber ist praktisch ein Problem. Deswegen sage ich noch einmal: Der faire Vergleich ist das Entscheidende. Der kann im Extremfall und in besonderen Fällen auch historisch sein. Historisch im Sinne von vielleicht nicht 1920, aber vor Einführung des Medikamentes, aber bitte: ein fairer Vergleich. Der ist meiner Ansicht nach gut möglich, aber hängt von der kompletten Situation ab. Der entscheidende Punkt – den ich kritisiere und in meinen Augen ein Anachronismus ist – ist der, an dieser Stelle Randomisierung zu untersagen.

Holzgreve: Wenn es auf den Zeitfaktor ankommt, weil man entscheiden muss, würden Sie zustimmen, dass man dann alles an verfügbaren Daten nimmt, selbst wenn die Evidenz nicht so groß ist oder wenn Sie gar nicht wissen, wie die Daten erhoben worden sind?

Müller: Das ist eben die Diskussion, was die Politik vom Bundesausschuss erwartet und was in der internen Diskussion im Bundesausschuss relevant ist. Wir möchten doch nur, dass beim Bundesausschuss mit dem Aufwand, den er als Spitzenorganisation der Selbstverwaltung mit allen Experten, Patientenvertretern und Gesellschaften betreibt, dieses Signal ankommt: Was können Patienten von einem neuen Arzneimittel erwarten? Wenn Sie dann zufügen, dass eine niedrige Aussagesicherheit besteht, weil eben zum Beispiel eine Kohortenstudie vorliegt und kein RCT, dann ist das genau das, was im Gesetz angelegt ist. Nur wenn eine Tendenz feststellbar ist, dass bei den Orphans und anderen Arzneimitteln in Richtung 100 Prozent „nicht quantifizierbar" herauskommt, dann werden wir auf der anderen Seite in den Empfehlungen der Fachgesellschaften und in den Diskussionen mit den Facharztgruppen völlig andere Einschätzungen bekommen. Wir können im Ministerium nicht entscheiden, wer recht hat, nur diese Diskrepanz ist auf Dauer für den Gesetzgeber problematisch, darauf weise ich nur hin. Wir können uns jetzt nicht in die Position zurückziehen und sagen, also wir – damit meine ich jetzt tatsächlich das IQWiG und den Bundesausschuss – wollen das RCT. Die Zulassung hat eine andere Entwicklung genommen, das muss man irgendwo akzeptieren. Ich verstehe, dass man an dem RCT als Goldstandard hängt, und der wird auch in vielen Indikationen weiter Goldstandard bleiben, aber wenn man in einer Indikation einen Mechanismus versteht oder eine Intervention wie zum Beispiel eine genetische Intervention hat, dann brauchen wir andere Instrumente. Dann erwarten Patienten auch, dass man Ihnen dazu eine Aussage gibt. Das ist das, was wir versuchen, zu befördern. Ich glaube aber, dass zum Beispiel Real-World-Daten in der Zukunft tatsächlich Vergleichsmöglichkeiten erlauben. Das sind einfach biometrische Weiterentwicklungen.

Windeler: Die 100 Prozent stimmen natürlich höchstens für ganz spezifische Orphan-Drug-Gruppen, sicherlich nicht für die normalen Bewertungen. Ich glaube, der Punkt ist, der mich ein bisschen irritiert, dass Sie, Herr Müller, am Anfang Ihres Vortrages, sehr nachvollziehbar beschrieben haben, wie schwierig die Situation durch die immer weiter vorverlagerten Zulassungen mit immer unreiferen Daten ist. Was sollen wir denn machen? Die Konsequenz ist doch, dass man über diese Arzneimittel nichts Konkretes sagen kann. Und das ist dann in der Arzneimittelnutzenverordnung mit „nicht quantifizierbar" abgebildet: Wir wissen es einfach nicht. Insofern gibt es eine sehr enge Verbindung zwischen den immer unreiferen Daten, die dazu führen, dass die Aussagen des IQWiG und die Aussagen des G-BA immer unpräziser werden müssen und immer achselzuckender – wenn ich das einmal so etwas flapsig sagen darf. Dabei geben wir uns große Mühe, möglichst nicht „nicht quantifizierbar" darauf zu schreiben. Aber wenn die Daten nicht da sind, sind die Daten nicht da.

Müller: Das ist eine schwierige Situation, aber auch eine, die begründet ist im Konsens der internationalen Zulassungsbehörden. Es ist eine Situation, bei der wir eine Entscheidung auf europäischer Ebene haben, dass damit ein neues Medikament die Patienten erreichen soll. Diese Diskussion um „Patient Access" wird eingeschränkt durch Pricing-Probleme in vielen Ländern, doch gibt es auf der medizinischen Seite überhaupt keine kontroverse Diskussion darüber, dass wir diese Beschleunigung in vielen Indikationen benötigen. Wir sehen auch nicht, dass das zu einem vermehrten Rückzug und Widerruf von Zulassungen führt. Natürlich gibt es Beispiele, bei denen wir korrigieren müssen, weil wir eben eine geringe Datenlage haben. Aber das das ein kompletter Irrweg ist, würde ich aus der Zulassungsdichte nicht bestätigen. Wir müssen uns eher im HTA-Bereich

überlegen, wie wir damit umgehen, um auch weiterhin das Ziel zu erfüllen, für Gesetzgeber Preisverhandlung und Patienten Aussagen zu neuen Arzneimitteln zu treffen. Das ist jetzt die Herausforderung. Das bedeutet auch, dass man sich von Konzepten, die man wie das RCT vor elf Jahren als Standard gesetzt hat, verabschieden muss. In einigen Bereichen gibt es biometrische Ansätze, wie man sich über große Datenmengen einer Validität nähert, auch wenn man sie vielleicht nicht erreicht.

„Wir werden im Rahmen der Ratspräsidentschaft die Orphan-Gesetzgebung evaluieren und neu fokussieren."
Thiomas Müller, Leiter der Abteilung 1 des BMG

Sprecher: Orphan Drugs haben einen besonderen Zulassungsstatus, weil man bereit ist, eine größere Unsicherheit in Kauf zu nehmen. Diese größere Unsicherheit wird im Prinzip in den G-BA und anderen Verfahren nicht abgebildet. Wäre es im Prinzip nicht sinnvoll, eine Stufe unter dem Anhaltspunkt einzuführen, um zu versuchen tiefer zu klassifizieren.

Windeler: Was wäre denn noch weniger Anhaltspunkt als ein Anhaltspunkt. Da könnte man sicher darüber diskutieren.

Müller: Wir werden im Rahmen der Ratspräsidentschaft die Orphan-Gesetzgebung evaluieren und neu fokussieren. Wir wollen, dass Orphans substanzielle Fortschritte für die Patienten bringen. Das ist das Anliegen. Deswegen ist das, was wir machen, kein IQWiG- oder G-BA-Bashing, sondern wir wollen eine Filterfunktion. Es gibt eben Orphans, die ganz tolle Fortschritte für Patienten bringen, aber es gibt auch solche, bei denen wir daran zweifeln. Nur kosten die dann 100.000 Euro und mehr und haben Nebenwirkungen und der Patient muss zudem jede Woche zum Arzt. Diese Differenzierung geht ein wenig unter, indem wir immer „nicht quantifizierbar" sagen. Das ist der Punkt.



Hoffmann: Wir haben keine fundamentale Diskussion, sondern egentlich die Anerkenntnis der Tatsache, dass wir hier keine Sprünge in der Evidenzbewertung haben, sondern ein kontinuierliches Spektrum. Wir müssen uns jenseits des ganz Klaren Gedanken darüber machen, wo es besser ist, dass man etwas mehr weiß als gar nichts. Das ist etwas, das mich als Methodiker und Versorgungsforscher extrem freut, auch die gewisse Nüchternheit, mit der man an die Fragestellung herangeht. Auf der einen Seite erkennt man an, dass es noch andere Dinge unterhalb der RCT-Schwelle gibt, doch bedeutet das auf der reziproken Seite eine sehr hohe Verantwortung, die wir ja damit haben - auch der Industrie! Das, was wir früher mit den Anwendungsbeobachtungen gemacht haben, war völlig unverantwortlich, unmoralisch und unethisch. Deswegen haben wir es auch abgeschafft. Heute sind wir aber in einer anderen Situation, in der man ehrlich zugeben muss, unsicher zu sein, aber bestimmte Entscheidungen trotzdem zu fällen, weil es genügend gute Argumente gibt. Aber dann muss man hinterher auch bereit sein zu korrigieren. Das ist die neue Situation, bei der es keine fundamentalen Sieger oder Verlierer gibt. Wir müssen aber auch ganz klar sagen: Wenn der rechtliche Datenschutzrahmen so bleibt wie im Moment, dann werden wir auch weiterhin diese Daten nicht haben oder unter Umständen sogar noch schlechtere. Wenn wir aber anfangen zu sagen, hier werden Patienten nicht gut behandelt, weil man bestimmte Daten nicht haben darf, dann ist die Abwägung Daten- gegen Lebensschutz eine neue. Die Frage dürfen Datenschützer dann alleine nicht mehr beantworten, sondern die sind Partei. Die größere Bedeutung von Daten aus der sogenannten realen Welt oder der Versorgung erzwingt, dass man Abstriche dabei macht, einen absoluten Datenschutz betreiben zu wollen. Ich empfehle, die Ratspräsidentschaft auch dafür zu nutzen.

Holzgreve: Ich glaube, das kann man so stehen lassen. Sie haben eben gesagt, dass Krebsregister sogar gesetzlich geregelt sind und trotzdem haben wir damit viele Probleme und können im Grunde genommen nichts damit machen. Macht es überhaupt Sinn Register auf eigene Initiative zu schaffen? Gibt es gute Register?

Windeler: Es gibt gute Register. Wir haben im Rahmen des Auftrages vom G-BA eine ganze Reihe von Registerbetreibern gesprochen, die wirklich ausgesprochen gute Register haben. Die nenne ich jetzt aber nicht, weil sie mit Inhalt unseres Berichtes sein werden. Das Problem mit guten Registern ist weniger der Datenschutz, sondern die Qualität. Sie müssen eine strukturelle Qualität liefern, was schon mit der Vollständigkeit und solche banalen Dinge angeht. Es gibt durchaus Register, mit denen man wirklich etwas anfangen kann. Aber es gibt auch eine ganze Reihe von Registern, mit denen man derzeit noch nicht sehr viel anfangen kann.

Sprecher: Ich habe noch eine kurze Frage zum Gesetzgebungsverfahren: Welche Patientengruppen oder Medikamente sind betroffen? Sie hatten Orphan Drugs in Ihrem Vortrag genannt, hatten dann aber auch Präzisionsmedizin gesprochen. Bedeutet das auch, dass es nicht nur die reine Orphan-Disease-Gruppe sein wird, sondern auch genetisch definierte Patientengruppen inkludiert sind?

Müller: Der Gesetzestext richtet sich an Orphans und an Zulassungen, bei denen Daten fehlen. Das sind die Zulassungen wie Conditional Approval und Exceptional Circumstances. Insofern sind da auch noch andere Gruppen inkludiert. Bei Conditionals sind typischerweise jene Fälle gemeint, bei denen noch Daten ausstehen, weil die Phase-3-Studien noch weiterlaufen. <<

Zitationshinweis

Stegmaier, P.: „Behring: ABDE ist nicht für alle Arzneimittel sinnvoll" (06/19), S. 36-39, doi: 10.24945/MVF.06.19.1866-0533.2188

Link zu den Vorträgen, die den Teilnehmern des Kongresses zum Download zur Verfügung stehen:
https://www.monitor-versorgungsforschung.de/kongresse/RWE_2019/referenten