

Das Plenum im Scharoun-Saal der AOK Nordost

9. MVF-Fachkongress zu Gast bei der AOK Nordost mit dem Thema „Real-World-Evidenz“

Windeler: „Ohne Vergleich geht gar nichts“

Der Erste gibt das Ziel vor. Die Zweite beschreibt den Weg dahin und der Dritte benennt Fährnisse und Hindernisse auf demselben. Der Zielvorgeber auf dem 9. MVF-Fachkongress „Real World Evidenz“ im Scharoun-Saal der AOK Nordost war Thomas Müller, Leiter der Abteilung 1 und damit innerhalb des BMG zuständig für Arzneimittel, Medizinprodukte und Biotechnologie und jener Bundesbeamte, der das vom Deutschen Bundestag beschlossene Gesetz für mehr Sicherheit in der Arzneimittelversorgung (GSAV) mit vorbereitet hat, in dem unter anderem der Weg für die Anwendungsbegleitende Datenerhebung frei gemacht wurde. Die Wegbeschreibende auf dem ebenso aktuellen wie inhaltlich dichten Kongress war Dr. Antje Behring, kommissarische Leiterin der Abteilung Arzneimittel des Gemeinsamen Bundesausschusses (G-BA) und damit Nachfolgerin von Müller, der vor rund eineinhalb Jahren zum BMG gewechselt war. Der Dritte im Bunde und in der Runde war Prof. Dr. Jürgen Windeler, seit September 2010 Leiter des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Ausgerechnet er, der im Titelinterview mit „Monitor Versorgungsforschung“ (1/18) vor zwei Jahren noch erklärt hatte, dass man mit guten Registern zwar Fragen zur Adhärenz der Patienten, zum Verschreibungsverhalten der Ärzte, zu Kontrolluntersuchungen und vielem anderen mehr beantworten könne, sie jedoch für Nutzenfragen nicht geeignet seien, muss den Weg dahin erarbeiten – im Auftrag des G-BA.

>> Hintergrund der nun von der Politik eingeführten Anwendungsbegleitenden Datenerhebung, die im G-BA Sprachgebrauch bereits kurz ABDE genannt wird, ist laut Müller der deutlich zunehmenden Zahl beschleunigter Zulassungsverfahren geschuldet. Bei diesen Verfahren ermöglichten die Zulassungsbehörden bei einer zu diesem Zeitpunkt begrenzten und oft auch unreifen Datenlage aufgrund einer reinen Abwägungsentscheidung den Marktzugang. Dies führe unter anderem dazu, dass das BMG die daraus resultierende Tendenz bei den Beschlüssen des Bundesausschuss sehe, dass die Kategorie „nicht quantifizierbar“ stark zunehme. Dies bezeichnete Müller als einen „Schmerzpunkt“. Er könne diese – aufgrund der geringeren Datenlage zustande gekommenen – Beschlüsse zwar verstehen, doch entsprächen diese nicht dem Auftrag des Gesetzgebers an den Bundesausschuss, „der bei neuen Arzneimitteln eine Einschätzung für Patienten und für die Preisverhandlung geben sollte, mit welcher Wahrscheinlichkeit ein zu erwartender Effekt eintritt oder nicht“.

Ausgehend davon, dass in der Arzneimittel-Nutzenbewertungs-Verordnung sechs Evidenzstufen aufgelistet sind, erwartet das BMG laut Müller, dass der Bundesausschuss mit dem IQWiG gemeinsam eine Methodik findet, diese Nutzenkategorien in Beschlüssen auch abzubilden, um so die Effekte, die Patienten von neuen Arzneimitteln zu erwarten haben, quantifizieren zu können. Ganz deutlich wandte er sich direkt an den G-BA, indem er wortwörtlich sagte: „Wenn im Bundesausschuss für viele Produkte immer nur ein ‚nicht quantifizierbar‘ beschlossen wird, fehlt an dieser zentralen Stelle der Indikator und die Richtschnur, die sich der Gesetzgeber vom Bundesausschuss verspricht.“

Nun ist es nach Müller so, dass die Möglichkeiten, die fehlenden Daten generieren zu können, seitens der europäischen Zulassung relativ schwach ausgeprägt seien. Genau das soll das seit dem 16. August dieses Jahres mögliche ABDE ändern. Das erklärte Ziel: Über eine Verknüpfung mit den Leistungserbringern und der Real-Versorgung ein neues, „sehr gutes Datenniveau zu generieren, um dann auf dieser Grundlage möglichst gute Entscheidungen zu treffen“. Aus diesem Grunde sei das Instrument der ABDE, so Müller „zugegeben schmerzhaft“, eingeführt worden. Dies sei aber auch deswegen geschehen, weil es Folgen in der Preisverhandlung mit möglichen Marktaustritten gegeben habe, eben weil der G-BA in seinen Entscheidungen nicht zu einer Quantifizierung gekommen sei. Müller: „Wir wissen, dass das ein Einschnitt ist, aber wir hoffen, dass beide oder alle Seiten damit verantwortungsvoll umgehen.“ Das, so Müller, wäre zumindest die Idee, ob diese funktioniere, müsse nun das System finden.

Dieses „System“ wird in erster Linie vom G-BA, in zweiter vom IQWiG gebildet. In ihrem Vortrag, den Dr. Antje Behring mit dem Titel „Sind bedingte Zulassungen reif für die Frühe Nutzenbewertung?“ über- und mit dem Untertitel „Umgang mit Evidenzlücken bei der Zusatznutzenbewertung“ unterschrieben hatte, antworte sie durchaus sybillisch: „Die Frage möchte ich ganz klar beantworten – mal ja, mal nein oder mal mehr, mal weniger.“ Das Problem, das die Politik mit dem Vorstoß der ABDE angehen wolle, sei indes kein neues, denn in der frühen Nutzenbewertung seien schon verschiedene Arzneimittel im Rahmen der bedingten Zulassung bewertet worden.

Einer der ersten Wirkstoffe war den Worten Behrings zufolge Crizotinib, bei dem schon 2013 im ersten Verfahren bei einer bedingten Zulassung vergleichende Studien vorgelegen hätten. Damals wäre der auswertbare Datenschnitt ausreichend genug gewesen, um einen Zusatznutzen zu quantifizieren. Als dann im zweiten Schritt die betreffenden Studien abgeschlossen und damit die Vollzulassung erreicht worden sei, hätte der Zusatznutzen bestätigt werden können.

Anders beim Wirkstoff Ceritinib, bei dem 2015, zum Zeitpunkt der ersten Nutzenbewertung, lediglich eine einarmige Studie vorgelegen hätte, was laut Behring dazu geführt habe, dass der Zusatznutzen mit einem „nicht belegt“ festgestellt worden sei. Beim „full approval“ im Jahre 2017 allerdings hätte durch die dann vorliegende vergleichende



Studie ein „beträchtlicher Zusatznutzen“ zuerkannt werden können.

Ein ähnliches Bild hätte sich auch beim Wirkstoff Osimertinib gezeigt, der zuerst 2016 mit einem „nicht belegten“ Zusatznutzen, 2017 dann allerdings mit einem „beträchtlichen“ aus der frühen Nutzenbewertung gegangen sei. Das Problem sei hier gewesen, dass nach der Erstbewertung, bei der nur Daten aus einarmigen Studien vorgelegen hätten, das in Markt bringende Pharmaunternehmen entschieden hatte, das Arzneimittel vom Markt zu nehmen. Was nichts anderes bedeutet hätte, als dass bis zur erfolgreichen Zweitbewertung der Wirkstoff in Deutschland – bis auf Einzelfälle – nicht verfügbar gewesen war.

Als weitere Beispiele für Nutzenbewertungen nach bedingter Zulassung führte Behring in ihrem Vortrag Blinatumumab (2016 „nicht quantifizierbar“, 2017 „beträchtlicher Zusatznutzen“) und den Wirkstoff Alectinib an, bei welchem 2017 ein Anhaltspunkt für einen „geringen Zusatznutzen“ beschlossen worden sei. Behring: „Alectinib ist ein Beispiel dafür, dass zum Zeitpunkt des Conditional Marketing Authorisation ausreichend Studien vorhanden waren, um einen Zusatznutzen zu quantifizieren.“ Hier sei es so gewesen, dass zwar zum Zeitpunkt der Zulassung die entsprechenden Phase-III-Studien noch nicht abgeschlossen waren, allerdings zum Zeitpunkt der AMNOG-Bewertungen vorgelegen hätten, weswegen diese dann auch berücksichtigt werden konnten.

Anders liegt laut Behring der Fall beim Wirkstoff Olaratumab, dem 2017 ein „beträchtlicher Zusatznutzen“ zuerkannt wurde. Hier sei die zum Zeitpunkt der Zulassung vorliegende Evidenz eigentlich sehr überzeugend gewesen, da in einer vergleichenden Phase-II-Studie sehr überzeugende Überlebensvorteile gezeigt wurden. Behring: „Man wäre im Leben nicht darauf gekommen, dass diese Daten nicht bestätigt werden können.“ So allerdings geschehen in der folgenden Phase-III-Studie („Announce“), welche die ursprünglich beschriebenen Überlebensvorteile nicht bestätigen konnte, worauf 2019 die Zulassung entzogen wurde.

Dies zeigt Behring zufolge eine den jeweils vorliegenden Datenlagen inhärente Unsicherheit. Sie gab offen zu, nicht zu wissen, was passiert wäre, wenn bei Olaratumab die Phase-III-Studie nicht beauflagt worden wäre, sondern stattdessen Daten aus der Versorgung herangezogen worden wären, die vielleicht andere Erkenntnisse ergeben und unter Umständen dazu geführt hätten, dass am Ende die Zulassung nicht zurückgezogen werden musste.

Auch das weist erneut auf das Problem der bedingten Zulassungen mit ihren begrenzten Datenlagen hin, was Behring zu der Frage bringt, welche Art von neuer Evidenz nach der Zulassung überhaupt erforderlich ist. Hier formuliert sie die Gretchenfrage: „Wenn man davon ausgeht, dass bei bedingten Zulassungen die nötige Evidenz schon durch die Zulassung gegeben ist, wozu braucht man denn dann noch Evidenz aus versorgungsnahen Studien?“ Oder sollten nicht besser, mit besonderem Blick auf das Beispiel Olaratumab, von den Zulassungsbehörden weitere vergleichende Studien gefordert werden?

Da das Deutschland im weltweiten Kontext nicht alleine durchsetzen kann, hat die Politik reagiert und für Deutschland mit dem GSAV die ABDE eingeführt. Die bald scharf geschaltete Maßnahme für die anwendungsbegleitenden Datenerhebungen kann nach Behrings Meinung „auch vor der Entscheidung der frühen Nutzenbewertung, parallel oder auch danach erfolgen“. Dies sei nicht mit dem Konzept der Befristung zu verwechseln, in deren Rahmen pharmazeutische Unternehmer schon lange beauflagt werden konnten, bestimmte Daten zu erheben. Behring: „Die ABDE ist etwas, was möglicherweise losgelöst davon erfolgen kann.“ Zudem sei die Beauflagung der ABDE nach Willen des Gesetzgebers „möglich, aber nicht zwingend“.

Der G-BA müsse darum künftig, so Behring, sehr genau analysieren, wo es überhaupt sinnvoll sei, zusätzliche Datenerhebungen zu beauflagen. Behring: „Das ist sicherlich nicht der Fall bei Arzneimitteln, bei denen wir ge-



„In den nächsten 5 Jahren werden uns 1.500 Mitarbeiter verlassen, weil sie in den Ruhestand gehen. Wir werden viele Aufgaben, vor allem in der Sachbearbeitung, durch digitale Prozesse abbilden. Das ist eine gewaltige Herausforderung.“

Frank Michalak, Vorstandsvorsitzender der AOK Nordost



„Wir brauchen eine breitere Diskussion zu den Fragen ‚Wie kann Evidenz bei hochspezialisierten Arzneimitteltherapien aussehen?‘ und ‚Gibt es eine sinnvolle Balance zwischen interner und externer Validität.‘“

Daniela Teichert, designierte Vorstandsvorsitzende der AOK Nordost

nug Evidenz haben und mit den vorliegenden Daten schon quantifizieren können." In eine zweite Kategorie würden ihrer Meinung nach all jene Wirkstoffe fallen, die in den Zulassungsstudien nicht unbedingt patientenrelevante Endpunkte erforscht haben. „Große Meilensteine innerhalb einer Erkrankung" will der G-BA in den zu beauflagenden ABDE sehen, denn es gehe nicht um „kleine Veränderungen", die auch später in der Nutzenbewertung möglicherweise nicht mehr interessieren werden. Ihr Rat: „Man möchte gerne wissen, was ein Wirkstoff tatsächlich für den Patienten bringt; das heißt, dass man sich in der anwendungsbegleitenden Datenerhebung auf gute Endpunkterhebungen konzentrieren muss."

Ein Dilemma ist nach Behrings Ansicht der zeitliche Ablauf. Der Wunsch des G-BA wäre es, so früh wie möglich mit der Datenerhebung im Rahmen der ABDE zu beginnen. Das Problem sei, dass hier zuerst einmal die Evidenzlücke definiert und möglichst genau beschrieben werden müsse, „welche Daten man eigentlich noch braucht, um den Zusatznutzen besser quantifizieren zu können". Ihre Frage: „Muss das erst nach der Nutzenbewertung erfolgen oder könnte man nicht vorher gemeinsam die Zulassungsbehörden konsentieren, um herauszufinden, welche Evidenzlücke es eigentlich gibt?" Hier sieht sie die pharmazeutischen Unternehmen in der Pflicht darzulegen, welche möglichen Gaps es gebe, die „nicht gerade zum Blumentopf in der Nutzenbewertung führen".

Hier gelte es aber auch, die Zusammenarbeit zwischen dem G-BA und den Zulassungsbehörden zeitnah zu gestalten, um sich zu einem frühen Zeitpunkt darüber abzustimmen, welche Daten einerseits die Zulassungsbehörden und welche möglicherweise andererseits der G-BA zusätzlich fordert.

Ist die ADBE einmal beauflagt, werde es darauf ankommen, wo die geforderten Daten überhaupt herkommen. Über das „Wo" hat sich der G-BA Gedanken gemacht, Behrings Antwort: „von Registern". Sie weiß nun aber auch ganz genau, dass hier der Faktor Zeit entscheidend sein wird; insbesondere dann, wenn es zu einer bestimmten Datenfrage überhaupt noch kein Register und auch kein Konzept gibt, wie eine darauf aufsetzende anwendungsbegleitende Datenerhebung aussehen könnte, das – so Behring – „dauert das eine Weile". Als Beispiel hierfür erwähnte sie eine aufzusetzende Registerstudie zur Dokumentation bestimmter Car-T-Zell-Therapien, die auf Wunsch der EMA mit dem Register der „**E**uropean Society for **B**lood and **M**arrow **T**ransplantation" (Anm. d. Red.: seit 1974 das Rückgrat der Forschungs- und Bildungsaktivitäten des EBMT mit mehr als einer halben Million Patienten, die eine hämatopoetische Stammzelltransplantation erhalten haben) erfolgen soll. Hier sei schon zu einem sehr frühen Zeitpunkt begonnen worden, mit den Registerbetreibern zu sprechen, ebenso wäre seitens der Zulassungs- und HTA-Behörde ein enormer Aufwand betrieben worden, Stakeholder mit einzubeziehen – und dennoch sei es „unheimlich zäh und schwierig gewesen, die Registerstudie aufzusetzen". Ihr Fazit aus dieser Erfahrung: „Das heißt, dass es, wenn man eine anwendungsbegleitende Datenerhebung erst mit dem Zeitpunkt der Nutzenbewertung beauflagt, das möglicherweise zu spät ist." Hier gelte: „je früher, desto besser."

Eine weitere Frage stelle sich nach der Art der Register, „geht es in Richtung Produktregister oder Anwendungsbeobachtungen, um konkrete Fragestellung beantworten zu können". Ersteren erteilt sie gleich eine deutliche Absage. Behring: „Die kurzfristige temporäre Einrichtung von Produktregistern ist nicht genau das, was wir uns wünschen würden." Eher sollte ihrer Ansicht nach daran gearbeitet werden, wie Fragestellungen der frühen Nutzenbewertung in möglicherweise vorhandene krankheitsspezifische Register eingebaut werden könnten, auch damit nicht parallel verschiedene Datenquellen bedient werden müssen. Das Problem dabei sei, dass bei indikationsbezogenen Registern nicht alle Daten erhoben würden, die für bestimmte Fragestellungen innerhalb der Nutzenbewertung geeignet sind.

Windelers Komma und Aber

All das führt zur Frage des „Wie?". Der Gesetzgeber habe, so IQWiG-Leiter Windeler, einen „hochsinnvollen Vorstoß unternommen, die Datenerhebung zu stimulieren", doch nun komme „das Komma und dann das Aber". Die ABDE sei seiner Meinung nach, wobei das eine sehr diplomatische Formulierung sei, „noch nicht optimal ausgestaltet". Hier denkt er vor allem an die praktische Machbarkeit, Wirksamkeit und die dazu nötigen Entwicklungsmaßnahmen.

So erfordere der laut Gesetzesbeschluss zentrale Punkt der anwendungsbegleitenden Datenerhebung, nämlich Daten zur Quantifizierung des Zusatznutzens für eine Nutzenbewertung oder eine Zusatznutzenbewertung zu generieren, bestimmte Vorgehensweisen und ebenso bestimmte Mindestvoraussetzungen. Er sei etwas skeptisch, ob der G-BA die Möglichkeiten habe, mit den bestehenden gesetzlichen Rahmenbedingungen das zu tun, was eigentlich erforderlich ist, was insbesondere den Ausschluss randomisierter, klinischer Studien betreffe, wobei er „die Begründung dafür nicht verstanden" habe.

Aus den beiden voran gegangenen Vorträgen von Müller und Behring habe er verstanden, und könne das auch so unterschreiben, dass „offenbar die Zielrichtung und die ganz wesentliche Datenquelle für die anwendungsbegleitende Datenerhebung Register sein werden". Gegen den im Gesetzesentwurf verwandten Begriff der Anwendungsbeobachtung legte er jedoch sein Veto ein, den brauche kein Mensch, „weder den Begriff, noch die sogenannten Studien" und „schon gar nicht, um damit in der frühen Nutzenbewertung Nutzen zu quantifizieren". Auch gebe es im Gesetzestext – „sagen wir einmal, schwer nachvollziehbare Formulierungen" –, wie etwa, dass durch Fallberichte oder Expertenmeinungen Nutzen nachgewiesen werden könne, das sei doch „irgendwie eine ganz neue Erkenntnis".

Das größte Problem sieht Windeler allerdings in einer anderen Sachlage begründet. Bei der Schwierigkeit der Quantifizierung sei es nicht prioritär, „dass wir nicht genug über das neue Arzneimittel wissen", vielmehr sei es der entscheidende Punkt, „dass wir nicht genügend Daten über Vergleichstherapien und den spontanen Krankheitsverlauf haben". Das heiße, dass eine Fokussierung der anwendungsbegleitenden Datenerhebung auf ein neues Medikament „uns in der Frage des Wirkungsnutzens und der Quantifizierung des Nutzens nicht weiterbringen" werde. Daher seine Forderung: „Wir brauchen den Vergleich." Aus diesem Grunde brauche man statt produktbezogenen indikationsbezogene Daten. Nun sei es jedoch ein entscheidender Punkt, dass sich zu untersuchenden Patien-tenpopulation „momentan relativ eng und strikt an den Zulassungen" orientieren würden, was auch für die anwendungsbezogenen Datenerhebungen gelte. Wenn der G-BA diese Anforderung in der bisherigen Weise aufrechterhalte, dass auch in der anwendungsbegleitenden Datenerhebung bei Registern ein Zulassungsbezug nötig sei, scheine das bei all dem, was er über Register wisse, „in keiner Weise eine triviale Anforderung zu sein".

Gleiches gelte für Dosierungen und Dosierungsschemata. Wenn der G-BA bei diesen Punkten bestehende Anforderungen beibehalte, müssten diese auch in der an-

„Es sollten nach dem Prinzip der bestverfügbaren Evidenz auch schon beim Einreichen der Dossiers historische Kontrollen oder andere Beobachtungstudien hinzugezogen werden."
Friedhelm Leverkus, Director Health Technology Assessment & Outcomes Research von Pfizer Deutschland



„Lange Zeit existierte keine eindeutige semantische Definition von Kausalität. Zudem fehlte es an passenden Daten, um kausale Antworten zu generieren. In den letzten zwei Dekaden hat jedoch ein Wandel stattgefunden: Kausalität wurde mathematisiert und die Datenräume erweitert."
Roland Nagel, Geschäftsführer der Leipziger Foren



„Die RCT bildet nicht die Real-Life-Situation ab"
Prof. Dr. Frank Griesinger, Universitat Oldenburg



„Die notwendige Stufe der Evidenz (und damit die Typen, Anzahl, Größe und Qualität der benötigten Studien) ist abhängig vom Schadenspotential für die Patienten und dem (wahrscheinlichen) Nutzen der Therapie."
Prof. Dr. Wolfgang Hoffmann, Direktor des Instituts für Community Medicine der Universitätsmedizin Greifswald





„Wenn wir als Entscheidungsbasis im Gesundheitssystem zusätzlich zur Efficacy (Proof of Principle) mit Randomized Controlled Trials (RCTs) unter Experimental Study Conditions (ESC) die Real-World-Effectiveness (RWE) mit Pragmatic Controlled Trials (PCTs) unter Real-World-Conditions (RWC) messen, leiten wir einen Paradigmenwechsel ein."
Prof. Dr. Franz Porzsolt, Vorstand des Institute of Clinical Economics (ICE) e.V.

wendungsbegleitenden Datenerhebung abgebildet sein.

Einen weiteren Problemkreis sieht Windeler bei Therapien, zum Beispiel bei Orphans zu genetischen Erkrankungen, bei denen es keine andere Vergleichstherapien gibt. Dennoch brauche man, wenn es schon keine Vergleichstherapien gebe, zumindest Vergleichssituationen, denn: „Ohne Vergleich geht gar nichts." Windeler: „Es ist ein Irrtum, eine Utopie zu meinen, dass man ohne Vergleich eine Aussage zum Zusatznutzen machen kann." Das sage schon der Begriff Zusatznutzen, der impliziere, „gegenüber etwas anderem". Daher sei es ausschlaggebend, sich zu überlegen, wie man einen solchen Vergleich ausgestalten kann, damit dieser die Quantifizierung eines Zusatznutzen erlaube.

Ebenso gelte aber, dass je kleiner die zu erwartenden Unterschiede sind, desto grösser die Anforderungen an die Fairness des Vergleichs würden. Dazu brauche man immer eine Ausrichtungsgleichheit sowie das Intention-to-Treat-Prinzip, das man dann auch in Registern umsetzen müsste.

Da das alles nicht einfach wäre, sei das IQWiG gerade dabei, mit verschiedenen Registerbetreibern zu diskutieren, wie Register darauf vorzubereiten sind, um vergleichende Studien in Registern implementieren zu können. Dieses Unterfangen seines Instituts stoße bei den Registerbetreibern auf Gegenliebe, weil diese daran interessiert wären, dass ihre Bemühungen, ein gutes Register zu führen, für irgendetwas Vernünftiges genutzt würden.

Doch was bedeutet: gutes Register? Es komme laut Windeler vor allen Dingen darauf an, welchen Anforderungsgrad der G-BA an deren Qualität stellen will. Dies beginne bei der Frage, wie vollständig Daten in Registern zu erheben sind. Und reiche bis zur Frage, ob die in Registern zu findenden Daten überhaupt eine Quantifizierung samt dazu gehörenden Bezugsgrößen, Populationen, Nebenwirkungen und auch Lebensqualität erlauben. Auf diese Datenquelle setze dann die Real World-Evidence als Forschungsmethodik, im MVF-Kongressverlauf beschrieben von Friedhelm Leverkus (Pfizer, Berlin), Prof. Dr. Wolfgang Hoffmann (Institut für Community Medicine, Greifswald), Prof. Dr. Frank Griesinger (Universitat Oldenburg) und Prof. Dr. Franz Porzsolt (ICE e.V., Ulm) überhaupt erst auf.

von:
MVF-Chefredakteur Peter Stegmaier

Zitationshinweis

Stegmaier, P.: „Windeler: Ohne Vergleich geht gar nichts" (06/19), S. 32-35, doi: 10.24945/MVF.06.19.1866-0533.2187