



Herbst-Symposium 2015

***Real World Data –
ein Gewinn für die Nutzenbewertung?***
**Welchen Beitrag können Register und Routinedaten
liefern?**

Einführungen und Schlusswort von Jürgen



Windeler, IQWiG

27./28. November 2015

Die IQWiG-Position (Einführung am 27.11.2015)

Wir haben in diesem Jahr das Thema „Real World Data“ gewählt, weil darüber seit einigen Jahren zunehmend diskutiert wird. Es wird auch vorgebracht in Zusammenhang mit den Diskussionen um Versorgungsforschung. Mit „Real World Data“, auch „Real World Evidence“ oder „Real Life Data“, um nur einige Varianten zu nennen, verbinden viele die Vorstellung von Datenschätzen, denen sie mit leuchtenden Augen und großer Wertschätzung begegnen. Wir waren der Auffassung, dass wir diesen Begriff und seine Inhalte zum Thema des Herbstsymposiums machen sollten, denn darüber zu sprechen, darüber zu diskutieren, ist der Mühe wert. Zur Definition werden wir sicher noch im Laufe des Nachmittags kommen.

Eigene Position hinterfragen und diskutieren

Als Leiter des IQWiG bin ich in diesem Thema nicht neutral, denn allen ist klar: Das IQWiG hat hierzu eine Position. Sie ist bekannt und nachlesbar im Methodenpapier des Instituts. Sie ist ebenso nachvollziehbar in der täglichen Arbeit des IQWiG. Diese Position zu hinterfragen und zu diskutieren, gehört trotzdem zu den Aufgaben des Instituts. Und deswegen haben wir für heute und morgen eine ganze Reihe von kompetenten und geschätzten Fachleuten eingeladen und sie gebeten, uns sowohl ihre Arbeitsergebnisse vorzustellen als auch ihren Standpunkt darzulegen und zu diskutieren.

Kleiner, aber feiner Unterschied

Ich will aber zunächst noch einmal die Gelegenheit nutzen, um die Hintergründe für die Position des IQWiG in Erinnerung zu rufen.

Zunächst zur Klarstellung: Bei der Themenstellung des Herbstsymposiums geht es nicht darum, ob und welche Fragen man mit Real World Data beantworten kann. Denn da gibt es ganz viele Informationen, die sich daraus generieren lassen und eventuell auch für Entscheidungen nutzen lassen. Wir reden hier speziell und nur darüber, ob man mit solchen Daten Erkenntnisse zum medizinischen Nutzen von Interventionen gewinnen kann. Das ist ein feiner, aber wichtiger Unterschied.

Wenn wir uns darauf verständigen,

- dass man den Nutzen von Interventionen bis auf ganz besondere Ausnahmefälle – nachlesbar im IQWiG-Methodenpapier – nicht durch Vorher-Nachher-Betrachtungen von Behandlungsergebnissen unter der Überschrift „Der Patient hat profitiert“ bewerten kann, sondern
- dass man für Nutzenbewertungen vergleichende Studien / vergleichende Daten braucht,

dann kann man – ich glaube, diese Position ist relativ wenig diskussionsbedürftig – auf dieser Basis folgende Feststellungen treffen, die sich dann zu der IQWiG-Position kondensieren:

1. Randomisierte kontrollierte Studien (RCTs) bieten für Nutzenfragen das fehlerärmste Studiendesign, weil sie eine hohe interne Validität aufweisen. RCTs sind fast immer – natürlich sage ich nicht immer – machbar. Sie sind fast immer vertretbar. Und sie sind schon allein deshalb auch ethisch gut vertretbar, weil man mit diesen Studiendesigns und ihren Ergebnissen zu validen Erkenntnissen kommen kann, was wiederum die besondere Studiensituation für Patienten rechtfertigen kann.
2. Alle anderen Designs – und es ist relativ egal, ob „Real Life“ oder andere Etiketten draufkleben – sind deutlich schlechter, und sie sind in der Praxis bedauerlicherweise noch deutlich schlechter, als sie es in der Theorie sein müssten. Das hängt mit dem außerordentlich hohen Aufwand zusammen, der in der Confounder-Kontrolle notwendig ist, um sie intern valide zu machen. Dieser Aufwand wird leider oftmals unterschätzt, gescheut oder nicht ausreichend umgesetzt.
3. Probleme, die es bereits bei der zuverlässigen Identifizierung von RCTs gibt (Suchmöglichkeiten in Datenbanken, Publication Bias) potenzieren sich bei anderen Studiendesigns. Und das bedeutet, dass man der Unsicherheit und den Verzerrungsmöglichkeiten anderer Designs noch Unsicherheiten und Fehlerquellen auf der Meta-Ebene hinzufügt.
4. Die Vorstellung, dass nicht-randomisierte Studien, diese sogenannten Real Life Data, per se eine höhere externe Validität haben, ist falsch. Damit gewinnt die erwähnte besondere Unsicherheit und Fehlergeneigtheit von Real Life Data gegenüber Highly Controlled Trials noch an Bedeutung.
5. Die externe Validität, also die Frage der Passung zwischen Studienergebnissen und konkreten Entscheidungssituationen, ist wertlos, solange ich sie auf schlechten, unzuverlässigen Daten aufbauen muss. Denke ich über die Frage der externen Validität – der Anwendbarkeit – nach, heißt die Voraussetzung, dass ich Studien vor mir habe, die eine hohe interne Validität aufweisen.
6. Das Konzept sogenannter „Effectiveness“, aus dem sich die Definition für „Alltag“ und von „Real-Life“-Ideen und deren Wertschätzung ableitet, steht – vorsichtig ausgedrückt – auf wackligen Füßen. Ich persönlich halte es konzeptionell für schlichten Unsinn. Das bedeutet – und ich will das hier auch so klar zuspitzen – dass man den Nutzen im „Alltag“ mit den dafür eingesetzten Methoden nicht nur nicht ermitteln kann, sondern gar nicht ermitteln sollte.

In der Konsequenz bedeutet dies, dass wir aussagefähige Studien brauchen, die bei Fragen des Nutzens RCTs sein sollten. Diese sollten gesetzte Rahmenbedingungen und nicht einfach „Alltag“ abbilden, allerdings auch auf „künstliche“ Elemente (Run-in) verzichten.

Lippenbekenntnisse

Nun ist bisher in Diskussionen oft die Meinung vertreten worden, dass man die Bedeutung von validen, prospektiven Studien, insbesondere RCTs, anerkenne, und dass Real Life Data solche RCTs nicht ersetzen, sondern nur ergänzen sollten. Ich persönlich habe das meist für Lippenbekenntnisse gehalten. Bisher ist nämlich völlig im Dunkeln geblieben, wie und was da ergänzt werden soll. Unklar ist daher auch, ob diese Ergänzung zu Erkenntnissen beiträgt und zu welchen.

Dazu passt auch, dass Autoren – z. B. bei Auswertungen des schwedischen Stent-Registers – zunächst Nutzaussagen aus diesen Daten ableiten, um dann am Ende einer entsprechenden Publikation oder in der Diskussion darüber festzustellen, dass die gefundenen Ergebnisse nun in einem RCT bestätigt werden müssten – was natürlich nicht passiert.

Kernfragen bleiben

Als zentrale Fragen ergeben sich für unser Symposium also:

- Wie soll die Relation sogenannter Real-World-Daten zu anderer Evidenz aussehen?
- Worin genau soll die Ergänzung zu vergleichenden Studien liegen?
- Welche Fragen sollen „Real World Data“ beantworten und welche Anforderungen sind an sie zu stellen, um diese Fragen beantworten zu können?
- Und schließlich: Können solche Daten einen Gewinn für die Nutzenbewertung darstellen?

Es gibt viele Äußerungen dazu, was man mit Real Life Data anfangen könnte, aber vergleichsweise wenig Beispiele dafür, was man konkret damit angefangen hat. „Where's the beef?“ könnte man also fragen. Wo ist das Fleisch an den vielleicht vorhandenen Knochen? Ich freue mich auf sicher anregende Beiträge und eine interessante Diskussion.

Adaptive Pathways (Einführende Worte am 28.11.2015)

Das Konzept von „Real World Evidence“ oder „Real World Data“ spielt in einer speziellen Diskussion eine wichtige Rolle, die sich um sogenanntes „Adaptive Licensing“ oder „Adaptive Pathways“ dreht. Dieser Verfahrensansatz wird seit einigen Jahren diskutiert. Es hat zahlreiche Veranstaltungen und viele Publikationen gegeben, in denen für dieses Konzept geworben wurde. Es ist aber auch niemandem verborgen geblieben, dass sehr viele HTA-Institutionen, darunter auch das IQWiG, diesem Konzept mit großer Skepsis, sogar mit Vorbehalten und Sorge gegenüber stehen und noch sehr viel mehr offene Fragen als Antworten sehen.

Offene Punkte

Ich frage mich zum Beispiel, wo die Evidenz dafür ist, dass es einen „growing patient demand for timely access to promising therapies“ gibt (abgesehen davon, dass „timely“ und „promising“ sehr interpretationsfähige Begriffe sind). Hat denn mal jemand die Patienten gefragt und sie dabei auch auf die Konsequenzen hingewiesen, wie etwa den Umstand der erhöhten Unsicherheit?

Ich frage mich, inwiefern die bisherigen Spezialwege (Stichworte Orphan, Conditional, Compassionate) nicht ausreichend sind, um bestimmte Spezialfälle zu lösen, wie sie als Rechtfertigung für Adaptive Pathways vorgetragen werden.

Ich frage mich, wie das Versprechen oder die Hoffnung begründet ist, dass Firmen nach der Zulassung noch Evidenz schaffen. Man kann eigentlich wissen, dass das – meistens nicht funktioniert – und man kann sogar wissen, warum.

Ich frage mich, wie man auf die Hoffnung setzen kann, dass man mit Real-Life-Daten Nutzaussagen aus RCTs ersetzen kann.

Ich frage mich, wie die international breit zunehmende Wertschätzung von Patient reported Outcomes damit vereinbar ist, bei der Zulassung noch mehr als bisher auf Surrogate zu setzen. Und in Registern mit Routinedaten wird man PROs sowieso nicht finden.

Ich frage mich, wie die beruhigenden Aussagen aus der Industrie, das sei ja alles gar nicht neu und nur auf wenige Spezialfälle mit besonderem „Medical Need“ beschränkt, mit den Aussagen von Befürwortern vereinbar sind. Beispielsweise denen von Hans-Georg Eichler, 2012: „However, Adaptive Licensing is envisioned as the ultimate replacement for the current development and authorization process/model, and as such would be applicable to most new products.“ Und 2015: „... likely make adaptive pathways the preferred approach in the future“.

Ich frage mich, wenn – wie die Industrie behauptet – alles gar nicht neu und nur auf wenige Spezialfälle beschränkt bleibt, warum brauchen wir es dann überhaupt so dringend?

Leere Versprechungen

Das Versprechen, nach der frühzeitigen Zulassung Studien durchzuführen, um die Evidenzbasis zu verbessern, ist für viele nicht glaubwürdig – mal abgesehen von dem Versprechen an die Industrie für „earlier revenue stream and less expensive and shorter clinical trials“. Dies mag teilweise mit den speziellen deutschen Gegebenheiten zusammenhängen, wo es keine Bewertung von Wirkstoffen im Sinne einer 4. Hürde gibt. Damit sind Arzneimittel in Deutschland direkt nach der Zulassung verordnungsfähig. Und: Die Anwendung wird – auch ohne ergänzende Evidenz – nicht begrenzt auf bestimmte Patientengruppen bleiben, und der Druck, weitere Evidenz vorzulegen, wird sehr schwach sein. Die Möglichkeit, nicht gelieferte Evidenz zu sanktionieren, ist faktisch Null.

Evidenz wird schwächer

Sicher ist also: Die Evidenz wird schwächer. Die Unsicherheit wird zunehmen. Die Kompensationen beruhen auf Hoffnungen. Dies hat in den letzten Wochen vermehrt Kritiker auf den Plan gerufen, die eine deutliche Verschlechterung der Patientenversorgung befürchten.

Eine Arbeit im BMJ zur amerikanischen Situation stellt die Beispiele von drei neuen Arzneimitteln vor, die man als „promising“ für die Demenzbehandlung angesehen hatte, bevor sie in Phase-III-Studien ihren fehlenden Nutzen oder erhebliche Schäden demonstrierten. Die Autoren stellen fest, dass die Bemühungen, die Zulassung vorzuverlegen „under pressure from congress and industry“ erfolgten. Sie geben zu Bedenken, dass die Begrenzung auf vorläufige Daten zu „high costs in morbidity and healthcare dollars“ (in Europa dann natürlich Euro) führen werde.

Ein bekannter deutscher Medizinjournalist hat es so auf den Punkt gebracht: Die Ankündigung, Interventionen immer schneller auf den Markt zu bringen, kann durchaus als Drohung verstanden werden.

Diskussion erforderlich

Aber auch diese Punkte sind diskussionswürdig und wir haben dazu kompetente Referenten eingeladen, unter ihnen den Hauptprotagonisten, Hans Georg Eichler, Chief Medical Officer bei der europäischen Zulassungsbehörde.

Wohin führt der Weg? (Schlusswort am 28.11.2015)

Wir sind jetzt gegen Ende dieser Veranstaltung und ich will ein paar Punkte resümieren, zusammenfassen oder kommentieren:

Begriff „Real World“ unzutreffend

Aus dieser Veranstaltung nehme ich mit, den Begriff „Real World Data“ und „Real World Evidence“ zu verlassen. Ich werde auch versuchen, ihn nicht mehr zu benutzen, weil er faktisch für Daten aus nicht-randomisierten Studien (Non-RCTs) steht. Ich glaube auch, dass wir gut daran tun, ihn deshalb beiseite zu legen, weil er diese alberne Implikation befördert, dass randomisierte Studien sozusagen nichts mit „Real World“ und nicht-randomisierte Studien alles mit „Real Life“ zu tun hätten. Er vermittelt ein falsches, auch ein interessegeleitetes Bild, und deswegen macht es Sinn, den Begriff zu verlassen.

Verschiedene Methoden für verschiedene Fragestellungen

In verschiedenen Beiträgen ist der Methodenbaukasten hervorgehoben worden als Sinnbild dafür, dass für verschiedene Fragestellungen verschiedene Methoden zur Verfügung stehen. Ich vermute, jeder kann diesem Bild folgen, das IQWiG natürlich auch. Ich würde sogar noch weiter gehen: Wenn Elisabeth George vom englischen National Institute for Health and Care Excellence (NICE), sagt, dass es beim NICE

die Norm ist, Daten aus nicht-randomisierten Studien zu verwenden, dann sage ich, beim IQWiG auch! – und zwar für Fragen, für die man keine randomisierten Studien braucht. Und Herr Augustin hat gestern beschrieben, dass man natürlich auf „Real-World“-Daten zurückgreifen muss, die sich im Übrigen in allen Dossiers des IQWiG finden, etwa bei folgenden Fragen:

- Welche Durchdringung erreichen Arzneimittel?
- Wie viele Patienten sind davon betroffen?
- Was kostet der Wirkstoff, die Vergleichstherapie oder Best supportive Care?

Für solche Fragen sind nicht-randomisierte Studiendaten die Norm. Für andere Fragestellungen aber, die zu Nutzen und Schaden, gibt es andere Instrumente in diesem Methodenbaukasten. Sie bestehen in vergleichenden aussagefähigen Studien. Davon ist die randomisierte Studie aus einer Reihe von Gründen die beste.

Wenn man über ausnehmend große Effekte redet, insbesondere über den Ausschluss bestimmter Effekte und nicht über die Quantifizierung von Effekten, dann – auch das ist IQWiG-Position – können auch Daten ausreichen, die möglicherweise ein relativ hohes Verzerrungspotenzial im Sinne einer hohen Unsicherheit haben.

Können Non-RCTs Daten aus RCTs ergänzen?

Ich habe öfter in den Vorträgen gehört, dass nicht-randomisierte Studien RCTs ergänzen sollen. Ich muss gestehen, auch nach meinem sorgfältigen Zuhören seit gestern Mittag, dass mir immer noch nicht so ganz klar ist, worin nun die Ergänzung eigentlich bestehen soll.

Viele Beispiele, die genannt worden sind, beziehen sich auf Schadensaspekte, auf Schadenssignale, auf die Quantifizierung unter der Fragestellung: „Sind große Effekte auszuschließen?“ Und da, glaube ich, gibt es auch kein wirkliches Problem.

Es gab Beispiele für Ergänzungen, die in Richtung Langzeitdaten gehen, die Situationen nannten, wo keine RCTs machbar sind – auch hier müssten in meinen Augen relativ hohe Anforderungen gelten. Aber trotz aller Beispiele und Vorschläge für Ergänzungsvarianten ist mir nicht wirklich klar geworden, worin nun eigentlich der besondere Beitrag der nicht-randomisierten Studien für die Nutzenbewertung bestehen soll.

Adaptive Pathways – wohin führt der Weg?

Elisabeth George richtete auch an das IQWiG den Appell, sich an Adaptive-Pathways-Aktivitäten zu beteiligen. Wir würden uns auch sehr viel leichter für eine Beteiligung entscheiden können, wenn wir wüssten, warum man diesen Weg eigentlich beschreitet und wohin er führen wird. Und dieses „Wohin?“ ist nicht trivial.

Herr Paar hat vorhin gesagt, wir redeten im Moment über sechs Beratungsfälle im Rahmen der Adaptive Pathways. Das ist natürlich noch kein Tsunami, wir sind ja ganz, ganz am Anfang dieses ganzen Prozesses. Aber wie man in den Publikationen von Herrn Eichler nachlesen kann, soll der Adaptive-Pathways-Prozess der präferierte Prozess für die Zukunft werden. Da habe ich keine Ausnahmen gelesen. Ich habe auch nicht gelesen, dass er nur für bestimmte Situationen gelten soll, für die wir übrigens ja Regelungen schon haben. Nach dem, was man lesen kann, scheint es ein Prozess zu sein, der einen weiten Anwendungsbereich haben will. Und wie weit der eigentlich gehen soll, weiß ich nicht.

Sind Studien nach der Zulassung realistisch?

Es fiel mir leichter, mich an Arbeiten im Zusammenhang mit Adaptive Pathways zu beteiligen, wenn es darum ginge, Folgendes zu klären:

1. Wie ist die Industrie dazu zu bringen, Studien nach der Zulassung zu machen?
2. Welche Sanktionen und Konsequenzen greifen, wenn keine Studien folgen?
3. Wie lässt sich Managed Prescription oder Managed Use im deutschen Gesundheitswesen umsetzen, denn absehbar sind gesetzliche Änderungen notwendig, um dies in Deutschland zu implementieren.

An solchen konzeptionellen Überlegungen kann sich das IQWiG gerne beteiligen. Nur an einer Umsetzung, die im Moment schon auf dem Weg ist, tun wir uns schwer aus den Gründen, die ich genannt habe.

Unterschreitung aller Evidenzanforderungen

Aktuell können wir bei den Regelungen zu Medizinprodukten und Methoden beobachten, wie sich die Entwicklung von einem Nutzenanspruch zu einem Potenzialanspruch hinbewegt. Und wir erleben im Moment alle, was Potenzial – ein anderes Wort für „Promise“ – im deutschen System bedeutet, nämlich eine Unterschreitung aller üblichen Evidenzanforderungen. Diesen Weg möchten wir im Arzneimittelsektor nicht so ohne weiteres mitgehen. Wir würden ihn an anderer Stelle auch gern zurückdrehen, aber da arbeiten wir dran.

Fragt die Bürger!

Meiner Ansicht nach ist das Argument „es gibt sowieso keine ergebnissicheren Studien und aus diesem Grund sind beschleunigte Zulassungsverfahren machbar“ nicht fair. Warum? Adaptive Licensing / Adaptive Pathways bedeutet in jedem Fall, dass wir die Unsicherheit erhöhen. Und dies mit einem Versprechen auf zukünftige Sicherheit, auf zukünftige, dann zu liefernde Evidenz, an deren Realisierung ich erhebliche Zweifel habe. Und ich meine, dass dieser Prozess, der zu einer Erhöhung der Unsicherheit führt, eigentlich kein Prozess ist, der von den Institutionen – weder vom IQWiG, noch vom G-BA, nicht von der EMA oder der Europäischen Kommission – ausgehen kann, sondern eigentlich ein Prozess ist, zu dem die Bürger befragt werden müssen. Dabei sollten sie auch informiert werden, dass sich die Arzneimittelsicherheit verringert. „Wollt ihr das?“ diese Frage hat sie bisher noch nicht erreicht.

Methodische Standards einhalten und einfordern!

Die Big-Data-Diskussion möchte ich nicht vermischt wissen mit der über „Real World Data“. Sie sind getrennt voneinander zu betrachten. Einen Aspekt haben sie gemeinsam: Wir sollten dafür sorgen, dass bei beiden Formen der Datengenerierung und -auswertung bestimmte Standards eingehalten werden. Wir sollten uns intensiv darum kümmern, dass methodische Standards, die dafür sorgen, dass die Fragen wirklich valide beantwortet werden können, auch eingehalten und eingefordert werden.