

Mündliche Anhörung



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

hier: Wirkstoff Tenofoviralfenamid

Sitzung im Hause des Gemeinsamen Bundesausschusses in Berlin
am 12. Februar 2019
von 12:33 Uhr bis 12:52 Uhr

– Stenografisches Wortprotokoll –

Angemeldete Teilnehmer der Firma **Gilead Sciences GmbH**:

Frau Dransfeld
Frau Hogger
Frau Dr. Hüging
Herr Kandlbinder

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

Herr Dr. Rasch
Herr Dr. Werner

Beginn der Anhörung: 12:32 Uhr

(Die angemeldeten Teilnehmer betreten den Raum.)

Herr Prof. Hecken (Vorsitzender): Meine sehr verehrten Damen und Herren, herzlich willkommen im Unterausschuss Arzneimittel des Gemeinsamen Bundesausschusses zu TAF Erwachsene und Jugendliche ab zwölf Jahren und Körpergewicht von mindestens 35 Kilogramm zur Behandlung chronischer Hepatitis B, Nutzenneubewertung nach Fristablauf, Dossierbewertung des IQWiG vom 20.12.2018, die Basis des Stellungnameverfahrens, der schriftlichen Stellungnahmen und der heutigen Erörterung. Stellung genommen haben zum einen Gilead Sciences GmbH als pharmazeutischer Unternehmer, die dagnä und der Verband forschender Arzneimittelhersteller.

Da wir wieder ein Wortprotokoll führen, muss ich zunächst für das Protokoll die Anwesenheit feststellen. Ich weise zugleich daraufhin, dass ich darum bitte, dass Sie das Mikrofon benutzen, jeweils Namen, entsendendes Unternehmen oder Verband nennen, bevor Sie das Wort ergreifen. Für Gilead müsste Frau Dransfeld anwesend sein – die habe ich schon gesehen –, Frau Hogger, Frau Dr. Hüging und Herr Kandlbinder, ebenfalls für Gilead. Ja. Herr Rasch und Herr Werner sind immer noch da. Seien Sie uns herzlich willkommen in dieser kleinen und überschaubaren Runde.

Wie gesagt, wir haben es mit einer erneuten Bewertung nach Fristablauf zu tun. Ich gebe Ihnen die Möglichkeit, aus Ihrer Sicht kurz einzuführen. Mich würde zur Methodik der Metaanalyse vom pU interessieren: Wie begründet sich die Einschätzung der Studienheterogenität? Weshalb wird von Ihnen ein Modell mit zufälligen Effekten als adäquat angesehen? Das ist eine relativ relevante Frage, über die wir sprechen sollten.

Wer macht die Einführung? – Frau Dransfeld, bitte schön.

Frau Dransfeld (Gilead): Schönen guten Tag. Sehr geehrter Herr Prof. Hecken! Sehr geehrte Damen und Herren! Zunächst möchte ich uns noch einmal kurz vorstellen: Frau Hogger ist Biostatistikerin. Sie hat uns bei der Erstellung des Dossiers unterstützt. Frau Dr. Hüging ist bei uns in der medizinisch-wissenschaftlichen Abteilung für die Lebererkrankungen zuständig, und Herr Kandlbinder leitet im Hause Gilead die Abteilung Market Access. Mein Name ist Sabine Dransfeld, und ich bin in der Abteilung Market Access für die Bereiche HIV und Lebererkrankungen verantwortlich.

Nach der befristeten Bewertung sprechen wir heute zum zweiten Mal über die Substanz Tenofoviralaftamid in der Indikation Hepatitis B. Ich kann es relativ kurz machen.

In der ersten Bewertung gab es vor allen Dingen Kritik hinsichtlich der vorgelegten Verträglichkeitsendpunkte und der Definition der Teilpopulation nach Behandlungsstatus. Diese Kritikpunkte haben wir mit dem aktuellen und umfangreichen Dossier adressiert. Die vom IQWiG in der Bewertung monierte Analyse der randomisierten Population der 144-Wochendaten haben wir mit der Stellungnahme nachgereicht, obwohl wir hier nach wie vor der Auffassung sind, dass die Daten nicht wirklich belastbar sind.

Der guten Ordnung halber haben wir mit der Stellungnahme mittlerweile vorliegende Daten aus einer neuen Studie bei transplantierten Patienten mit chronischer Nierenerkrankung eingereicht. Die ist zwar aufgrund der nicht umgesetzten zweckmäßigen Vergleichstherapie unserer Auffassung nach nicht nutzenbewertungsrelevant, würde aber das Ergebnis so oder so nicht verändern. Wir haben es der Transparenz halber jetzt mit eingereicht.

Das Ergebnis der Bewertung lässt sich schnell zusammenfassen und ist im Hinblick auf die Produktprofile nicht überraschend, denn außer bei Niere und Knochen sind sich naturgemäß die beiden Substanzen TAF und TDF sehr ähnlich, was ihre Sicherheitsprofile betrifft. Entsprechend zeigt sich bei

den klinischen Endpunkten kein Unterschied. Die renalen und ossären Endpunkte, bei denen sich die Unterschiede zeigen, sind allesamt Surrogatparameter. Wir haben es vielfach diskutiert. Sie sind in der Bewertung nicht berücksichtigt worden. Da diese Surrogatparameter jedoch in der klinischen Praxis und auch in der Patientenversorgung relevant sind, haben wir hier einen Zusatznutzen abgeleitet, den wir für gerechtfertigt halten. TAF ist ein deutlich verbessertes TDF, jedoch lassen sich die Vorteile nur über die Surrogate abbilden.

Damit möchte ich von meiner Seite die kurze Einleitung schließen und würde gerne, wenn Sie damit einverstanden sind, an Frau Hogger übergeben, die als Biostatistikerin kurz die Frage adressiert, die Sie in der Einleitung aufgeworfen haben.

Herr Prof. Hecken (Vorsitzender): Herzlichen Dank, Frau Dransfeld. – Frau Hogger, bitte.

Frau Hogger (Gilead): Bei einem metaanalytischen Modell mit den festen Effekten wird angenommen, dass allen Studien ein gleicher Effekt zugrunde liegt und die Unterschiede nur durch zufällige Abweichungen zustande kommen. Im Gegensatz dazu bezieht ein Modell mit zufälligen Effekten die Heterogenität mit ein. Die Heterogenität kann viele Ursachen haben, beispielsweise Unterschiede im Design, im Aufbau der Studie, unterschiedliche Studienpopulationen etc.

Das IQWiG erachtet in seiner Nutzenbewertung die Verwendung von festen Effekten als adäquat, da die Studien hinsichtlich der Methodik nahezu identisch sind. Außerdem gehen in die Metaanalyse vier Studien ein. Das spricht eher für ein Modell mit festen Effekten.

Auf der anderen Seite unterscheiden sich aber die Patienten zwischen den Studien systematisch voneinander. Es befinden sich Patienten in unterschiedlichem Krankheitsstadium. Zum Beispiel sieht man das an der mittleren Visuslast. Sie unterscheidet sich dadurch deutlich. Zwei Studien wurden ausschließlich in China durchgeführt. Aufgrund dieser unterschiedlichen Patientenpopulationen kann die Heterogenität nicht ausreichend sicher ausgeschlossen werden. Daher wurde im Nutzendossier ein Modell mit zufälligen Effekten gewählt. Auch nach der Modulvorlage sind in der Regel Modelle mit zufälligen Effekten durchzuführen. Hierbei handelt es sich um einen konservativeren Ansatz.

Beim Vergleich der Metaanalysen mit den festen Effekten, die das IQWiG im Rahmen der Nutzenbewertung selbst durchgeführt hat, mit den Metaanalysen im Modell mit den zufälligen Effekten sieht man, dass sich die Effektschätzer und die Konfidenzintervalle kaum unterscheiden. Schlussendlich führen beide Modelle in allen klinischen Endpunkten zu denselben Aussagen.

Herr Prof. Hecken (Vorsitzender): Danke schön. Gibt es Fragen? – Herr Kaiser, bitte.

Herr Dr. Kaiser: Ich habe zwei, drei Anmerkungen. Vielleicht zunächst einmal zu Ihren Ausführungen zu den nachgereichten Daten: Sie haben zwei Komplexe nachgereicht. Das eine ist die von Ihnen erwähnte 25-Personen-Studie. Dazu haben wir in unserer Bewertung schon beschrieben – sie war aufgrund der Dinge, die man zu dieser Studie wusste, schon identifizierbar –, dass sie für diese Fragestellungen nicht relevant ist. Das haben Sie gerade selber beschrieben. Insofern haben Sie das jetzt dargestellt. Aber für die eigentliche Fragestellung ist das keine zusätzliche Evidenz.

Der zweite Punkt ist, dass Sie Daten zu 144 Wochen und den unerwünschten Ereignissen nachgereicht haben. Ich sage einmal, es ist Ihre Einschätzung, dass Sie nicht mehr bringen, aber man muss erst einmal auf die Daten schauen, und dann kann man sich entscheiden, was das bringt. Sie haben durch die Art der Auswertung leider folgendes Problem verursacht: Sie haben die Auswertung auf diejenigen Patienten und Patientinnen beschränkt, die tatsächlich 144 Wochen beobachtet wurden. Damit haben Sie die ursprüngliche Randomisierung nicht mehr. Wir haben häufig einen vorzeitigen Studienabbruch, selbst in der 96-Wochen-Studie, da machen Sie natürlich trotzdem ein gesamtes Analy-

setzt und bezieht das auf alle Patienten. Jetzt haben Sie an der Stelle keine ITT-Analyse mehr. Das ist ein Problem mit einer solchen Auswertung. Es wäre eigentlich sinnvoll gewesen, das insgesamt zu haben.

Wenn ich auf die Diskussion mit festen und zufälligen Effekten eingehen darf: Wenn ich Sie richtig verstanden habe, haben Sie am Ende gesagt, im Grunde genommen kommt sowieso nichts Unterschiedliches dabei heraus. Es ist so, dass Sie an der Stelle aus dem Methodenpapier bei uns zitieren. In der Version 2017 haben wir beschrieben, was seit einigen Jahren – Das betrifft nicht nur Deutschland, sondern auch andere, unter anderem die Cochrane Collaboration, die eine zentrale Kollaboration für systematische Übersichten ist. Man hat gesehen, dass das Modell für zufällige Effekte, das man über Jahre verwendet hat, völlig falsche Ergebnisse liefert, nämlich dieses Modell nach der DerSimonian and Laird.

Aus diesem Grunde sieht man in einer Situation mit wenigen Studien – bei vielen Studien hat man nicht das Problem, das ist unter fünf, vier ist kleiner als fünf, bei weniger als fünf Studien –, dass man im Grunde genommen eine Abwägung zwischen festen und zufälligen Effekten hat, auf der einen Seite ein Methodenproblem, vielleicht auf der anderen ein inhaltliches Argumentationsproblem, nämlich ob die Studien ähnlich genug sind oder nicht.

Was aber ganz sicher gar nicht mehr geht, ist, zufällige Effekte mit diesen DSL, DerSimonian-und-Laird-Modell zu rechnen. Das haben Sie aber gemacht, obwohl in dem Methodenpapier, das Sie zitiert haben, beschrieben ist, dass dieses Modell völlig ungeeignet ist. Wenn Sie Knapp-Hartung – das ist ein anderes Modell für zufällige Effekte – gewählt und damit agiert hätten, hätte man sehen können, welche Auswirkungen das hat oder nicht. Ich glaube, es ist letztendlich insofern alles vergebene Liebesmüh, als dass – wie Sie richtig beschrieben haben – nichts Unterschiedliches herauskommt. In Ihrem Dossier ist zunächst etwas Unterschiedliches in Ihrer Beurteilung herausgekommen, weil Sie, wie Sie in den Stellungnahmen beschrieben haben, eine falsche Anwendung des Peto Odds Ratio gemacht haben. Das haben Sie mit den Stellungnahmen korrigiert, auch aufgrund der Angaben, die wir gemacht haben. Insofern ist die Diskussion um die Methodik, glaube ich, zwar für zukünftige Verfahren wichtig, aber für dieses konkrete Verfahren nicht entscheidend.

Für mich wäre vielmehr die Frage, die ich noch einmal wiederholen möchte: Warum haben Sie die 144 Wochen-Daten nicht in einer ITT-Population vorgewählt?

Herr Prof. Hecken (Vorsitzender): Wer möchte antworten? – Frau Dransfeld, einer muss antworten. Bitte.

Frau Dransfeld (Gilead): Ja, ich muss. – Wir haben die 144-Wochen-Daten, die Gesamtpopulation. Sie haben völlig Recht. Das ITT ist in der jetzigen Analyse nicht mehr gewährleistet. Wir haben die Nachfrage und Kritik des IQWiG so verstanden, dass es hier tatsächlich um die randomisierte Teilpopulation geht. Wir hatten die kompletten Daten zu den 144 Wochen mit dem Dossier eingereicht, aber nicht weiter ausgewertet, weil wir auch da transparent sein wollten, haben aber da die Analyse nicht vorgenommen. Wir waren der Auffassung, es war eine Analyse für die noch immer randomisierte Population gefordert.

Herr Prof. Hecken (Vorsitzender): Herr Kaiser, bitte.

Herr Dr. Kaiser: Ich weiß nicht, ob wir uns jetzt missverstehen. Es gibt die Situation von Gesamtstudien und Teilpopulationsstudien. Das ist für mich nicht die Frage. Es muss natürlich eine treffgenaue Teilpopulation sein, aber innerhalb der Teilpopulation natürlich insofern eine ITT-Analyse, als dass alle diejenigen, die in diese Teilpopulation zu Studienbeginn eingeschlossen worden sind, am Ende auch nach 144 Wochen ausgewertet werden. Diejenigen, die nicht bis Woche 144 beobachtet wer-

den, weil sie früher aus der Studie ausscheiden, werden – das ist das Übliche – mit quasi LOCF (last observation carried forward) ausgewertet, indem man den letzten Beobachtungszeitpunkt zum Beispiel auf 96 Wochen setzt. Das ist etwas ganz Normales. Das haben wir vermisst. Sie haben das auf die Population beschränkt, die tatsächlich 144 Wochen beobachtet wurde. Das ist ein Problem, weil das kein ITT-Ansatz mehr ist. Das hat mit der Gesamt- und Teilpopulation nichts zu tun.

Herr Prof. Hecken (Vorsitzender): Okay. Das nehmen Sie zur Kenntnis. – Frau Bickel und dann Frau Müller.

Frau Bickel: Ich bin etwas über Ihr Eingangsstatement gestolpert, dass Sie letztendlich den Zusatznutzen nur noch bei Knochendichtefrakturen und den Nierenerkrankungen sehen. Ich habe das in dem Dossier und in Ihrer Stellungnahme anders verstanden. Da versuchen Sie noch, einen Zusatznutzen aufgrund der ALT-Normalisierung abzuleiten. Dann sagen Sie in Ihrer Stellungnahme, dass Sie eigentlich nur zwei korrigieren, wo Sie keinen Zusatznutzen mehr sehen, weil Sie da das Peto Odds Ratio anders berechnet haben. Da muss das relative Risiko kleiner als Eins sein. Aber in Ihrem Dossier sehe ich noch deutlich mehr Endpunkte, die Sie zur Ableitung des Zusatznutzens heranziehen. Mich wundert das. Oder nehme ich das nur als Widerspruch wahr?

Herr Prof. Hecken (Vorsitzender): Frau Dransfeld, bitte.

Frau Dransfeld (Gilead): Das hört sich vermutlich wirklich widersprüchlich an, ist aber so zu verstehen: Zum einen haben wir Fehler korrigiert. Das hat Herr Kaiser gerade angemerkt. Zum anderen geht es um die primären Vorteile TAF gegen TDF. Das sind Vorteile. Ungeachtet der Tatsache, ob sich noch statistisch signifikante Unterschiede zeigen, aus denen wir etwas abgeleitet haben, teilweise nur in Subpopulationen, geht es bei dem wirklich primär patientenrelevanten Profil von TAF um den Bereich Niere und Knochen. Das ist der Hintergrund, warum schlussendlich Tenofoviralfenamid auf Basis von Tenofovirdisoproxil entwickelt wurde. Deshalb haben wir das hier ganz klar, um es zu fokussieren, primär auf die renalen, ossären Vorteile beschränkt. Ansonsten sind es die gleichen Wirksubstanzen – das muss man klar sagen – in unterschiedlicher Produktformulierung, die aber ihre Effekte primär in Niere und Knochen ausweisen.

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

Frau Dr. Müller: Eine Frage ist schon beantwortet. Sie sind dem IQWiG gefolgt, unter welchen Bedingungen man Peto Odds Ratio anwenden kann und haben das für alle, bei denen Sie das angewendet haben, mit der Stellungnahme korrigiert. Das ist geklärt.

Ich habe noch eine Frage zu den nachgereichten UE-Daten. Es ging darum, dass Sie nur die Patienten ausgewertet haben, für die Sie bis zum Ende des Beobachtungszeitraumes Daten haben. Herr Kaiser hat eine Ersetzungsstrategie angemerkt, was ein adäquates Vorgehen gewesen wäre. Könnten Sie zu einer entsprechenden Ersetzungsstrategie etwas nachliefern? Es ist nur eine Frage, beispielsweise „last observation carried forward“ oder eine andere Ersetzungsstrategie, je nachdem was geeignet wäre, oder nicht, um das möglicherweise verwertbar zu machen? Es ist schon ein gravierendes Problem, wenn die Patienten herausfallen, bei denen man keine Daten mehr hat. Das gibt ein Bias. Das ist einfach eine Frage.

Herr Prof. Hecken (Vorsitzender): Frau Dransfeld.

Frau Dransfeld (Gilead): Eine Rückfrage: Bezieht sich Ihre Nachfrage auf die Wirksamkeitsendpunkte oder auf die UE?

Frau Dr. Müller: Ich habe es so verstanden, dass das von Herrn Kaiser zu den nachgereichten Daten auf die unerwünschten Ereignisse ging. Aber das ist eigentlich egal. Egal, was man macht, muss man – –

Frau Dransfeld (Gilead): Ehrlicherweise kann man immer noch Daten nachreichen und Analysen machen. Ich würde in dem Fall aber die Sinnfrage stellen wollen. Es zeigen sich weder – das hat das IQWiG auch in seiner Bewertung festgestellt – in der gesamten Population irgendwelche Unterschiede, noch in der zugegebenermaßen nicht-mehr-ITT-Analyse, die wir nachgereicht haben, noch ist irgendetwas zu erwarten, vor allen Dingen vor dem Hintergrund, dass im Hinblick auf die Substanzen als solche nichts zu erwarten ist. Außer bei Nieren und Knochen, wo wir auch bei 96 Wochen schon genug sehen, was aber leider nur über Surrogate abgebildet werden kann, würde ich mich an dieser Stelle trauen, die Sinnfrage zu stellen.

Frau Dr. Müller: Sie würden die Daten nicht nachliefern? Aber Sie sehen ein, dass das, was Sie an Daten geliefert haben, für die Fragestellung nicht verwertbar ist?

Frau Dransfeld (Gilead): Das haben wir von vornherein gesagt. Wir glauben, dass die Analyse bzw. die Bewertung auf Basis der 96-Wochen-Daten erfolgen kann. Erfahrungsgemäß muss man ganz klar sagen, dass sich zwischen Woche 96 und 144 eigentlich nichts mehr tut. Insofern sind die Analyse bzw. die Bewertung auf Basis der vorgelegten vollumfänglichen 96-Wochen-Daten sicherlich möglich.

Herr Prof. Hecken (Vorsitzender): Herr Kaiser.

Herr Dr. Kaiser: Ich möchte das insofern unterstützen, als dass wir in unserer Dossierbewertung nicht gefordert haben, dass diese Daten zwingend erforderlich sind, um eine Bewertung machen zu können, sondern wir haben nach Sichtung der Daten eine Zusammenfassung gemacht. Nach Sichtung der UE-Daten zu Woche 144, die in gewissem Ausmaß aber nicht vollständig vorhanden waren, zeigen sich keine relevanten Änderungen durch den längeren Beobachtungszeitraum. Aus unserer Sicht kann man mit den vorhandenen und bewerteten 96-Wochen-Daten eine ausreichend sichere Aussage treffen.

Herr Prof. Hecken (Vorsitzender): Okay. Dann hat sich das erledigt. Gibt es weitere Fragen? – Keine. Dann brauchen wir es nicht ohne Not zu verlängern. Möchten Sie noch einmal zusammenfassen? – Sonst bedanken wir uns wechselseitig.

Frau Dransfeld (Gilead): Genau. Vielen Dank.

Herr Prof. Hecken (Vorsitzender): Dann bedanken wir uns wechselseitig für die schöne halbe Stunde und für die Beantwortung der Fragen und die Darstellung dessen, was Sie uns präsentiert haben. Dann können wir diese Anhörung beenden. Herzlichen Dank, dass Sie hier waren.

Schluss der Anhörung: 12:52Uhr