

IQWiG-Berichte – Nr. 278

Aflibercept
(Addendum zum Auftrag A14-32)

Addendum

Auftrag: A15-05
Version: 1.0
Stand: 13.02.2015

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Aflibercept
(Addendum zum Auftrag A14-32)

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

28.01.2015

Interne Auftragsnummer:

A15-05

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8 (KölnTurm)
50670 Köln

Tel.: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

ISSN: 1864-2500

An dem Addendum beteiligte Mitarbeiter des IQWiG¹:

- Beate Wieseler
- Catharina Brockhaus
- Petra Kohlepp

Schlagwörter: Aflibercept, Makulaödem, Nutzenbewertung

Keywords: Aflibercept, Macular Edema, Benefit Assessment

¹ Aufgrund gesetzlicher Datenschutzbestimmungen haben Mitarbeiter das Recht, ihrer Namensnennung nicht zuzustimmen.

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	iv
Abkürzungsverzeichnis.....	v
1 Hintergrund.....	1
2 Bewertung.....	2
2.1 Generelle Überlegungen zur Auswahl von Endpunkten zur Sehschärfe.....	2
2.2 Bewertung der Responderanalyse zur Verbesserung bzw. Verschlechterung der Sehschärfe um ≥ 15 EDTRS-Buchstaben.....	4
2.3 Bewertung der stetigen Daten zur Beschreibung der Sehschärfe	5
2.4 Zusammenfassung.....	7
3 Literatur	8

Tabellenverzeichnis

	Seite
Tabelle 1: Responderanalyse Verbesserung bzw. Verschlechterung der Sehschärfe ≥ 15 ETDRS-Buchstaben – RCT, indirekter Vergleich: Aflibercept vs. Ranibizumab.....	5

Abkürzungsverzeichnis

Abkürzung	Bedeutung
BCVA	best corrected visual acuity (bestkorrigierte Sehschärfe)
DMÖ	diabetisches Makulaödem
ETDRS	Early Treatment Diabetic Retinopathy Study
G-BA	Gemeinsamer Bundesausschuss
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
KI	Konfidenzintervall
MID	Minimal Important Difference
pU	pharmazeutischer Unternehmer
RCT	Randomisierte kontrollierte Studie
RR	relatives Risiko
SMD	standardisierten Mittelwertdifferenzen

1 Hintergrund

Der Gemeinsame Bundesausschuss (G-BA) hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) am 28.01.2015 mit einer ergänzenden Bewertung zum Auftrag A14-32 (Aflibercept (neues Anwendungsgebiet) – Nutzenbewertung gemäß § 35a SGB V [1]) beauftragt.

Mit seiner Stellungnahme hat der pharmazeutische Unternehmer (pU) weitere Angaben zu Ergebnissen des indirekten Vergleichs von Aflibercept und Ranibizumab vorgelegt [2]. Der G-BA hat daher das IQWiG mit der Bewertung der Angaben zu Endpunkten zur Beschreibung der Sehschärfe für den genannten indirekten Vergleich beauftragt.

Die Verantwortung für die vorliegende Bewertung und das Bewertungsergebnis liegt ausschließlich beim IQWiG. Die Bewertung wird dem G-BA übermittelt. Über den Zusatznutzen beschließt der G-BA.

2 Bewertung

Mit seiner Stellungnahme hat der pharmazeutische Unternehmer (pU) weitere Angaben zum indirekten Vergleich von Aflibercept und Ranibizumab bei diabetischem Makulaödem für Endpunkte zur Sehschärfe vorgelegt [2].

Abschnitt 2.1 enthält generelle Überlegungen zur Auswahl von Endpunkten zur Sehschärfe in der vorliegenden Bewertung. Die Bewertung weiterer Endpunkte zur Sehschärfe wird in den Abschnitten 2.2 und 2.3 dargestellt. In Abschnitt 2.4 wird zusammenfassend beschrieben, ob und wenn ja welche Aussagen der ursprünglichen Dossierbewertung A14-32 sich durch diese Bewertung ändern.

2.1 Generelle Überlegungen zur Auswahl von Endpunkten zur Sehschärfe

In den Studien zu Aflibercept und Ranibizumab, die in den indirekten Vergleich zur Beurteilung des Zusatznutzens von Aflibercept bei DMÖ eingehen, wurde die Sehschärfe mit einer Sehtafel gemäß dem Standard der ETDRS gemessen. Die Anzahl der korrekt gelesenen Einzelbuchstaben wurde als bestkorrigierte Sehschärfe (best corrected visual acuity, BCVA) berichtet, nachfolgend als „ETDRS-Buchstaben“ bezeichnet. Der pU legt im Dossier und in seiner Stellungnahme zur Dossierbewertung eine Vielzahl von Auswertungen verschiedener Operationalisierungen der Sehschärfe vor. Er präsentiert Ergebnisse zu folgenden Operationalisierungen (jeweils für die 3 Effektmaße Odds Ratio, relatives Risiko und absolute Risikoreduktion):

- Mittlere Veränderung der BCVA nach 52 Wochen (Differenz zwischen Ausgangs- und Endwert der Sehschärfe)
- Mittlere Veränderung der BCVA von Woche 4 bis Woche 52 (Durchschnitt der mittleren Veränderung über die Zeit; für jeden Patienten wird für jeden Monat die Veränderung zum Ausgangswert berücksichtigt, anschließend wird der Durchschnitt dieser Werte gebildet)
- Anteil an Patienten mit einer Verbesserung der Sehschärfe um ≥ 10 ETDRS-Buchstaben nach 52 Wochen
- Anteil an Patienten mit einer Verschlechterung der Sehschärfe um ≥ 10 ETDRS-Buchstaben nach 52 Wochen
- Anteil an Patienten mit einer Verbesserung der Sehschärfe um ≥ 15 ETDRS-Buchstaben nach 52 Wochen
- Anteil an Patienten mit einer Verschlechterung der Sehschärfe um ≥ 15 ETDRS-Buchstaben nach 52 Wochen

Im Dossier zieht der pU zur Beschreibung des Zusatznutzens auf Basis der Sehschärfe damit eine Vielzahl von Effektschätzern heran (insgesamt 18). Aus diesem Vorgehen entsteht bereits innerhalb der Bewertung des Endpunkts Sehschärfe das Problem der Multiplizität

(weitere Multiplizität entsteht durch die Betrachtung weiterer Endpunkte z. B. zur gesundheitsbezogenen Lebensqualität). Ein solches Vorgehen ist nicht sinnvoll und führt ggf. zu einer erhöhten Rate falsch-positiver Ergebnisse.

Über das Problem der Multiplizität hinaus stellt sich für Endpunkte zur Sehschärfe die Frage der Relevanz der Effekte. Nicht jede Änderung der Sehschärfe, die bei Erhebung mit dem ETDRS-Verfahren gemessen wird, ist für den Patienten spürbar und damit relevant. Bei der Bewertung von Aflibercept im Vergleich zu Ranibizumab geht es damit um die Frage, ob der beobachtete Unterschied an EDTRS-Buchstaben zwischen den beiden Behandlungen überhaupt relevante Unterschiede anzeigt. Diese Relevanzbewertung kann auf Basis von Mittelwertdifferenzen und Responderanalysen vorgenommen werden [3].

Um skalen- und indikationsspezifischen Besonderheiten gerecht zu werden, wendet das Institut für die Relevanzbewertung regelhaft folgende Hierarchie an, deren Stufen durch das Vorliegen verschiedener Relevanzkriterien determiniert sind [4].

- 1) Falls für die jeweilige Skala eine begründete Irrelevanzschwelle für den Gruppenunterschied (Mittelwertdifferenz) vorliegt bzw. ableitbar ist, wird diese für die Relevanzbewertung herangezogen. Liegt das zum beobachteten Effekt korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle, ist statistisch abgesichert, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Das Institut wertet dies als ausreichend für den Nachweis eines relevanten Effekts, da sich in diesem Fall die beobachteten Effekte in der Regel deutlich oberhalb der Irrelevanzschwelle (und zumindest nahe an einer Relevanzschwelle) realisieren.
- 2) Falls skalenspezifische begründete Irrelevanzschwellen nicht vorliegen oder ableitbar sind, kommen Responderanalysen in Betracht. Hier ist es erforderlich, dass bei diesen Analysen ein validiertes bzw. etabliertes Responsekriterium angewendet wurde (z. B. im Sinne einer individuellen Minimal Important Difference [MID]) [5]. Liegt bei einer solchen Auswertung ein statistisch signifikanter Unterschied bezüglich der Anteile an Respondern zwischen den Gruppen vor, wird dies, sofern keine spezifischen Gründe dagegen sprechen, als Nachweis eines relevanten Effekts angesehen, da die Responserdefinition bereits eine Schwelle der Relevanz beinhaltet.
- 3) Falls weder skalenspezifische Irrelevanzschwellen noch Responderanalysen vorliegen, wird für die Relevanzbewertung auf ein allgemeines statistisches Maß in Form von standardisierten Mittelwertdifferenzen (SMD in Form von Hedges' g) zurückgegriffen. Als Irrelevanzschwelle wird dann 0,2 verwendet: Liegt das zum Effektschätzer korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Dies soll gewährleisten, dass der Effekt hinreichend sicher mindestens als „klein“ angesehen werden kann [6].

Der pU legt im Dossier und in den Stellungnahmen keine Angaben vor, aus denen eine begründete Irrelevanzschwelle für den Gruppenunterschied abgeleitet werden kann. Gemäß der oben beschriebenen Hierarchie hat das IQWiG aus den vorgelegten Analysen deshalb eine Responderanalyse für die Bewertung des Zusatznutzens herangezogen [1]. Die Wahl fiel dabei auf die Responderanalyse mit dem Kriterium ≥ 10 Buchstaben, weil dieser Wert in der Literatur als MID beschrieben wird [7].

2.2 Bewertung der Responderanalyse zur Verbesserung bzw. Verschlechterung der Sehschärfe um ≥ 15 EDTRS-Buchstaben

Der pU legt in seinem Dossier Responderanalysen mit 2 verschiedenen Responsekriterien (≥ 10 EDTRS-Buchstaben und ≥ 15 EDTRS-Buchstaben) vor. Für die Dossierbewertung wurde die Responderanalyse mit dem Kriterium ≥ 10 Buchstaben herangezogen, weil dieser Wert in der Literatur als MID beschrieben wird [7]. In der Anhörung zur Dossierbewertung wurde diskutiert, ob zusätzlich die Responderanalyse mit dem Kriterium ≥ 15 Buchstaben betrachtet werden sollte, da bei Patienten mit starker Seheinschränkung eine Verbesserung (bzw. Verschlechterung) um ≥ 10 EDTRS-Buchstaben ggf. nicht zu einer spürbaren Veränderung führe.

Im Fall der vorliegenden Bewertung erscheint die Responderanalyse mit dem Kriterium ≥ 10 Buchstaben angemessen, weil die mittlere Sehstärke zu Studienbeginn in den Studien der Nutzenbewertung (59 bis 65 EDTRS-Buchstaben) geringfügig besser war als in der Studie, die zur Ermittlung der MID herangezogen wurde (55 EDTRS-Buchstaben). Es bleibt unklar, ob die Betrachtung der Responderanalyse mit ≥ 15 Buchstaben aus den Studien der Nutzenbewertung tatsächlich Rückschlüsse auf Patienten mit stärkerer Seheinschränkung erlaubt.

Generell erhöht die Betrachtung mehrerer Responderanalysen mit verschiedenen Responsekriterien die Multiplizität in der Nutzenbewertung und sollte deshalb vermieden werden.

Unabhängig von diesen Überlegungen zeigt auch die Analyse mit einem Responsekriterium von ≥ 15 EDTRS-Buchstaben keinen Unterschied zwischen Aflibercept und Ranibizumab (siehe Tabelle 1).

Tabelle 1: Responderanalyse Verbesserung bzw. Verschlechterung der Sehschärfe ≥ 15 ETDRS-Buchstaben – RCT, indirekter Vergleich: Aflibercept vs. Ranibizumab

Endpunktkategorie Endpunkt Vergleich Studie	Aflibercept bzw. Ranibizumab		Laserfotokoagulation		Gruppenunterschied RR [95 %-KI]; p-Wert
	N	Patienten mit Ereignis n (%)	N	Patienten mit Ereignis n (%)	
Morbidität					
Verbesserung der Sehschärfe ≥ 15 ETDRS-Buchstaben					
Aflibercept vs. Laser					
VISTA	151	47 (31,1)	154	12 (7,8)	3,99 [2,21; 7,23]; k. A.
VIVID	135	45 (33,3)	132	12 (9,1)	3,67 [2,03; 6,61]; k. A.
Gesamt					3,83 [2,52; 5,81]; < 0,001 ^a
Ranibizumab vs. Laser					
RESTORE	115	26 (22,6)	110	9 (8,2)	2,76 [1,36; 5,63]; < 0,001
REVEAL	133	25 (18,8)	128	10 (7,8)	2,41 [1,20; 4,81]; k. A.
Gesamt					2,57 [1,57; 4,23]; < 0,001 ^a
Indirekter Vergleich^b:					
Aflibercept vs. Ranibizumab (mit REVEAL)					1,49 [0,78; 2,84]; k. A.
Aflibercept vs. Ranibizumab (ohne REVEAL)					1,38 [0,61; 3,16]; k. A.
Verschlechterung der Sehschärfe ≥ 15 ETDRS-Buchstaben					
Aflibercept vs. Laser					
VISTA	151	1 (0,7)	154	14 (9,1)	0,07 [0,01; 0,55]; k. A.
VIVID	135	0 (0)	132	14 (10,6)	0,03 [0,00; 0,56]; k. A.
Gesamt					0,06 [0,01; 0,29]; < 0,001 ^a
Ranibizumab vs. Laser					
RESTORE	115	1 (0,9)	110	9 (8,2)	0,11 [0,01; 0,83]; k. A.
REVEAL	133	2 (1,5)	128	5 (3,9)	0,38 [0,08; 1,95]; k. A.
Gesamt					0,23 [0,07; 0,84]; 0,03 ^a
Indirekter Vergleich^b:					
Aflibercept vs. Ranibizumab (mit REVEAL)					0,24 [0,03; 1,90]; k. A.
Aflibercept vs. Ranibizumab (ohne REVEAL)					0,53 [0,04; 7,27]; k. A.
a: berechnet aus Meta-Analyse					
b: adjustierter indirekter Vergleich nach Bucher [8]					
ETDRS: Early Treatment Diabetic Retinopathy Study; k. A.: keine Angabe; KI: Konfidenzintervall; N: Anzahl ausgewerteter Patienten; n: Anzahl Patienten mit mindestens einem Ereignis; RCT: randomisierte kontrollierte Studie; RR: relatives Risiko; vs.: versus					

2.3 Bewertung der stetigen Daten zur Beschreibung der Sehschärfe

Zusätzlich zu den Responderanalysen legt der pU im Dossier 2 Auswertungen der stetigen Daten zur Sehschärfe vor. In diesen Auswertungen wird die mittlere Veränderung der BCVA

nach 52 Wochen bzw. die mittlere Veränderung der BCVA von Woche 4 bis Woche 52 herangezogen (siehe Abschnitt 2.1).

Der pU legt keine Irrelevanzschwelle für einen Gruppenunterschied für die mit einer Sehtafel gemäß dem Standard der ETDRS gemessenen Sehschärfe vor. Wie im Methodenpapier des IQWiG beschrieben (siehe auch Abschnitt 2.1), wurde deshalb aus der Vielzahl von vorgelegten Analysen der Sehschärfe die Responderanalyse mit der empirisch erhobenen MID von 10 Buchstaben [7] für die Bewertung des Zusatznutzens herangezogen.

Unabhängig von der Hierarchie für die Auswahl der Endpunkte würde die Auswertung der stetigen Daten eine Relevanzbewertung mithilfe des Hedges' g benötigen. Die untere Grenze des Konfidenzintervalls des Hedges' g liegt für beide Analysen unterhalb von 0,2 (Gruppenunterschied der Veränderung der BCVA nach 52 Wochen: Mittlerer Unterschied [95 %-KI] Aflibercept vs. Ranibizumab: 4,81 [2,52; 7,11], Hedges' g: 0,37 [0,12; 0,62]; Gruppenunterschied der mittleren Veränderung der BCVA von Woche 4 bis Woche 52: Mittlerer Unterschied [95 %-KI] Aflibercept vs. Ranibizumab: 2,95 [1,16; 4,73], Hedges' g: 0,19 [-0,06; 0,44]). Ein irrelevanter Effekt kann deshalb für beide Analysen nicht ausgeschlossen werden.

Die Werte zum Hedges' g wurden vom IQWiG neu berechnet. Die Ergebnisse des pU aus dem Dossier zum Hedges' g waren zunächst nicht nachvollziehbar. Aus der Stellungnahme des pU wurde deutlich, dass der pU für die Berechnung des Hedges' g zunächst die studienspezifischen nicht-standardisierten Mittelwertsdifferenzen berechnet und mit diesen jeweils für die Vergleiche Aflibercept vs. Laserfotokoagulation sowie Ranibizumab vs. Laserfotokoagulation Meta-Analysen durchführt. In einem weiteren Schritt versucht der pU anhand der meta-analytischen Schätzer der Mittelwertdifferenzen einen indirekten Vergleich nach Bucher, in dem gleichzeitig durch Standardisierung versucht wird, das standardisierte Effektmaß Hedges' g zu schätzen. Hierbei benutzt der pU jedoch nur die Standardabweichungen der einzelnen Therapiearme von Aflibercept und Ranibizumab. Bei den Standardabweichungen der einzelnen Therapien, welche in die gepoolte Standardabweichung eingehen, handelt es sich um die gemeinsame Variabilität in den jeweiligen Therapie-Armen (Aflibercept in VISTA und VIVID bzw. Ranibizumab-Arm in RESTORE und REVEAL). Bei diesem Vorgehen wird weder die Stichprobengröße noch die Variabilität in den jeweiligen Kontroll-Armen berücksichtigt.

Die vom pU geschätzte gepoolte Standardabweichung stellt wegen der Nichtberücksichtigung der Variabilität in den Kontrollarmen eine Unterschätzung der tatsächlichen Variabilität dar. Darüber hinaus setzt sich die Varianz des Effektschätzers des adjustierten indirekten Vergleiches nach Bucher aus der Summe der Varianzen der meta-analytischen Therapievergleiche (Aflibercept vs. Laserfotokoagulation bzw. Ranibizumab vs. Laserfotokoagulation) zusammen. Da beim Vorgehen des pU die gepoolte Standardabweichung, statt wie von Bucher gefordert die Summe der Varianzen, in die

Schätzung des Effektes im indirekten Vergleich eingeht, und diese zudem die Variabilität in den Kontrollarmen nicht berücksichtigt, ergibt sich:

- 1) eine Verzerrung des Effektschätzers (die Effektschätzung wird fälschlich größer)
- 2) die Unsicherheit, gemessen anhand der Breite der resultierenden Konfidenzintervalle, wird unterschätzt.

Die Berechnung des Hedges' g durch den pU ist deshalb aus Sicht des IQWiG zur Bewertung der Relevanz des Effekts nicht geeignet.

Insgesamt belegen die Analysen der stetigen Daten keine relevanten Vorteile von Aflibercept im Vergleich zu Ranibizumab.

2.4 Zusammenfassung

Die im vorliegenden Addendum diskutierten Daten und Auswertungen ändern die Aussage der Dossierbewertung A14-32 nicht. Ein Zusatznutzen von Aflibercept im Vergleich zu Ranibizumab ist nicht belegt.

3 Literatur

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Aflibercept: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A14-32 [online]. 11.12.2014 [Zugriff: 10.02.2015]. (IQWiG-Berichte; Band 263). URL: https://www.iqwig.de/download/A14-32_Aflibercept-neues-Anwendungsgebiet_Nutzenbewertung-35a-SGB-V.pdf.
2. Bayer Vital. Stellungnahme zum IQWiG-Bericht Nr. 263: Aflibercept; Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A14-32. [Demnächst verfügbar unter: <https://www.g-ba.de/informationen/nutzenbewertung/139/#tab/beschluesse> im Dokument "Zusammenfassende Dokumentation"].
3. Schünemann HJ, Akl EA, Guyatt GH. Interpreting the results of patient reported outcome measures in clinical trials: the clinician's perspective. *Health Qual Life Outcomes* 2006; 4: 62.
4. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 4.1. Köln: IQWiG; 2013. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4-1.pdf.
5. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; 61(2): 102-109.
6. Fayers P, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. Chichester: Wiley; 2007.
7. Koch KR, Muether PS, Hermann MM, Hoerster R, Kirchhof B, Fauser S. Subjective perception versus objective outcome after intravitreal ranibizumab for exudative AMD. *Graefes Arch Clin Exp Ophthalmol* 2012; 250(2): 201-209.
8. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997; 50(6): 683-691.