IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

IQWiG Reports – Commission No. A15-05

Aflibercept (Addendum to Commission A14-32)[1]

Addendum

Commission: A15-05
Version: 1.0
Status: 13 February 2015

[1] Translation of addendum A15-05 *Aflibercept* (*Addendum zum Auftrag A14-32*) (Version 1.0; Status: 13 February 2015). Please note: This translation is provided as a service by IQWiG to English-language readers. However, solely the German original text is absolutely authoritative and legally binding.

Publishing details

Publisher:
Institute for Quality and Efficiency in Health Care

Topic:
Aflibercept (Addendum to Commission A14-32)

Commissioning agency:
Federal Joint Committee

Commission awarded on:
28 January 2015

Internal Commission No.:
A15-05

Address of publisher:
Institute for Quality and Efficiency in Health Care
Im Mediapark 8 (KölnTurm)
50670 Cologne
Germany

Tel.: +49 (0)221 – 35685-0
Fax: +49 (0)221 – 35685-1
E-Mail: berichte@iqwig.de
Internet: www.iqwig.de

IQWiG employees involved in the addendum[2]:

- Beate Wieseler
- Catharina Brockhaus
- Petra Kohlepp

Keywords: aflibercept, macular edema, benefit assessment

[2] Due to legal data protection regulations, employees have the right not to be named.

Table of contents

List of tables

Page

List of abbreviations

Abbreviation	Meaning
BCVA	best corrected visual acuity
CI	confidence interval
DMO	diabetic macular oedema
ETDRS	Early Treatment Diabetic Retinopathy Study
G-BA	Gemeinsamer Bundesausschuss (Federal Joint Committee)
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care)
MID	minimally important difference
SGB	Sozialgesetzbuch (Social Code Book)
SMD	standardized mean difference

1 Background

On 28 January 2015, the Federal Joint Committee (G-BA) commissioned the Institute for Quality and Efficiency in Health Care (IQWiG) to conduct a supplementary assessment for Commission A14-32 (Aflibercept (new therapeutic indication) – Benefit assessment according to §35a Social Code Book (SGB) V [1]).

With its comment, the pharmaceutical company (hereinafter referred to as "the company") presented further information on the results of the indirect comparison of aflibercept and ranibizumab [2]. The G-BA therefore commissioned IQWiG to assess the information on outcomes describing visual acuity for the indirect comparison mentioned.

The responsibility for the present assessment and the results of the assessment lies exclusively with IQWiG. The assessment is forwarded to the G-BA. The G-BA decides on the added benefit.

2 Assessment

With its comment, the company presented further information on the indirect comparison of aflibercept and ranibizumab in diabetic macular oedema (DMO) for the outcomes on visual acuity [2].

Section 2.1 contains general considerations on the choice of outcomes on visual acuity in the present assessment. The assessment of further outcomes on visual acuity is presented in Sections 2.2 and 2.3. Section 2.4 summarizes whether, and, if any, which conclusions of the original dossier assessment A14-32 were changed by this assessment.

2.1 General considerations on the choice of outcomes on visual acuity

Visual acuity was measured with a vision chart according to the Early Treatment Diabetic Retinopathy Study (ETDRS) standard in the studies on aflibercept and ranibizumab that were included in the indirect comparison for the assessment of the added benefit of aflibercept in DMO. The number of individual letters that can be read correctly was reported as best corrected visual acuity (BCVA), hereinafter referred to as "ETDRS letters". In the dossier and in its comment on the dossier assessment, the company presented a large number of analyses of different operationalizations of visual acuity. It presented results on the following operationalizations (in each case for the 3 effect measures odds ratio, relative risk and absolute risk reduction):

- mean change in BCVA after 52 weeks (difference between baseline and final value of visual acuity)
- mean change in BCVA from week 4 to week 52 (average of the mean change over time; the change from the baseline value is considered for each patient for each month, then the average of these values is calculated)
- proportion of patients with improvement of visual acuity by ≥ 10 ETDRS letters after 52 weeks
- proportion of patients with worsening of visual acuity by ≥ 10 ETDRS letters after 52 weeks
- proportion of patients with improvement of visual acuity by ≥ 15 ETDRS letters after 52 weeks
- proportion of patients with worsening of visual acuity by ≥ 15 ETDRS letters after 52 weeks

In the dossier, the company used a large number of effect estimates to describe the added benefit on the basis of visual acuity (18 in total). The problem of multiplicity resulted from this approach already within the assessment of the outcome "visual acuity" (further multiplicity resulted from the consideration of further outcomes, e.g. on health-related quality

of life). Such an approach is not meaningful and may result in an increased rate of false positive results.

Beyond the problem of multiplicity, there is the question of the relevance of the effects for outcomes on visual acuity. Not every change in visual acuity that is measured in the recording with the ETDRS method is tangible for the patient and therefore relevant. The question in the assessment of aflibercept in comparison with ranibizumab is therefore whether the observed difference in ETDRS letters between the 2 treatments actually indicates relevant differences. This evaluation of relevance can be made on the basis of differences in mean values as well as responder analyses [3].

In order to do justice to characteristics specific to scales and therapeutic indications, the Institute as a rule uses the following hierarchy for the evaluation of relevance, the corresponding steps being determined by the presence of different relevance criteria [4].

1) If a justified irrelevance threshold for the group difference (mean difference) is available or deducible for the corresponding scale, this threshold is used for the evaluation of relevance. If the corresponding confidence interval (CI) for the observed effect lies completely above this irrelevance threshold, it is statistically ensured that the effect size does not lie within a range that is certainly irrelevant. The Institute judges this to be sufficient for demonstration of a relevant effect, as in this case the effects observed are normally realized clearly above the irrelevance threshold (and at least close to the relevance threshold).
2) If scale-specific justified irrelevance criteria are not available or deducible, responder analyses may be considered. It is required here that a validated or established response criterion was used in these analyses (e.g. in terms of an individual minimally important difference [MID]) [5]. If a statistically significant difference is shown in such an analysis in the proportions of responders between groups, this is seen as demonstrating a relevant effect (unless specific reasons contradict this), as the responder definition already includes a threshold of relevance.
3) If neither scale-specific irrelevance thresholds nor responder analyses are available, a general statistical measure for evaluating relevance is drawn upon in the form of standardized mean differences (SMD expressed as Hedges' g). An irrelevance threshold of 0.2 is then used: If the CI corresponding to the effect estimate lies completely above this irrelevance threshold, it is assumed that the effect size does not lie within a range that is certainly irrelevant. This is to ensure that the effect can be regarded at least as "small" with sufficient certainty [6].

The company presented no information in the dossier and in the comments, from which a justified irrelevance threshold for the group difference can be derived. According to the hierarchy described above, IQWiG therefore used a responder analysis from the analyses presented for the assessment of the added benefit [1]. The responder analysis with the

criterion of ≥ 10 letters was chosen because this score is described as MID in the literature [7].

2.2 Assessment of the responder analysis for improvement or worsening of visual acuity by ≥ 15 EDTRS letters

The company presented responder analyses with 2 different response criteria (≥ 10 EDTRS letters and ≥ 15 EDTRS letters) in its dossier. The responder analysis with the criterion of ≥ 10 letters was used for the dossier assessment because this score is described as MID in the literature [7]. In the hearing on the dossier assessment it was discussed whether the responder analysis with the criterion of ≥ 15 should be additionally used because in patients with severe visual impairment improvement (or worsening) by ≥ 10 ETDRS letters might not result in a tangible change.

In case of the present assessment, the responder analysis with the criterion of ≥ 10 letters appeared appropriate because the mean visual acuity at baseline was marginally better in the studies of the benefit assessment (59 to 65 ETDRS letters) than in the study that was used to determine the MID (55 ETDRS letters). It remained unclear whether the consideration of the responder analysis with ≥ 15 letters from the studies of the benefit assessment actually allows to draw conclusions on patients with more severe visual impairment.

Generally, the consideration of several responder analyses with different response criteria increases the multiplicity in the benefit assessment and should therefore be avoided.

Regardless of these considerations, the analysis with a response criterion of ≥ 15 ETDRS letters also showed no difference between aflibercept and ranibizumab (see Table 1).

Table 1: Responder analysis improvement or worsening of visual acuity ≥ 15 ETDRS letters – RCT, indirect comparison: aflibercept vs. ranibizumab

Outcome category outcome comparison study	Aflibercept or ranibizumab N	Aflibercept or ranibizumab Patients with event n (%)	Laser photocoagulation N	Laser photocoagulation Patients with event n (%)	Group difference RR [95% CI]; p-value
Morbidity					
Improvement of visual acuity by ≥ 15 ETDRS letters					
Aflibercept vs. laser					
VISTA	151	47 (31.1)	154	12 (7.8)	3.99 [2.21; 7.23]; ND
VIVID	135	45 (33.3)	132	12 (9.1)	3.67 [2.03; 6.61]; ND
Total					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]; ND
Total					2.57 [1.57; 4.23]; < 0.001[a]
Indirect comparison[b]:					
Aflibercept vs. ranibizumab (with REVEAL)					1.49 [0.78; 2.84]; ND
Aflibercept vs. ranibizumab (without REVEAL)					1.38 [0.61; 3.16]; ND
Worsening of visual acuity by ≥ 15 ETDRS letters					
Aflibercept vs. laser					
VISTA	151	1 (0.7)	154	14 (9.1)	0.07 [0.01; 0.55]; ND
VIVID	135	0 (0)	132	14 (10.6)	0.03 [0.00; 0.56]; ND
Total					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]; ND
REVEAL	133	2 (1.5)	128	5 (3.9)	0.38 [0.08; 1.95]; ND
Total					0.23 [0.07; 0.84]; 0.03[a]
Indirect comparison[b]:					
Aflibercept vs. ranibizumab (with REVEAL)					0.24 [0.03; 1.90]; ND
Aflibercept vs. ranibizumab (without REVEAL)					0.53 [0.04; 7.27]; ND

a: Calculated from meta-analysis.
b: Adjusted indirect comparison according to Bucher [8].
CI: confidence interval; ETDRS: Early Treatment Diabetic Retinopathy Study; N: number of analysed patients; n: number of patients with at least one event; ND: no data; RCT: randomized controlled trial; RR: relative risk; vs.: versus

2.3 Assessment of the continuous data for the description of visual acuity

In addition to the responder analyses, the company presented 2 analyses of the continuous data on visual acuity in the dossier. The mean change in BCVA after 52 weeks and the mean change in BCVA from week 4 to week 52 were used in these analyses (see Section 2.1).

The company presented no irrelevance threshold for a group difference for the visual acuity measured with a vision chart according to the ETDRS standard. As described in the IQWiG method paper (see also Section 2.1), the responder analysis with the empirically determined MID of 10 letters [7] was therefore chosen from the numerous analyses of visual acuity presented and was used for the assessment of the added benefit.

Irrespective of the hierarchy for the choice of outcomes, the analysis of the continuous data would require an evaluation of relevance using Hedges' g. The lower limit of the CI of Hedges' g was below 0.2 for both analyses (group difference of the change in BCVA after 52 weeks: mean difference [95% CI] aflibercept vs. ranibizumab: 4.81 [2.52; 7.11], Hedges' g: 0.37 [0.12; 0.62]; group difference of the mean change in BCVA from week 4 to week 52: mean change [95% CI] aflibercept vs. ranibizumab: 2.95 [1.16; 4.73], Hedges' g: 0.19 [−0.06; 0.44]). Hence for both analyses, an irrelevant effect cannot be excluded.

The values on Hedges' g were recalculated by the Institute. The company's results from the dossier on Hedges' g were initially not comprehensible. It was clear from the company's comment that the company initially calculated the study-specific non-standardized mean differences for the calculation of Hedges' g, and used these to calculate meta-analyses for the comparisons of aflibercept versus laser photocoagulation and of ranibizumab versus laser photocoagulation. In a next step, the company tried to conduct an indirect comparison according to Bucher using the meta-analytical estimates of the mean differences, in which at the same time it was tried to estimate the standardized effect measure Hedges' g by means of standardization. The company only used the standard deviations of the individual treatment arms of aflibercept and ranibizumab to do this. The standard deviations of the individual treatments that were included in the pooled standard deviation were the joint variability in the respective treatment arms (aflibercept in VISTA and VIVID, or ranibizumab arm in RESTORE and REVEAL). Neither the sample size nor the variability of the respective control arms are considered in this approach.

The pooled standard deviation estimated by the company represents an underestimation of the actual variability because of the non-consideration of the variability in the control arms. Moreover, the variance of the effect estimate of the adjusted indirect comparison according to Bucher is the sum of the variances of the meta-analytical treatment comparisons (aflibercept versus laser photocoagulation, and ranibizumab versus laser photocoagulation). Since, in the company's approach, the pooled standard deviation, and not the sum of the variances as required by Bucher, was included in the estimation of the effect in the indirect comparison, and this also did not consider the variability in the control arms, this resulted in:

1) a bias of the effect estimate (the estimation of the effect is incorrectly increased)
2) an underestimation of the uncertainty, measured with the width of the resulting CIs

Hence, from the Institute's point of view, the calculation of the Hedges' g by the company was unsuitable for the assessment of the relevance of the effect.

Overall, the analyses of the continuous data proved no relevant advantages of aflibercept in comparison with ranibizumab.

2.4 Summary

The data and analyses discussed in the present addendum do not change the conclusion of dossier assessment A14-32. An added benefit of aflibercept compared with ranibizumab is not proven.

3 References

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Aflibercept: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A14-32 [online]. 11 December 2014 [accessed: 10 February 2015]. (IQWiG-Berichte; Volume 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. [Soon available under: https://www.g-ba.de/informationen/nutzenbewertung/139/#tab/beschluesse in the document "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. Institute for Quality and Efficiency in Health Care. General Methods: version 4.1 [online]. 28 November 2013 [accessed: 1 August 2014]. URL: https://www.iqwig.de/download/IQWiG_General_Methods_Version_%204-1.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.