CrossMark

ORIGINAL PAPER

Four years of early benefit assessment of new drugs in Germany: a qualitative study on methodological requirements for quality of life data

Christine Blome[1] · Matthias Augustin[1] · Hidayet Metin[2] · David Lohrberg[1]

Abstract

Background Since 2011, an early benefit assessment (EBA) of new drugs constricts free price setting in Germany. According to the Pharmaceutical Market Restructuring Act (AMNOG), pharmaceutical companies are obliged to demonstrate added benefit of new drugs over comparative treatment. Benefit is usually evaluated by the Institute for Quality and Efficiency in Health Care (IQWiG). The final appraisal is made by the Federal Joint Committee, Germany's highest-ranking decision body in the health sector, triggering drug prize negotiations between companies and statutory health insurance funds. One of four evaluation criteria is quality of life (QoL). QoL outcomes have, however, only rarely been pivotal in EBAs.
Objective This study determined methodological requirements for QoL measurement and data presentation in the EBA.
Design In a qualitative content analysis, documents of all EBAs completed by 2014 were searched for the term QoL. Relevant passages of all EBAs of 2011–2013 were independently extracted and reduced to key content by two researchers. Recurring patterns were identified and verified through comparison with EBAs of 2014.
Results We identified a range of requirements regarding QoL assessment, analysis, presentation, and interpretation, which go beyond official regulations. Disease-specific questionnaires are preferred and have to be validated according to certain standards and in the respective patient group. Effects must exceed the minimal important difference, which in turn must be validated in compliance with specific requirements. Often, instruments were not accepted as QoL measures, sometimes inconsistently across EBAs. Another frequent reason for non-acceptance of QoL data was that more than 30 % of randomized patients could not be analyzed due to missing data.
Conclusions Non-compliance with methodological requirements for QoL evidence impairs chances for positive benefit evaluation.

Keywords AMNOG · Quality of life · Qualitative research · Health technology assessment · Early benefit assessment · Comparative effectiveness research

JEL Classification I18

Introduction

Since January 2011, an early benefit assessment ("frühe Nutzenbewergung"; EBA) is required for new pharmaceuticals in Germany, introduced by the Pharmaceutical Market Restructuring Act ("Gesetz zur Neuordnung des Arzneimittelmarktes"; AMNOG). Manufacturers now have to demonstrate that a drug provides higher patient-relevant benefit than the appropriate comparator therapy. The price at which the new drug will be reimbursed by statutory health insurance funds now depends on this evidence of added benefit—as has already been the case in many European countries including, for example, France and Sweden [1]. Prior to the AMNOG, manufacturers had been free to set prices in the German market [2], and almost all

✉ Christine Blome
c.blome@uke.de

1   German Center for Health Services Research in Dermatology (CVderm), Institute for Health Services Research in Dermatology and Nursing (IVDP), University Medical Center Hamburg-Eppendorf, Martinistraße 52, 20246, Hamburg, Germany

2   Janssen-Cilag GmbH, Neuss, Germany

2 Springer

licensed pharmaceuticals were covered by statutory health insurance [3]. The EBA has been introduced in order to reduce the statutory health funds' high expenses for pharmaceuticals [4]. EBA outcomes are likely to have a high impact on pharmaceutical prices in Europe, as Germany is the country most frequently referenced by other countries using external reference pricing [5].

The evaluation of added benefit is based on four patient-relevant outcomes: mortality, morbidity, side effects, and quality of life (QoL), as defined in the Social Code Book V ("Sozialgesetzbuch V"; SGB V). Among these, QoL is the outcome most recently established in research. Using patient-reported outcomes (such as QoL) in comparative effectiveness research can be challenging, for example when it comes to selecting the appropriate instrument or interpreting results [6]. In this study, we therefore looked at methodological requirements regarding QoL data in the EBA.

The procedure of an EBA is as follows. At launch of a new drug in Germany (or within 1 month after indication change), the pharmaceutical company submits a dossier with data on the product's added benefit to the Federal Joint Committee ("Gemeinsamer Bundesausschuss", G-BA, the supreme decision-making body of physicians, dentists, psychotherapists, hospitals, and health care funds in Germany). The G-BA by convention commissions the Institute for Quality and Efficiency in Health Care (IQWiG) to evaluate the added benefit; only if the dossier relates to a drug for a rare disease with revenues not exceeding 50 million euros in the past 12 months, the G-BA itself performs the benefit assessment. In the latter case, added benefit is already proved through market authorization. However, after exceeding the 50 million sales mark, a new evaluation is performed by the IQWiG. In this case, an added benefit is not legally assured. Stakeholders can comment on the IQWiG's evaluation in a hearing process.

The final decision on added benefit is made by the G-BA and may differ between patient subgroups. The agreement between G-BA's and IQWiG's benefit ratings is substantial, with the IQWiG's ratings tending to be lower (i.e., less added benefit) [3]. If added benefit has been proven, the G-BA's decision on added benefit forms the basis for reimbursement price negotiations between the pharmaceutical company and the Central Federal Association of Health Insurance Funds ("Spitzenverband Bund der Krankenkassen"). If no added benefit has been proven, the drug will be subject to reference pricing provided a reference pricing group exists (i.e., a group of active ingredients with a defined maximum price according to §35 SGB V). If no such group exists, the drug's yearly cost to the statutory health insurance may not be higher than those of the most economic comparator treatment [7].

The G-BA publishes a synopsis, summarizing its decision and the main underlying reasons ("Tragende Gründe") in addition to a detailed decision and documentation document. In only five cases between 2011 and 2014 an added benefit due to QoL data was explicitly stated, as opposed to 20 cases showing a benefit due to mortality data, 36 due to morbidity data, and 25 due to data on side effects (computation based on the G-BA's sub-group-level decisions; the G-BA created 183 subgroups in 105 EBAs). This raises the question of why QoL data were the basis for an added benefit in so markedly fewer cases than the other outcomes categories.

Requirements for proving an added benefit are defined in the legislations on EBA, in the G-BA's rules of procedure and, particularly, in the IQWiG's "General Methods" paper [8]. However, these documents can only provide general guidance on benefit evaluation; they cannot foresee and specify each individual question that may arise from future EBAs. It might also be possible that IQWiG and G-BA deviate from these specifications. It can therefore be informative to look at the individual benefit evaluations conducted so far in addition to specifications in legislations and guidelines. For example, a definition of QoL is neither to be found in the legislations nor in the IQWiG's "General Methods". In single EBAs, however, the IQWiG explicitly defines QoL as "a complex construct comprising psychological, physical, and social domains" (aclidinium bromide; vandetanib par. 5b; axitinib), citing Schipper et al. [9].

This study analyzed the methodological requirements regarding QoL data in the German EBA, conducting a qualitative content analysis [10] of documents from all EBAs completed in the first 4 years after introduction of the AMNOG.

Methods

We conducted a systematic analysis of documents from all EBAs completed between January 2011 and December 2013 using the approach of qualitative content analysis according to Philipp Mayring [10]. This method is suitable for analyzing large amounts of text exploratively, i.e., without pre-specified hypotheses to be tested [11]. In the second step, we verified the results of this analysis through comparison with all EBAs completed in 2014.

The qualitative analysis was based on documents publicly available on the G-BA website [12] including:

– benefit dossier of the pharmaceutical company (module 1: summary of dossier content)
– dossier evaluation and benefit assessment by IQWiG or G-BA (the latter only for pharmaceuticals for rare diseases)
– protocol of the oral hearing (verbatim report)
– rationale of the G-BA decision ("Tragende Gründe" = main justifications)

The documents were downloaded and imported into the software MAX-QDA (VERBI, Berlin, Germany). Using the search terms "quality of life", the German equivalent "Lebensqualität", and their respective abbreviations, relevant passages were identified and extracted "en bloc" into Microsoft Excel 2013 spreadsheets.

Each text extract was reduced to one or more sentences, representing its core content regarding our research question. As reduction and abstraction of text material is necessarily influenced by subjective interpretation, it was independently performed by two researchers (CB and DL), the two versions were each discussed, and a consensus was found. This process was intended to enhance objectivity, avoid text misinterpretations, and minimize the risk of missing important aspects. Text extracts representing content already covered in previous extracts within the same EBA were excluded from further analysis (e.g., information given in both text and table form within the same document). For each paraphrase, the respective author (pharmaceutical company, IQWiG, G-BA member, G-BA chairperson, others) and EBA procedure (i.e., the drug to be evaluated) were documented.

Both researchers assigned one or more codes to each extract. A code represented the paraphrase's content or topic in a keyword or a catch phrase. Codes were developed inductively based on text content and were successively refined during the analytic process. All codes were listed in a document detailing scope and definition of each code and giving anchoring examples. Codings by the two researchers were discussed and harmonized.

All paraphrases, together with the respective text extracts, were grouped by code. For each code, a summary with overall conclusions was written. This was done partly by DL, partly by CB, with a subsequent critical review by the other respective researchers followed by discussion and consensus finding.

For verification of these results, documents of all EBAs completed in 2014 were searched for mentions of QoL, determining whether any of this new material contradicted or complemented the codes and conclusions found in the analysis of 2011–2013 EBAs. If so, the text passage was extracted, and a consensus was found between DL and CB on respective modifications or amendments to the code summaries and conclusions.

Results

According to the G-BA website, 105 EBA assessments had been completed by December 2014. In 14 of these procedures, no dossier had been submitted. In this study, we analyzed those 91 EBAs with a dossier.

In Table 1, the instruments used for QoL measurement in these EBAs are listed. Up to five different measures were used with the EQ-5D [13], EQ VAS [13], SF-36 [14], EORTC QLQ-C30 [15], and different versions of the FACT questionnaire [16] being most prevalent. Twenty-three dossiers did not include any QoL data.

Using the predefined search terms, 18,630 hits were found in the documents of 2011–2013 of which 1769 hits were relevant text passages to be extracted. After reduction and abstraction, 44 different codes were created. Code summaries comprised between 92 (code name "QoL included") and 883 words (code name "minimal important difference"); an example for a short summary (158 words in the German original) is given below.

Code name Preference and satisfaction.
Code description Relates to the relationship between patient preference, patient satisfaction, and QoL.
Summary In the EBA on aclidinium bromide, the manufacturer included treatment satisfaction (with the inhalator) as an endpoint in the section on QoL. In an oral hearing, he justified this with the fact that both G-BA's rules of procedures and AM-NutzenV ("Arzneimittel-Nutzenbewertungsverordnung"; Regulation for Early Benefit Assessment of New Pharmaceuticals) viewed "especially mortality, morbidity and QoL as paramount"; thus, the manufacturer argued that additional patient-relevant endpoints such as patient preference might be possible. The IQWiG disagreed: higher satisfaction would need to be reflected in better QoL. In accordance with the IQWiG's benefit assessment, the G-BA stated:

"Furthermore, the pharmaceutical manufacturer analyzes patient preferences regarding usage of an inhalator as a surrogate for QoL with regard to treatment satisfaction. A questionnaire determining patient preferences is not suitable for a valid assessment of health-related QoL. Advantaged resulting from its handling should reflect in clinical effects such as reduction of side effects or COPD symptoms." (aclidinium bromide)

Similarly, the G-BA stated in another EBA:

"Treatment satisfaction is not specified as a patient-relevant endpoint in the AM-NutzenV, and it also cannot be equated with health-related QoL." (saxagliptin/metformin)

Conclusion Patient preference and treatment satisfaction are considered neither patient-relevant endpoints nor surrogate parameters for QoL.

In the following, we will report those findings that related to our research question, i.e., the methodological requirements regarding QoL data in the EBA.

Table 1 Instruments used for QoL measurement in 91 early benefit assessments completed between January 2011 and December 2014. Instruments are listed if they were regarded an operationalization of quality of life by manufacturer, IQWiG, and/or G-BA

Active ingredient	Start date	Date of G-BA decision	Indication	Instruments used as QoL measures	Number of subgroups[a]	Extent of additional benefit[a,b] 1 = lower than comparator 2 = not demonstrated 3 = not quantifiable 4 = minor 5 = considerable	Certainty of additional benefit[a,b] 1 = not demonstrated 2 = unclear 3 = indication 4 = hint 5 = proof 888 = orphan drug	Added benefit due to QoL explicitly stated by G-BA[c]
Abiraterone acetate	1-Oct-11	29-Mar-12	Prostate carcinoma	FACT-P	2	5/2	4/1	0/0
Abiraterone acetate (new indication)	15-Jan-13	4-Jul-13	Prostate carcinoma	ECOG-PS; FACT-P V.4	1	5	4	0
Aclidinium bromide	1-Oct-12	21-Mar-13	COPD	EQ-5D; EQ VAS; SGRQ	1	2	1	0
Afatinib	15-Nov-13	8-May-14	Non-small cell lung cancer	EORTC QLQ-C30; EQ-5D	3	5/4/1	4/3/4	1/0/0
Aflibercept (age-related macular degeneration)	15-Dec-02	6-Jun-13	Neovascular, age-related macular degeneration	EQ-5D; EQ VAS; NEI VFQ-25	1	2	1	0
Aflibercept (macula edema)	1-Oct-13	20-Mar-14	Macular edema due to central or branch retinal vein occlusion	NEI VFQ-25	1	2	1	0
Aflibercept (metastatic colorectal carcinoma)	1-Mar-13	15-Aug-13	Metastatic colorectal cancer	None	1	4	4	0
Aliskiren/amlodipin	15-May-11	3-May-12	Essential hypertension	None	1	2	1	0
Apixaban	15-Jun-11	7-Jun-12	Prevention of venous thromboembolic events after knee or hip replacement operations	None	2	2/4	1/4	0/0
Apixaban (new indication)	1-Jan-13	20-Jun-13	Stroke prevention	None	1	4	4	0
Axitinib	1-Oct-12	21-Mar-13	Renal cell carcinoma	EQ-5D; EQ-VAS; FKSI-15; FKSI-DRS; SF-36	2	2/4	1/4	0/0
Belatacept	15-Jul-11	5-Jul-12	Kidney transplantation	SF-36	2	4/4	4/4	0/0
Belimumab	27-Jul-11	2-Aug-12	Systemic lupus erythematosus (SLE)	EQ-5D; EQ VAS; FACIT-Fatigue; SF-36	1	5	4	0
Boceprevir	1-Sep-11	1-Mar-12	Chronic hepatitis C	EQ-5D; EQ VAS	2	3/3	4/4	0/0
Bosutinib	1-May-13	17-Oct-13	Chronic myeloid leukemia	EQ-5D; EQ VAS; FACT-Leu	1	3	5	0
Brentuximab vedotin	1-Dec-12	16-May-13	Hodgkin lymphoma, anaplastic large cell lymphoma	None	3	3/3/3	888/888/888	0/0/0
Cabazitaxel	15-Apr-11	29-Mar-12	Prostate carcinoma	None	2	4/2	4/1	0/0

Table 1 continued

Active ingredient	Start date	Date of G-BA decision	Indication	Instruments used as QoL measures	Number of subgroups[a]	Extent of additional benefit[a,b] (1 = lower than comparator, 2 = not demonstrated, 3 = not quantifiable, 4 = minor, 5 = considerable)	Certainty of additional benefit[a,b] (1 = not demonstrated, 2 = unclear, 3 = indication, 4 = hint, 5 = proof, 888 = orphan drug)	Added benefit due to QoL explicitly stated by G-BA[c]
Canagliflozin	15-Mar-14	4-Sep-14	Diabetes mellitus type II	IWQoL-Lite; SF-36; EQ-5D	5	2/2/2/2/2	1/1/1/1/1	0/0/0/0/0
Cholic acid	15-May-14	6-Nov-14	Inborn errors in primary bile acid synthesis	None	1	3	2	0
Cobicistat	1-Apr-14	18-Sep-14	HIV infection	None	1	2	1	0
Colestilan	1-Apr-13	1-Oct-13	Hyperphosphatemia	None	1	2	1	0
Crizotinib	15-Nov-12	2-May-13	Non-small cell lung cancer	EORTC QLQ-C30; EORTC QLQ-LC13; EQ-5D; EQ VAS; VSAQ	2	5/2	3/1	1/0
Dabrafenib	1-Oct-13	3-Apr-14	Melanoma	EORTC QLQ-C30; EQ-5D; FACT-M	1	2	1	0
Dapagliflozin	15-Dec-12	6-Jun-13	Diabetes mellitus type II	DTSQ; EQ-5D	4	2/2/2/2	1/1/1/1	0/0/0/0
Dapagliflozin/Metformin	15-Feb-14	7-Aug-14	Diabetes mellitus type II	None	3	2/2/2	1/1/1	0/0/0
Decitabine	1-Nov-12	2-May-13	Myeloid leukemia	EORTC QLQ-C30	1	4	888	0
Dimethyl fumarate	1-May-14	16-Oct-14	Multiple sclerosis	None	1	2	1	0
Dolutegravir	15-Feb-14	7-Aug-14	HIV infection	None	5	5/2/4/2/2	5/1/4/1/1	0/0/0/0/0
Elosulfase alfa	1-Jun-14	20-Nov-14	Mucopolysaccharidosis type IVA (Morquio's syndrome, MPS IVA)	None	1	4	2	0
Elvitegravir, cobicistat, emtricitabine, tenofovir disoproxil	15-Jun-13	5-Dec-13	HIV infection	None	2	2/2	1/1	0
Emtricitabine, rilpivirine, tenofovir disoproxil	15-Jan-12	5-Jul-12	HIV infection	SF-36	1	4	5	0
Emtricitabine, rilpivirine, tenofovir disoproxil (new indication)	1-Jan-14	19-Jun-14	HIV-1-infection and viral load $\leq$100.000 HIV-1-RNA-copies/ml, without resistance-associated mutations of the HIV-1	None	1	2	1	0

Table 1 continued

Active ingredient	Start date	Date of G-BA decision	Indication	Instruments used as QoL measures	Number of subgroups[a]	Extent of additional benefit[a,b] 1 = lower than comparator 2 = not demonstrated 3 = not quantifiable 4 = minor 5 = considerable	Certainty of additional benefit[a,b] 1 = not demonstrated 2 = unclear 3 = indication 4 = hint 5 = proof 888 = orphan drug	Added benefit due to QoL explicitly stated by G-BA[c]
Enzalutamide	1-Sept-13	20-Feb-14	Prostate carcinoma	FACT-P; EQ-5D	1	5	4	0
Eribulin	1-May-11	19-Apr-12	Breast cancer	None	2	4/1	3/3	0/0
Extract from cannabis sativa	1-Jul-11	21-Jun-12	Spasticity in multiple sclerosis	EQ-5D; EQ VAS; MSQOL-54; SF-36	1	4	3	0
Fampridine	29-Jul-11	2-Aug-12	Multiple sclerosis	MSWS-12	1	2	1	0
Fidaxomicin	15-Jan-13	4-Jul-13	Clostridium infection	None	2	2/5	1/5	0
Fingolimod	15-Apr-11	29-Mar-12	Multiple sclerosis	EQ-5D; EQ VAS; PRIMUS-QoL	3	2/2/4	1/1/3	0/0/0
Fingolimod (new indication)	1-Jul-14	18-Dec-14	Multiple sclerosis	Primus QoL; EQ-5D; mFIS; Primus Activities	1	2	1	0
Fluticasone furoate/vilanterol trifenatate	1-Jan-14	20-Mar-14	Asthma, COPD	None	1	2	1	0
Indacaterol/glycopyrronium	15-Nov-13	8-May-14	COPD	SGRQ-C	2	4/4	3/4	0/0
Ingenol mebutate	15-Jan-13	4-Jul-13	Actinic keratosis	Skindex-16; TSQM	1	2	1	0
Insulin degludec	1-May-14	16-Oct-14	Diabetes mellitus	SF-36	4	2/2/2/2	1/1/1/1	0/0/0/0
Insulin degludec (new indication)	15-Jun-14	4-Dec-14	Diabetes mellitus	None	1	2	1	0
Ipilimumab	1-Aug-11	2-Aug-12	Melanoma	EORTC QLQ-C30	1	5	4	0
Ipilimumab (new indication)	15-Dec-13	5-Jun-14	Advanced (non-operable or metastasized) melanoma	EORTC QLQ-C30	2	2/2	1/1	0/0
Ivacaftor	15-Aug-12	7-Feb-13	Cystic fibrosis	CFQ-R; EQ-5D	2	4/5	888/888	0/1
Linaclotide	1-May-13	17-Oct-13	Irritable bowel syndrome with obstipation	EQ-5D; EQ VAS; IBS-QoL; SF-12; WPAI:IBS-C	1	2	1	0
Linagliptin	1-Oct-11	29-Mar-12	Diabetes mellitus type II	DTSQs; EQ VAS	1	2	1	0
Linagliptin (par. 5b)	1-Sep-12	21-Feb-13	Diabetes mellitus type II	DTSQ; EQ VAS	1	2	1	0
Lisdexamfetamine dimesylate	1-Jun-13	14-Nov-13	Attention-deficit/hyperactivity syndrome (ADHS) in children from the age of 6	HUI-Mark 2 and 3; WFIRS-P	1	2	1	0

Table 1 continued

Active ingredient	Start date	Date of G-BA decision	Indication	Instruments used as QoL measures	Number of subgroups[a]	Extent of additional benefit[a,b] 1 = lower than comparator; 2 = not demonstrated; 3 = not quantifiable; 4 = minor; 5 = considerable	Certainty of additional benefit[a,b] 1 = not demonstrated; 2 = unclear; 3 = indication; 4 = hint; 5 = proof; 888 = orphan drug	Added benefit due to QoL explicitly stated by G-BA[c]
Living maggots of *Lucilia sericata*	1-Jun-14	20-Nov-14	Wound healing/debridement	None	1	2	1	0
Lixisenatide	15-Mar-13	5-Sep-13	Diabetes mellitus type II	PAGI-QOL	4	2/2/2/2	1/1/1/1	0/0/0/0
Lomitapide	15-Dec-14	5-Jun-14	Hypercholesterolemia	None	1	2	1	0
Macitentan	1-Feb-14	17-Jul-14	Pulmonary arterial hypertension	SF-36; CAMPHOR	1	4	2	0
Microbial collagenase from *Clostridium histolyticum*	1-May-11	19-Apr-12	Dupuytren's contracture	None	4	2/2/2/2	1/1/1/1	0/0/0/0
Mirabegron	1-Jun-14	20-Nov-14	Overactive bladder	OABq; KHQ; PPBC	1	2	1	0
Ocriplasmin	1-May-13	17-Oct-13	Vitreomacular traction	NEI VFQ-25	3	5/5/2	3/3/1	0/0/0
Olodaterol	15-May-14	17-Jul-14	COPD	None	1	2	1	0
Pasireotide	15-Jun-12	6-Dec-12	Pituitary dysfunction	CushingQoL	1	4	888	0
Perampanel	15-Sep-12	7-Mar-13	Partial epilepsy	QOLIE-31-P	1	2	1	0
Perampanel (par. 5b)	15-May-14	6-Nov-14	Partial epilepsy	QOLIE-31-P	1	2	1	0
Pertuzumab	1-Apr-13	1-Oct-13	Breast cancer	FACT-B	3	5/2/2	3/1/1	0/0/0
Pirfenidone	15-Sep-11	15-Mar-12	Idiopathic pulmonary fibrosis	SGRQ; WHO QoL	1	3	888	0
Pixantrone	1-Dec-12	16-May-13	Non-Hodgkin's lymphoma	None	1	2	1	0
Pomalidomide	1-Sep-13	20-Feb-14	Multiple myeloma	EORTC QLQ-MY20; EORTC QLQ-C30; EQ-5D	1	5	2	0
Ponatinib	1-Aug-14	23-May-14	Lymphoblastic leukemia, myelogenous leukemia	None	2	3/3	2/2	0/0
Radium-223-dichloride	1-Jan-14	19-Jun-14	Prostate carcinoma	FACT-P	2	2/5	1/4	0/0
Regadenoson	15-Apr-11	29-Mar-12	Analysis of myocardial perfusion	None	1	2	1	0
Regorafenib	1-Oct-14	20-Mar-14	Colorectal carcinoma	EORTC QLQ-C30; EQ-5D	1	4	3	0
Retigabine	15-May-11	3-May-12	Epilepsy	QOLIE-31-P	1	2	1	0
Retigabine (par. 5b)	15-Jan-14	7-Jul-14	Epilepsy	None	1	2	1	0
Rilpivirine	15-Jan-12	5-Jul-12	HIV infection	SF-36	1	4	5	0
Riociguat	1-May-14	16-Oct-14	Pulmonary arterial hypertension	EQ-5D; LPH	2	4/4	2/2	0/0

Table 1 continued

Active ingredient	Start date	Date of G-BA decision	Indication	Instruments used as QoL measures	Number of subgroups[a]	Extent of additional benefit[a,b] 1 = lower than comparator 2 = not demonstrated 3 = not quantifiable 4 = minor 5 = considerable	Certainty of additional benefit[a,b] 1 = not demonstrated 2 = unclear 3 = indication 4 = hint 5 = proof 888 = orphan drug	Added benefit due to QoL explicitly stated by G-BA[c]
Ruxolitinib	15-Sep-12	7-Mar-13	Chronic myeloproliferative disorders	ECOG status; EORTC QLQ-C30; FACT-Lym	1	4	888	0
Ruxolitinib (par. 5b)	15-May-14	6-Nov-14	Chronic myeloproliferative diseases	EORTC QLQ-C30; FACT-Lym	1	5	3	1
Saxagliptin	1-Apr-13	1-Oct-13	Diabetes mellitus type II	EQ-5D	4	4/2/2	3/1/1/1	0/0/0/0
Saxagliptin/metformin	15-Nov-12	2-May-13	Diabetes mellitus type II	EQ-5D	2	4/2	3/1	0/0
Saxagliptin/metformin (new indication)	1-Apr-13	1-Oct-13	Diabetes mellitus type II	None	1	2	1	0
Siltuximab	15-Jun-14	4-Dec-14	Multicentric Castleman's disease (MCD)	SF-36; FACIT-F; MCD-SS	1	3	2	1
Simeprevir	1-Jun-14	20-Nov-14	Chronic hepatitis C	EQ-5D; WPAI	7	5/5/4/2/4/2	4/4/4/3/1/3/1	0/0/0/0/0/0/0
Sitagliptin	1-Apr-13	1-Oct-13	Diabetes mellitus type II	EQ VAS	5	4/4/2/2/2	3/3/1/1/1	0/0/0/0/0
Sitagliptin/metformin	1-Apr-13	1-Oct-13	Diabetes mellitus type II	EQ VAS	3	4/2/2	3/1/1	0/0/0
Sofosbuvir	1-Feb-14	17-Jul-14	Chronic hepatitis C	SF-36	1	5	4	0
Tafamidis meglumine	15-Dec-11	7-Jun-12	Amyloidosis	SF-36; TQoL	1	4	888	0
Tegafur, gimeracil, oteracil	1-Jul-12	20-Dec-12	Gastric cancer	None	1	2	1	0
Telaprevir	15-Oct-11	29-Mar-12	Hepatitis C	EQ-5D; FSS; questions on health care utilization (visits to physician); WPQ	2	3/3	2/2	0/0
Teriflunomide	1-Oct-13	20-Mar-14	Multiple sclerosis	FIS	1	2	1	0
Ticagrelor	1-Jan-11	15-Dec-11	Acute coronary syndrome	EQ-5D	4	5/2/2/2	5/1/1/1	0/0/0/0
Trastuzumab emtansine	1-Jan-14	19-Jun-14	Breast cancer	TOI-PFB; FACT-B	2	2/5	1/4	0/0
Turoctocog alfa	15-Jan-14	3-Jul-14	Hemophilia A	None	1	2	1	0
Vandetanib	15-Mar-12	6-Sep-12	Neoplasia of the thyroid gland	FACIT-D; FACT-G	1	2	1	0
Vandetanib (par. 5b)	15-Mar-13	5-Sep-13	Neoplasia of the thyroid gland	FACIT-D; FACT-G; WHO PS	1	4	3	0

Table 1 continued

Active ingredient	Start date	Date of G-BA decision	Indication	Instruments used as QoL measures	Number of subgroups[a]	Extent of additional benefit[a,b] 1 = lower than comparator 2 = not demonstrated 3 = not quantifiable 4 = minor 5 = considerable	Certainty of additional benefit[a,b] 1 = not demonstrated 2 = unclear 3 = indication 4 = hint 5 = proof 888 = orphan drug	Added benefit due to QoL explicitly stated by G-BA[c]
Vemurafenib	15-Mar-12	6-Sep-12	Melanoma	FACT-M; VAS pain	1	5	4	0
Vemurafenib (renewed assessment after expiration of the deadline)	15-Sep-13	6-Mar-14	Melanoma	FACT-M	1	5	4	0
Vildagliptin	1-Apr-13	1-Oct-13	Diabetes mellitus type II	DSC-r; DTSQ; SF-36; WPAI-GH	5	2/2/2/2	1/1/1/1	0/0/0/0/0
Vildagliptin/metformin	1-Apr-13	1-Oct-13	Diabetes mellitus type II	DSC-r; DTSQ; SF-36; WPAI-GH	3	2/2/2	1/1/1	0/0/0
Vismodegib	15-Aug-13	6-Feb-14	Basal cell carcinoma	SF-36	2	2/4	1/3	0/0

[a] According to the G-BA

[b] In case subgroups were defined, values for each subgroup are given

[c] In the "Tragende Gründe" document, which summarizes the decision and the main underlying reasons

Examples for EBAs on which the findings are based are given in brackets. In some cases, the same drug was subject of two EBAs because it was evaluated for two different conditions or because it was re-evaluated according to §35b par. 5b SGB V subsequent to a formerly incomplete dossiers. In these cases, the second EBA is identified with "new condition" or "par. 5b", respectively.

Construct measured by a questionnaire

The IQWiG understands QoL as "a complex construct comprising psychological, physical, and social domains", as discussed above. Accordingly, instruments or subscales of an instrument evaluating symptoms only were classified as morbidity measures by the IQWiG and thus could not prove benefit regarding QoL (e.g., the Prostate Cancer Subscale of the FACT-P [17], abiraterone acetate new condition, or symptom scales of the EORTC QLQ-C30 [15], dabrafenib). Pain was also classified a morbidity outcome (vemurafenib) instead of being a QoL measure. Fatigue was classified as an aspect of morbidity by the IQWiG (telaprevir), whereas the G-BA subsumed it sometimes under morbidity and sometimes under QoL within the same EBA (ruxolitinib)—the latter probably only due to an error in phrasing, as the G-BA also subsumed fatigue under morbidity in the decision on siltuximab. However, a questionnaire on the influence of fatigue on the patient's life—instead of fatigue intensity itself—was regarded a measure of QoL (teriflunomide).

Appraisal of the EQ-5D index score [13] was inconsistent: While accepted as a QoL measure in 2012 (fingolimod), it was later rejected because it was regarded as a utility measure with weightings not based on patient judgments (simeprevir, regorafenib). The single items of the EQ-5D, however, were still accepted as a valid QoL measure (regorafenib). The visual analogue scale (EQ VAS) on patient-rated current health state was now regarded a measure of morbidity (radium-223-dichlorid; dabrafenib).

Questionnaires measuring treatment satisfaction (saxagliptin/metformin) or patient preferences regarding route of drug administration (aclidinium bromide) were not accepted as QoL measures or surrogate measures of QoL by either IQWiG or G-BA. The same was true for measures of work productivity and health care utilization (telaprevir; simeprevir) and for assessments made by the study physician instead of the patient (fampridine; vandetanib par. 5b). However, if a questionnaire only measured subdomains of QoL without covering all areas of QoL from the IQWiG's point of view, it was still accepted as a QoL measure (e.g., PRIMUS QoL [18], fingolimod; TOI-PFB subscale of the FACT-B [19], trastuzumab emtansine). Assessing only some of an instrument's subscales was criticized by IQWiG

and G-BA (pertuzumab; ivacaftor), but did not necessarily lead to the disregard of the respective data (decitabine; trastuzumab emtansine).

Disease specificity of QoL instruments

Generic instruments measure QoL irrespective of the specific disease. They seemed to be accepted by IQWiG and G-BA as long as they had been correctly validated in studies including patients with the diagnosis in question (see below; sitagliptin/metformin). However, diagnosis-specific instruments were regarded more suitable, because they allow for drawing inferences on the specific impact the respective disease has on QoL (ruxolitinib; ipilimumab).

Questionnaires measuring QoL in cancer patients in general (instead of in specific types of cancer) such as the FACT-G [16] were regarded generic (ruxolitinib; ipilimumab), showing a rather narrow understanding of disease-specificity by IQWiG and G-BA.

Validation

Only validated QoL instruments were accepted. If references to validation publications were missing in the dossier, the respective data was disregarded in some EBAs (belatacept; vandetanib par. 5b), whereas in other cases the IQWiG conducted an own literature search (sitagliptin; telaprevir). At least some of the participants in the validation studies should have had the disease in question (a share of 86 % being accepted by the G-BA, pasireotide), but not meeting this requirement did not necessarily lead to the G-BA disregarding the data (ruxolitinib).

The G-BA also noted critically if a validation had been conducted with the data of the clinical trial included in the dossier itself, if the validation study had not been published yet (pasireotide) or if the questionnaire had been validated as a health utility instrument instead of a QoL measure (fingolimod). It was further mentioned if unfavorable validation results had been reported in the literature (ceiling effects, decitabine; low sensitivity to change, riociguat).

If any changes had been made to the QoL questionnaires, the IQWiG did not accept the data anymore—be it adding an item (pertuzumab) or calculating a global score for all but one subscale that related to surgery in a study population that did not have any surgery (vemurafenib).

However, from the IQWiG's point of view, an instrument did not need to be established in order to be relevant for benefit assessment, as long as it was validated (fingolimod), whereas the G-BA explicitly noted for some questionnaires that they were established (belimumab), widely used (ivacaftor), or recommended for the respective indication (decitabine).

Risk of bias

In dossier evaluations, the IQWiG regularly rated the risk of bias for each outcome. As a rule, QoL data from unblinded studies were rated potentially highly biased (belatacept).

Another frequent reason for a high potential of bias in QoL data were missing values. If data could be analyzed for less than 90 % of randomized study participants (for example because baseline, but not follow-up data were available), the outcome was rated potentially highly biased (telaprevir), which can reduce the certainty of added benefit determined by the IQWiG (vemurafenib). A percentage higher than 30 % (boceprevir; radium-223-dichlorid) or a percentage difference between study groups higher than 15 % (Fingolimod, new condition) led to the disregard of the outcome. The institute, however, took into account if missing data were due to mortality (ipilimumab). Moreover, QoL data with more than 30 % missing values may still be accepted if conservative sensitivity analyses show favorable results (radium-223-dichlorid).

Study design

QoL did not need to be the primary endpoint of the clinical trial in order to be accepted as evidence (ivacaftor), as long as the analyses had been specified in the study protocol (see below). QoL also did not need to be assessed beyond disease progression and thus longer than other outcomes, although this was recommended by IQWiG employees in oral hearings (pertuzumab; abiraterone acetate, new condition).

Another design issue discussed in the oral hearings was the frequency of QoL assessments within a trial. The IQWiG, however, did not take a stand on this issue, except for discussing a risk of low data quality due to weekly assessment schedules (vandetanib).

Concordance with study protocol

As with other outcomes, it was important for IQWiG and G-BA that QoL outcomes presented in the dossier concurred with the specifications in the study protocol or statistical analysis plan. Deviations were explicitly—and apparently critically—noted in dossier evaluations, especially if they had not been justified in the dossier. Deviations included non-reporting of assessed QoL data (crizotinib; pertuzumab), differences between planned and reported statistical analyses (ipilimumab; crizotinib), and unplanned, post hoc analyses (pertuzumab; pirfenidone).

Effect sizes

QoL benefits had to be statistically significant at the 5 % level in order to be accepted (e.g., axitinib). QoL effects could consist in group differences in mean QoL scores, in differences in the number of QoL responders per group, or in group differences in the time until QoL improvement or deterioration (abiraterone acetate new condition; ivacaftor; trastuzumab emtansine).

For all three analyses, the minimal size of a QoL change that is patient-relevant needs to be determined. The relevance of differences in mean QoL scores was determined with a validated minimal important difference (MID) on group level (ruxolitinib) or, if no MID was available, with Hedges' g as an effect size measure with the lower level of the 95 % interval of confidence for g having to exceed 0.2 (axitinib; macitentan). For differences in the number of responders or in time to improvement, an individual level MID was needed (pirfenidone).

MIDs were only accepted if they had been validated and published (abiraterone acetate, new condition), ideally in more than one scientific article (ivacaftor) and in an original paper instead of in a congress abstract only (ruxolitinib).

There were also a range of requirements for the methodology of MID validation. Anchor-based approaches seemed to be preferred over distribution-based approaches (ruxolitinib), with the anchor having to be a patient-relevant parameter such as patients' global assessment of change (ruxolitinib) or visual acuity measured by number of lines (ocriplasmin). Physicians' global assessment and laboratory parameters, in contrast, were not accepted as patient-relevant anchor variables (abiraterone acetate, new condition).

MIDs had to be validated in a patient population similar to the trial population (e.g., regarding age: ivacaftor; and diagnosis: ruxolitinib), because they were regarded sensitive to context. However, an MID validation conducted with the trial population itself was criticized by the G-BA (pasireotide). Validation studies should furthermore be longitudinal instead of cross-sectional (abiraterone acetate, new condition).

As with QoL analysis in general, the MID used and the MID analyses to be conducted had to be specified in the study protocol which the dossier should strictly adhere to (decitabine). The MID had to be validated for the exact version of a questionnaire that was actually used in the trial (pertuzumab).

Comparing the number of responders or the time to a relevant QoL change, an individual level MID as a response criterion is required. G-BA and IQWiG critically noted that analyzing the percentage of patients with QoL

improvement does not account for a possible QoL deterioration (which was to be expected in the respective disease, radium-223-dichlorid). Conversely, an analysis of time until QoL deterioration does not account for possible QoL improvements, but was accepted with reference to the natural disease course where QoL worsens (trastuzumab emtansine).

Missing information on QoL analysis

In several EBAs, the IQWiG and, less often, the G-BA criticized that QoL parameters were presented incompletely in the dossiers. This included sample sizes and statistical tests (aclidinium bromide), standard errors (boceprevir), intervals of confidence (boceprevir), QoL levels at study onset (ivacaftor) or at both study onset and end while only presenting delta values (crizotinib), as well as effect sizes measures (boceprevir). Missing or merely graphically presented sub-group analyzes were also criticized (vemurafenib; vandetanib par. 5b).

If QoL data were regarded as insufficient, this resulted in a temporal limitation of the G-BA's decision on additional benefit (eribulin, ocriplasmin, pertuzumab).

Discussion

From our analysis of EBAs conducted within the first 4 years after introduction of the AMNOG, it became apparent that there are quite rigorous requirements for the assessment, analysis, and presentation of QoL data. The IQWiG's methodological standards have often been criticized as being too rigid [20, 21], a point that was also discussed in the oral hearings (critique by pharmaceutical manufacturer: abiraterone acetate; critique by patient representative of the G-BA: pertuzumab). The fact that strict standards are also applied to QoL—a subjective and relatively new outcome—seems to reflect an appreciation of QoL equivalent to the appreciation of more traditional outcomes.

By contrast, QoL data were not included in 23 out of 91 benefit dossiers. In only five benefit dossiers, the G-BA stated an added benefit regarding QoL. The reason for this discrepancy is probably that clinical trials, forming the basis of benefit dossiers, take years to plan and conduct and thus had been designed in pre-AMNOG times where QoL played a minor role. Moreover, the G-BA did not explicitly define QoL, which resulted in different understandings of QoL. Manufacturers often had a wider understanding of QoL than the IQWiG or G-BA. As a consequence, there is often a lack of QoL data for benefit assessment [22].

When QoL data were included in the dossiers, they often were not accepted as evidence for added benefit for methodological reasons. Reliable QoL assessment can be perceived as methodologically challenging [23] and it may not have been possible for manufacturers to anticipate the IQWiG's and G-BA's expectations regarding QoL evidence. The requirements identified on the basis of the first 91 benefit evaluations go far beyond what is specified in the SGB V §35a, the corresponding regulation (AM-NutzenV), the G-BA's rules of procedures, and the IQWiG's General Methods paper. This applies, for example, to the definition of QoL (and hence the instruments accepted for measuring it) and to the methods accepted for questionnaire and MID validation.

QoL in the EBA has been the subject of previous studies. On the basis of 43 completed EBAs, Klose et al. [22] described the role of QoL in an unsystematic review. They highlight, for example, that validated measures and response criteria need to be used. Other unsystematic reviews on the EBA did not focus on QoL, but found that QoL data were only partly included in EBAs [4], possibly due to methodological challenges in measuring this outcome reliably [23]. However, to our knowledge, our study has been the first to use a systematic qualitative approach to analyze the EBA in Germany.

A different, complementary approach was taken by Fischer and Stargardt [3] who analyzed inputs and outputs of EBAs quantitatively. They found that there was no statistically significant association between the inclusion of QoL data in benefit dossiers and the G-BA's rating decision. Our qualitative approach showed that this may in part be explained by non-compliance with the various methodological requirements found in our analysis, so that in most cases, the mere inclusion of QoL data in the dossier did not lead to a positive evaluation of QoL benefit. In addition, many EBAs did include QoL outcomes, but there were no statistically or clinically significant effects.

The methodological requirements of QoL assessment we found are based on individual benefit evaluations, implying that changes in the IQWiG's and G-BA's decision-making in future EBAs are certainly possible, as has been the case in the past with the EQ-5D index not being accepted anymore as a QoL endpoint. However, our study did show that various, and often recurring, standards have been applied in the past. Compliance with these requirements when compiling a benefit dossier—and, ideally, when designing, conducting, and analyzing clinical trials—will improve the chances of a positive evaluation in the early assessment of benefit.

References

1. Franken, M., Le Polain, M., Cleemput, I., Koopmanschap, M.: Similarities and differences between five European drug reimbursement systems. Int. J. Technol. Assess. Health Care **28**(4), 349–357 (2012)

2. Henschke, C., Sundmacher, L., Busse, R.: Structural changes in the German pharmaceutical market: price setting mechanisms based on the early benefit evaluation. Health Policy **109**(3), 263–269 (2013)

3. Fischer, K.E., Stargardt, T.: Early benefit assessment of pharmaceuticals in Germany: manufacturers' expectations versus the Federal Joint Committee's decisions. Med. Decis. Making **34**(8), 1030–1047 (2014)

4. Hörn, H., Nink, K., McGauran, N., Wieseler, B.: Early benefit assessment of new drugs in Germany—results from 2011 to 2012. Health Policy **116**(2–3), 147–153 (2014)

5. Leopold, C., Vogler, S., Mantel-Teeuwisse, A.K., de Joncheere, K., Leufkens, H.G.M., Laing, R.: Differences in external price referencing in Europe: a descriptive overview. Health Policy **104**(1), 50–60 (2012)

6. Ahmed, S., Berzon, R.A., Revicki, D.A., Lenderking, W.R., Moinpour, C.M., Basch, E., Reeve, B.B., Wu, A.W.: The use of patient-reported outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. Med. Care **50**(12), 1060–1070 (2012)

7. Kreis, J., Busse, R.: From evidence assessments to coverage decisions? The case example of glinides in Germany. Health Policy **104**(1), 27–31 (2012)

8. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG): General Methods (Benefit Assessment). https://www.g-ba.de/informationen/nutzenbewertung/ (2015). Accessed 25 March 2015

9. Schipper, H., Clinch, J.J., Olweny, C.L.: Quality of life studies: definitions and conceptual issues. In: Spilker, B. (ed.) Quality of Life and Pharmacoeconomics in Clinical Trials, 2nd edn, pp. 15–18. Lippincott-Raven, Philadelphia (1996)

10. Mayring, P.: Qualitative content analysis. In: Flick, U., Kardorff, E.V., Steinke, I. (eds.) A Companion to Qualitative Research, pp. 266–269. Sage Publications, London (2004)

11. Meyer, T., Karbach, U., Holmberg, C., Güthlin, C., Patzelt, C., Stamer, M.: Qualitative Studien in der Versorgungsforschung— Diskussionspapier, Teil 1: gegenstandsbestimmung [Qualitative research in health services research—discussion paper, part 1: what is the idea?]. Gesundheitswesen **74**(8–9), 510–515 (2012)

12. Gemeinsamer Bundesausschuss (G-BA): Übersicht der Wirkstoffe. https://www.g-ba.de/informationen/nutzenbewertung/ (2015). Accessed 25. March 2015

13. The EuroQol Group: EuroQol—a new facility for the measurement of health-related quality of life. Health Policy **16**(3), 199–208 (1990)

14. Ware, J.E.: The SF-36 health survey. In: Spilker, B. (ed.) Quality of Life and Pharmacoeconomics in Clinical Trials, pp. S337–S346. Lippincott-Raven, Philadelphia (1996)

15. Aaronson, N.K., Ahmedzai, S., Bergman, B., et al.: The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J. Natl. Cancer Inst. **85**(5), 365–376 (1993)

16. Cella, D.F., Tulsky, D.S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., Silberman, M., Yellen, S.B., Winicour, P., Brannon, J., et al.: The functional assessment of cancer therapy scale: development and validation of the general measure. J. Clin. Oncol. **11**(3), 570–579 (1993)

17. Esper, P., Mo, F., Chodak, G., Sinner, M., Cella, D., Pienta, K.J.: Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. Urology **50**(6), 920–928 (1997)

18. McKenna, S.P., Doward, L.C., Twiss, J., et al.: International development of the patient-reported outcome indices for multiple sclerosis (PRIMUS). Value Health **13**, 946–951 (2010)

19. Brady, M., Cella, D.F., Mo, F., Bonomi, A.E., Tulsky, D.S., Lloyd, S.R., Deasy, S., Cobleigh, M., Shiomoto, G.: Reliability and validity of the functional assessment of cancer therapy—breast quality-of-life instrument. J. Clin. Oncol. **15**, 974–986 (1997)

20. Perleth, M., Gibis, B., Göhlen, B.: A short history of health technology assessment in Germany. Int. J. Technol. Assess. Health Care **25**(1), 112–119 (2009)

21. Sattelmeier, J., Prenzler, A., Frank, M.: Das Arzneimittelmarkt-neuordnungsgesetz (AMNOG) in der Praxis—erste Erfahrungen, Kritikpunkte und Weiterentwicklungsmöglichkeiten. Gesundheitsökonomie & Qualitätsmanagement **18**, 213–220 (2013)

22. Klose, K., Witte, J., Kreimeier, S., Greiner, W.: Lebensqualität als Wettbewerbsvorteil in der frühen Nutzenbewertung [Quality of life as advantage in competition within the early benefit assessment]. Public Health Forum **21**(4), 14. e1–14. e3 (2013)

23. Ruof, J., Schwartz, F., Schulenburg, J.-M., Dintsios, C.-M.: Early benefit assessment (EBA) in Germany: analysing decisions 18 months after introducing the new AMNOG legislation. Eur. J. Health Econ. **15**(6), 577–589 (2014)