Early Benefit Assessment of Pharmaceuticals in Germany: Manufacturers' Expectations versus the Federal Joint Committee's Decisions

Katharina E. Fischer, Dr., Tom Stargardt, Prof. Dr.

Background. *Since 2011, when the German Pharmaceutical Market Restructuring Act (AMNOG) came into effect, newly licensed pharmaceuticals must demonstrate an added benefit over a comparator treatment to be reimbursed at a value greater than the reference price. Evidence submitted by manufacturers is assessed by the Institute for Quality and Efficiency in Health Care (IQWiG) and subsequently appraised by the German Federal Joint Committee (FJC) as part of so-called early benefit assessments (EBA). This study aims to explain the decisions made, clarify the roles of the parties (manufacturers, IQWiG, FJC) involved, and guide manufacturers in developing future submissions by analyzing 42 EBAs concluded since January 2011.* ***Methods.*** *We developed a variable list representing the essential components of the EBA: the rating decisions of manufacturers, IQWiG, and the FJC regarding each pharmaceutical's added benefit; the characteristics of the pharmaceutical; the characteristics of the EBA process; the types of evidence submitted; the methods used to generate evidence; and the pharmaceutical's maximum possible budget impact. We used Cohen's kappa to analyze agreement between the rating decisions of the different parties. The chi-square test and bivariate regression were used to identify associations between components of the EBA process and the rating decisions of the FJC.* ***Results.*** *We observed a low level of agreement between manufacturers and the FJC (kappa = 0.21; 95% CI 0.107–0.31) and a substantial level of agreement between IQWiG and the FJC (kappa = 0.64; 95% CI 0.451–0.827) in their rating decisions. The characteristics of the EBA process—for example, duration of the process (*$P = 0.357$*), participation in the official hearing (*$P = 0.227$*), and the pharmaceutical's budget impact (*$P = 0.725$*)—did not have a significant effect on the rating decisions of the FJC. There was, however, an association between the type of evidence submitted and the FJC's rating decision when the manufacturer's dossier reported outcomes related to morbidity (*$P = 0.009$*) or adverse events (*$P < 0.001$*) but not mortality (*$P = 0.718$*) or quality of life (*$P = 0.783$*).* ***Conclusions.*** *While the FJC tends to disagree with the rating of benefit by manufacturers, it softens IQWiG's decisions, potentially to make the final outcome more acceptable. Concerns voiced that the FJC might be exceeding its statutory authority by taking cost or procedural considerations into account appear to be unfounded. Choosing appropriate evidence to submit for each endpoint remains a challenge, as submission of health outcomes evidently influences decisions.* ***Key words:*** *pharmaceutical; coverage; reimbursement; Germany; Federal Joint Committee; Institute for Quality and Efficiency in Health Care; IQWiG, AMNOG; pharmaceutical market regulation; fourth-hurdle decision making.* ***(Med Decis Making 2014;34:1030–1047)***

Until 2011, virtually any newly licensed pharmaceutical launched on the German market was covered by statutory health insurance regardless of its price.[1] The only exceptions were over-the-counter medications and so-called lifestyle drugs, such as those used to treat erectile dysfunction or male pattern baldness in generally healthy individuals. Since 1 January 2011, however, all newly introduced pharmaceuticals in Germany with new active ingredients, or new combinations of active ingredients, are systematically assessed to determine whether they have an added benefit over an appropriate comparator treatment. Pharmaceuticals that show an added benefit are subject to price

DOI: 10.1177/0272989X14546377

negotiations between the manufacturer and statutory health insurers. Pharmaceuticals that do not show an added benefit are priced in reference to a group of pharmaceuticals that, while not identical, are considered roughly interchangeable (i.e., through a system known as ''reference pricing'').[2]

The requirement to demonstrate an added benefit was introduced by the Pharmaceutical Market Restructuring Act (*Arzneimittelmarktneuordnungsgesetz*, or AMNOG), which came into effect on 1 January 2011. The act specifies rules, responsibilities, and a timetable for all parties involved. At market launch, the manufacturer is still free to set the price of a new pharmaceutical as it sees fit. This price is valid for 1 year. Within 3 months after market launch, however, the pharmaceutical must undergo a process known as an early benefit assessment (EBA) by the German Federal Joint Committee (FJC).[3,4]* The FJC is a non-state self-governance body that includes payer, provider, and patient representatives. It is responsible for making coverage decisions for statutory health insurance.[5]†

As part of the EBA, the manufacturer submits a dossier of evidence. Among a range of materials, this dossier includes information on the manufacturer's choice of comparator treatment and its own rating of the new pharmaceutical's added benefit. The FJC subsequently commissions a third party to assess the added benefit of the pharmaceutical based on the manufacturer's dossier (stage 1). By convention, this third party is the Institute for Quality and Efficiency in Health Care (IQWiG). Within 3 months, IQWiG's assessment is made public, and a range of stakeholders may comment on it in writing or orally during an official hearing (see Table 1). This assessment includes IQWiG's own rating of the pharmaceutical's added benefit. Within a further 3 months, the FJC makes a final decision in which it, too, rates the pharmaceutical's added benefit (stage 2). It bases this decision on IQWiG's assessment, the manufacturer's dossier, and the stakeholders' comments. The process by which the FJC makes this decision is described in the literature as an appraisal.‡

If the FJC concludes that the pharmaceutical has an added benefit, the manufacturer and the Federal Association of Statutory Health Insurers negotiate a price (or, more precisely, a discount on the ex-factory price) at the substance level. This price goes into effect 1 year after the pharmaceutical was launched (stage 3). If an agreement cannot be reached within 6 months, an arbitration board settles a price (stage 4). A cost-effectiveness analysis using IQWiG's methods for economic evaluation may be requested by either party.[8] If no added benefit is found in stage 2, the pharmaceutical may become subject to reference pricing or to restrictions on the indications for which it can be used.

Because of the novelty of AMNOG, research to date has only described the reform and discussed the first wave of EBAs.[3–5,9,10] From an international perspective, Germany's 2-tiered approach to EBA, allocating as it does the processes of assessment and appraisal to 2 separate institutions (i.e., IQWiG and the FJC, respectively), has similarities and differences to the approaches taken in the UK or Australia.[11] Examples of similarities include the use of separate institutions for assessment and appraisal and the breadth of the

Received 5 August 2013 from University of Hamburg, Hamburg Center for Health Economics, Hamburg, Germany (KEF, TS). The data used in the statistical analysis can be obtained from the authors upon request. Financial support for this study was provided entirely by a grant from the Federal Ministry for Research and Education in Germany (grant number: BMBF 01EH1101A). The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. One author (TS) states that he is a member of a market access advisory board for Daiichi Sankyo. In this context, no financial support was provided for conducting this study. Revision accepted for publication 22 June 2014.

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://mdm.sagepub.com/supplemental.

Address correspondence to Katharina E. Fischer, Hamburg Center for Health Economics, Universität Hamburg, Esplanade 36, Hamburg, 20354, Germany; telephone: + 49 40 428 38 9296; fax: + 49 40 428 38 8043; e-mail: katharina.fischer@uni-hamburg.de.

*The term ''early benefit assessment'' has become established in the literature and is a direct translation of the German term ''frühe Nutzenbewertung.'' Although we will continue to use the established term, it should be noted that a more appropriate translation would be ''early assessment of [added] benefit,'' as it is the assessment that is early, not the benefit.

†Substances with a negligible impact on this expenditure (i.e., < €1 million per year) are excluded from the early benefit assessment.[6]

‡The distinction between the assessment and the appraisal of health technologies was formalized with the creation of the National Institute for Clinical Excellence in England and Wales in 1999.[7] *Assessment* in this context refers to the analytical process of gathering and summarizing information on the relevant aspects of a health technology, whereas *appraisal* refers to the political process of making a decision about health technologies while taking assessment information, values, and other factors into account (Stevens and Milne 2004, as cited by Velasco-Garrido and others[7]). Despite its name, the process known as ''early benefit assessment'' in Germany includes both an assessment stage (undertaken by IQWiG) and an appraisal stage (undertaken by the FJC).

Table 1 Variable Definitions

Variable	Definition	As Stated by . . .	Source Document
Pharmaceutical's added benefit as rated by the different parties			
Added benefit	Rating of benefit: major added benefit, considerable added benefit, minor added benefit, not quantifiable added benefit, no added benefit, or less benefit than comparator (patient subgroup level)	Manufacturer, IQWiG, FJC	Manufacturer: dossier, part 1, Table 1-10 IQWiG: assessment, summary FJC: appraisal documents
Added benefit binary	Rating of benefit: added benefit or no added benefit (patient subgroup or indication level)	Manufacturer, IQWiG, FJC	
Characteristics of the pharmaceutical			
Indication	Indication by ICD-10-GM	Manufacturer	Dossier, part 1, 1.3
ATC	Therapeutic main group of pharmaceutical defined by ATC code level 2 that defines the pharmacological or therapeutic subgroups	Manufacturer	Dossier, part 1, 1.2
Orphan drug	Pharmaceutical orphan drug or not orphan drug	Manufacturer	Dossier, part 4, 4.4.4
Characteristics of the EBA process			
Status	Status of early benefit assessment: completed, appraisal in preparation, discharged, no status, EBA started, or hearing started	FJC	http://www.g-ba.de/informationen/nutzenbewertung/
Duration	Number of days from EBA submission to decision by FJC (legally defined duration: 180 days, excluding submissions before 1 August 2011 because of exceptional regulations for length of procedure)	FJC	http://www.g-ba.de/informationen/nutzenbewertung/
Year	Year of the EBA decision: 2011, 2012, or 2013	FJC	http://www.g-ba.de/informationen/nutzenbewertung/
Hearing participants	Number of participants in hearing (2 representatives per each stakeholder that is eligible to submit written statements; see below for eligibility criteria)	FJC	Appraisal documents, general documentation, C.4
Written statements	Number of statements submitted during the hearing process (statements may be submitted for 3 weeks by medical and pharmaceutic experts; umbrella organizations of doctors, the pharmaceutical industry, pharmacists; the manufacturer)	FJC	Appraisal documents, general documentation, C.3
Preliminary decision	Decision was preliminary, or decision was not preliminary (FJC may state that a pharmaceutical is reappraised if a change in the evidence base is expected according to certain criteria or because of conditional market approval)	FJC	Appraisal documents

(continued)

Table 1 (continued)

Variable	Definition	As Stated by . . .	Source Document
Consultation requested	Manufacturer has requested consultation by FJC, or manufacturer has not requested consultation (advice meeting within 8 weeks prior to submission, only advice regarding content of submission documents, definition of comparator, studies to be submitted)	FJC	Appraisal documents
Methods used to generate evidence on a pharmaceutical's added benefit			
Systematic literature search presented	Manufacturer presented a systematic literature search or did not present a systematic search	Manufacturer	Dossier, part 4, 4.2.3.2
Meta-analysis presented	Manufacturer presented a meta-analysis or did not present a meta-analysis	Manufacturer	Dossier, part 4, 4.2.5.3
Sensitivity analysis presented	Manufacturer presented a sensitivity analysis or did not present a sensitivity analysis	Manufacturer	Dossier, part 4, 4.2.5.4
Indirect comparison presented	Manufacturer presented an indirect comparison between pharmaceuticals or did not present an indirect comparison	Manufacturer	Dossier, part 4, 4.2.5.6
Registry studies (number)	Total number of registry studies presented by manufacturer	Manufacturer	Dossier, part 4, Table 4-3
RCTs (number)	Number of randomized clinical trials presented by manufacturer	Manufacturer	Dossier, part 4, 4.3.1.1.4
Head-to-head trials (number)	Number of head-to-head trials presented by manufacturer	Manufacturer	Dossier, part 4, Table 4-6
Other studies (number)	Number of other studies (e.g. nonrandomized, observational) presented by manufacturer	Manufacturer	
Patients in intervention arms (number)	Total number of patients in studies as sum of patients in intervention arms	Manufacturer	Dossier, part 4, Table 4-5
Patients in intervention arms per study (normalized)	Number of patients per study, normalized to size of patient subgroup	Manufacturer	Dossier, part 4, Table 4-5
Type of evidence submitted			
Morbidity presented	Manufacturer has presented information about the endpoint morbidity or has not presented information	Manufacturer	Dossier, part 4
Mortality presented	Manufacturer has presented information about the endpoint mortality or has not presented information	Manufacturer	Dossier, part 4
Adverse events presented	Manufacturer has presented information about the endpoint adverse events or has not presented information	Manufacturer	Dossier, part 4
QoL presented	Manufacturer has presented information about the endpoint quality of life or has not presented information	Manufacturer	Dossier, part 4

(continued)

Table 1 (continued)

Variable	Definition	As Stated by . . .	Source Document
Patient subgroup rejected	Definition of patient subgroup rejected, or definition not rejected	FJC	Appraisal documents
Comparator rejected	Definition of comparator rejected, or definition not rejected	FJC	Appraisal documents
Endpoint assessment (1 variable per endpoint)	Assessment of endpoints morbidity, mortality, adverse events, QoL, 1 variable for each endpoint category: no statement about added benefit in documents, evidence not appraised, no data from manufacturer, dossier considered incomplete, added benefit, or no added benefit	IQWiG, FJC	IQWiG assessment (section assessment of added benefit at endpoint level) and FJC appraisal documents
Maximum possible budget impact			
Patient subgroup size	Number of patients in patient subgroup	Manufacturer, FJC	Dossier, part 1, Table 1-10 and appraisal documentation
Annual therapy cost	Total annual therapy cost in patient subgroup of new pharmaceutical in euros	Manufacturer, FJC	Dossier, part 1, Table 1-11 and appraisal documentation
Maximum possible budget impact	Patient subgroup size times annual therapy cost per patient in euros	Manufacturer, FJC	Dossier, part 1, Tables 1-10 and 1-11 and appraisal documentation

Note: ATC = Anatomical Therapeutic Chemical Classification System; EBA = early benefit assessment; FJC = Federal Joint Committee; ICD-10-GM = International Classification of Diseases Version 10, German Modification; IQWiG = Institute of Quality and Efficiency in Health Care; QoL = quality of life; RCT = randomized controlled trial.

procedures. Examples of differences include the timing of assessment and appraisal within the product life cycle of a pharmaceutical, the criteria used for appraisal to determine whether there is an added benefit, the method for selecting substances to be assessed, and the effect of these factors on subsequent price setting. Other unusual aspects of Germany's approach are the volume and the type of information that are used and made publicly available at stages 1 and 2. In other settings, only information on the final appraisal and the assessment report are published.[12] AMNOG therefore offers the opportunity to analyze differences in the roles of institutions that are involved in the separate stages of assessment and appraisal but draw conclusions using the same data.

There has been much speculation about the impact of the AMNOG on the budgets of statutory health insurers and the availability of new pharmaceuticals in Germany.[10] Manufacturers, who previously were able to set their prices freely and were ensured coverage of their products through statutory health insurance, have reported a high degree of process uncertainty because the processes of assessment and appraisal are allocated to 2 separate institutions.[4] Manufacturers have also argued that they are generally unable to provide the required level of evidence at the time of dossier submission.[13] Last, some manufacturers have voiced concerns that the FJC may be exceeding its statutory authority and taking account of nonclinical evidence, such as the budget impact of a pharmaceutical, in its rating decisions.[13] Gaining a clearer understanding of the associations between the components of the EBA process and the outcomes of this process could help explain the rating decisions made to date by the FJC. It could also clarify the roles of the participating parties, guiding manufacturers in developing future submissions and highlighting areas where the decision-making process could be improved. To meet these informational needs, our approach was twofold: 1) We evaluated differences in the rating decisions made by manufacturers, the IQWiG, and the FJC with regard to each pharmaceutical's added benefit. 2) We identified whether the FJC's rating decisions had been influenced by information submitted in the dossiers, including information on individual endpoints, the characteristics of the EBA process, and the budget impact of each pharmaceutical. Our analysis is based on data extracted from EBAs for which the FJC made a rating decision between 2011 and 2013.

METHODS

Data Sampling and Extraction

We first reviewed all 73 EBAs listed on the website of the FJC[14] as of 15 June 2013 (see Appendix 2 in the online supplementary material for this article). The documentation for an EBA is structured in a standardized fashion and, if complete, consists of the manufacturer's dossier, IQWiG's assessment of the evidence in the dossier, the documentation of the official hearing, the decision made by the FJC in its rating of a given pharmaceutical's added benefit, and the FJC's rationale for this decision. We excluded EBAs from our analysis if the dossier was not available or the EBA process was ongoing.

To extract data, we developed a list of variables that represented the essential components of the EBA process, as defined previously in the literature.[15,16] These components are as follows: the extent of a pharmaceutical's added benefit as rated separately by the manufacturer, IQWiG, and the FJC; the characteristics of the pharmaceutical being evaluated; the duration and other characteristics of the EBA process; the types of evidence submitted by the manufacturer; the methods used to generate evidence; and the budget impact of the pharmaceutical.

Table 1 provides operational definitions for each variable. Whenever possible, we based our English translations on the terms used in the relevant legal statutes.[17–19] Two independent reviewers extracted the data using the list of variables and a worksheet template. Once completed, the worksheets were compared to identify any deviations. Interrater reliability was good, with an average Cohen's kappa coefficient of 0.63 (range, 0.28 to 1.00) for categorical variables and an average Pearson's correlation coefficient of 0.80 for continuous variables (range, –0.18 to 1.00). Any disagreement was resolved through discussion between the authors.

A pharmaceutical's added benefit as rated by the different parties. When one is looking at the way in which the manufacturer, IQWiG, and the FJC rated the added benefit of any given pharmaceutical, it is important to distinguish between 4 different levels: 1) the substance level (i.e., the active ingredient or combination of active ingredients, which is used in this paper synonymously with the term *pharmaceutical*); 2) the indications for which the substance has been licensed; 3) for each indication, the patient subgroups in which the substance may be used; and 4) for each patient subgroup, the endpoints chosen

by the manufacturer to demonstrate the added benefit of a pharmaceutical in that subgroup.[§]

If a pharmaceutical has been licensed for use in multiple indications, the FJC rates the extent of the pharmaceutical's added benefit separately for each indication (Level 2). The manufacturer must submit the appropriate documentation demonstrating an added benefit for each indication separately. This evidence makes up what is known as Part 4 of the manufacturer's dossier. In the case of a pharmaceutical with 2 indications, for example, the manufacturer must submit 2 subsections in Part 4: 4A and 4B. For the purposes of our analysis, we defined each subsection as an EBA in its own right. At this level of decision making, the FJC rates the added benefit of a pharmaceutical in a binary manner: ''added benefit'' or ''no added benefit.'' If a pharmaceutical is given a positive rating by the FJC for any of the indications for which it has been licensed, then the pharmaceutical as a whole (i.e., at the substance level; Level 1) is given a positive rating (also binary: added benefit or no added benefit) and price negotiations can take place.

Of course, for each indication, there may be multiple patient subgroups (Level 3). For each subgroup (Level 3), IQWiG and the FJC rate the extent of a pharmaceutical's added benefit. In doing so, they distinguish between 6 categories when making their rating decisions: major added benefit, considerable added benefit, minor added benefit, nonquantifiable added benefit, no evidence of added benefit, and less benefit than the appropriate comparator (see Appendix 1 for detailed description). If, for any patient subgroup within a given indication, the FJC rates the added benefit of a pharmaceutical using any of the first 4 of these categories, then the FJC gives a positive binary rating (i.e., added benefit as opposed to no added benefit) for that indication as a whole. As noted above, a positive rating for an indication means that the pharmaceutical as a whole is given a positive rating and price negotiations can take place. In our paper, we use rating decisions made by the different parties at the patient subgroup level as our unit of analysis. We do this because the more granular decisions made at the patient subgroup level are taken into account by the Federal Association of Statutory Health Insurers in subsequent price negotiations to settle on a price at the substance level.

Finally, in any given patient subgroup within a given indication, the manufacturer chooses 1 or more endpoints to demonstrate the added benefit of a pharmaceutical in that subgroup (Level 4). These endpoints are morbidity (i.e., the medical conditions or complications of a disease), mortality, adverse events, quality of life, or some combination of these. At this level of observation, the decision-making criteria used by IQWiG and the FJC are not explicit. In some cases, the manufacturer provides data on an endpoint, but neither IQWiG nor the FJC will make any statement on this particular endpoint. In other cases, the evidence for a certain endpoint is discussed by IQWiG or the FJC, but the added benefit of a pharmaceutical for this endpoint is not rated by either party. In yet other cases, the added benefit of a pharmaceutical for each endpoint is rated according to the 6 categories described above. In short, there are no explicit rules for translating ratings of added benefit for individual endpoints to ratings of added benefit for the corresponding patient subgroups.

When exploring differences in the rating decisions made by the different parties, we had to consider that the FJC does not always conduct its appraisal using the same number or type of subgroups as those given in the manufacturer's dossier or the assessment provided by IQWiG. To ensure comparability between the rating decisions, we chose the perspective of the FJC, as it is this institution whose decision is binding. An example can be seen in the case of telaprevir, a protease inhibitor for the treatment of hepatitis C. In its dossier, the manufacturer of telaprevir defined 5 patient subgroups; in its subsequent assessment, IQWiG defined 8 patient subgroups; and in its appraisal, the FJC defined 2 subgroups and rated the extent of telaprevir's added benefit for these 2 subgroups only. To compare the rating decisions made by the 3 different parties in this and similar cases, we imputed a median rating of benefit for any patient subgroup that had been defined by the FJC but had not been explicitly defined by the manufacturer or by IQWiG.

Characteristics of the pharmaceutical. Several studies have analyzed the impact of a pharmaceutical's characteristics on coverage decision making, particularly for cancer treatment or orphan drugs.[20–22] Their findings suggest that coverage decisions can vary by the chemical characteristics of the active ingredient as well as by the indications for which a pharmaceutical has been approved. For each EBA, we used the German modification of the International Classification of Diseases, Version 10,

[§]The levels described here are used for explanatory purposes and are not used explicitly by the FJC or any other party in the early benefit assessment process.

German Modification (ICD-10-GM) to record the indication for which an EBA was being undertaken,[23] the therapeutic main group (i.e., second level) of the Anatomical Therapeutic Chemical (ATC) classification system,[24] and whether the pharmaceutical had orphan drug status.

Characteristics of the early benefit assessment process. To capture variation in the EBA process, we recorded information on the duration of each EBA, whether the rating decision by the FJC was preliminary, and the year in which the FJC's decision was published. To account for stakeholder involvement, we captured the number of individuals participating in the official hearing, the number of written statements submitted during the official hearing, and whether the manufacturer had requested a consultation with the FJC, IQWiG, or both institutions prior to submitting a dossier. The aim of providing consultations to manufacturers is to support them in choosing the appropriate documentation and comparator treatment.

Methods used to generate evidence on a pharmaceutical's added benefit. To describe the methods used to generate evidence on added benefit, we used variables that captured which type of studies had been presented by the manufacturer in its dossier, the number of studies, and the size of patient cohorts. We also captured whether indirect comparisons had been presented or sensitivity analyzes had been used.

Type of evidence submitted. This variable represents the type of evidence submitted by the manufacturer to demonstrate added benefit. We extracted data on the endpoints used by the manufacturer and on whether IQWiG and the FJC had found an added benefit for each endpoint. These endpoints were mortality, morbidity, adverse events, quality of life, or some combination of these. For some patient subgroups, we were unable to identify any explicit statements on added benefit by IQWiG or the FJC. For other patient subgroups, IQWiG, the FJC, or both institutions had assessed the evidence but had not considered it sufficient, or they had evaluated the evidence but made no explicit decision.

When exploring the types of evidence submitted, we also had to consider that the FJC can reject the manufacturer's choice of patient subgroups or comparator treatment. We therefore captured such rejections at the patient subgroup level to analyze associations with the rating decisions of the FJC.

Maximum possible budget impact. Although AMNOG does not require a full cost-effectiveness, cost-utility, or budget-impact analysis, manufacturers must nevertheless submit information on annual treatment costs per patient and the size of patient subgroups. To calculate a proxy for the maximum possible budget impact, we multiplied the number of patients in each patient subgroup by the annual cost per patient in that subgroup. If the size of patient subgroups or the cost of standard treatment was given in intervals, we calculated mean values.

Statistical Analyses

First, we undertook descriptive analysis of the rating decisions made by the different parties with regard to a pharmaceutical's added benefit. To begin, we determined the status of each of the EBAs listed on the website of the FJC as of 15 June 2013 (e.g., whether it was finished or still in the appraisal stage), as well as its length and the characteristics of the pharmaceutical being evaluated. We subsequently excluded EBAs that were ongoing, had been terminated, or involved pharmaceuticals that were ultimately exempted from the process (e.g., because they were found to have a negligible impact on the annual expenditure of statutory health insurers). Using data from the remaining EBAs, we analyzed agreement between the rating decisions made by the FJC and IQWiG and between the rating decisions made by the FJC and the manufacturers. In doing so, we used the decisions made at the level of patient subgroups, because, as described above, a positive rating by the FJC for any patient subgroup leads to a positive rating for the pharmaceutical as a whole.

To quantify agreement between the parties, we calculated a weighted Cohen's kappa, which ranges from –1 to 1 and is a measure used to determine whether agreement among raters is by chance.[25] A value less than or close to 0 indicates that agreement is due to chance, whereas a value of 1 indicates perfect agreement. Values higher than 0.4 indicate moderate agreement. The weighted kappa is used for ordinal variables with more than 2 categories. In the present analysis, we weighted each of the 6 categories of added benefit equally. When measuring agreement between the FJC and IQWiG, we had to exclude a total of 9 patient subgroups that had been defined by the FJC but for which there was no equivalent in IQWiG's assessment. The patient subgroups all belonged to substances that are orphan drugs for which IQWiG typically does not provide an assessment of benefit.

Second, within each patient subgroup, we examined how the evidence presented in the manufacturer's dossier on various endpoints had been evaluated by IQWiG and the FJC. For this part of the analysis, we considered only those patient subgroups for which manufacturers had presented evidence on specific endpoints. For each endpoint category, we calculated how frequently a binary rating of ''added benefit'' or ''no added benefit'' had been stated explicitly. If we were unable to identify a rating of added benefit in the documents, we classified the statements made by IQWiG and/or the FJC about the evidence.

Third, we used bivariate analysis to quantify the impact of the variables from our data extraction form on the rating decisions made by the FJC at the patient subgroup level. Although we were looking at decisions made by the FJC at the patient subgroup level, we classified these decisions in a binary manner rather than using the 6 categories of added benefit described above.

To account for nonindependent observations, test statistics were adjusted for clustering (i.e., multiple patient subgroups within a substance). For categorical data, we used the Rao-Scott chi-square test, a design-adjusted version of the Pearson chi-square test that accounts for clustering by substance.[26] For continuous data, we used t statistics from bivariate regressions with the rating decision of the FJC as independent variable. We accounted for clustering by substance using generalized least squares estimation techniques, adjusting the variance-covariance matrix for complex sample designs by applying Taylor expansion theory to estimate sampling errors of estimators.[27] Significance was defined at $P < 0.05$. Data preparation and statistical analysis were performed using SAS version 9.3.

RESULTS

In total, 73 EBAs had been initiated by 15 June 2013. These represented 73 indications for which 59 substances had been licensed for use. A total of 33 manufacturers had submitted dossiers for these 59 substances. Figure 1 and Appendix 1 provide a breakdown of the EBAs according to their status, substance, and indication.

For the 59 substances considered by the FJC during the study period, the most common indications according to ICD-10-GM were endocrine, nutritional, and metabolic diseases (16 substances); neoplasms (16 substances); diseases of the circulatory system (6 substances); and certain infectious and parasitic diseases (6 substances). A total of 10 substances were orphan drugs. For these same 59 substances, the most common groups according to the ATC system were antineoplastic agents (14 substances), drugs used in diabetes (10 substances), antivirals for systemic use (4 substances), and immunosuppressants (4 substances).

Of the 73 EBAs that had been initiated by 15 June 2013, we excluded a total of 31 (representing 23 substances) from subsequent analysis: 21 were ongoing and therefore no final decision had been made by the FJC; 3 had been concluded with exemptions from the EBA process as the substances in question were found by IQWiG or the FJC to have a negligible budget impact on statutory health insurers; 6 lacked a manufacturer's dossier but had nevertheless been concluded with a (negative) rating decision by the FJC; and 1 involved a substance that had been listed on the website of FJC without a status or an ending date. Of the 9 EBAs that were excluded because they lacked a manufacturer's dossier or were exempted from the EBA process, 2 were concluded in 2011, 6 in 2012, and 1 during the first half of 2013.

For our subsequent analyses, we used the 42 remaining EBAs (representing 36 substances, 68 patient subgroups) (Figure 1). Among these 42 EBAs, the average duration of the EBA process was 211 days. One of these EBAs was concluded in 2011, 23 were concluded in 2012, and 18 were concluded in the first half of 2013. The FJC found evidence of benefit in at least 1 patient subgroup of 24 substances that it appraised in 25 EBAs.

Agreement between the FJC and Manufacturers, and Between the FJC and IQWiG, in the Rating of Added Benefit

Within 42 EBAs, the FJC appraised a total of 68 patient subgroups (as counted according to the FJC definition of subgroups). Of these subgroups, 42 (61.8%) were identical to those defined by the manufacturer and 49 (72.1%) were identical to those defined by IQWiG. Tables 2 and 3 compare the rating decisions made by the FJC to those made by the manufacturer and IQWiG at the patient subgroup level.

In general, the FJC's rating was lower than that of the manufacturer. Indeed, both parties agreed in their rating for only 5 of the 68 patient subgroups. In most cases (36 subgroups), the FJC decided that there was no evidence of added benefit. Although the highest rating (i.e., major added benefit) was used by manufacturers for 36 subgroups (52.9%), it was never

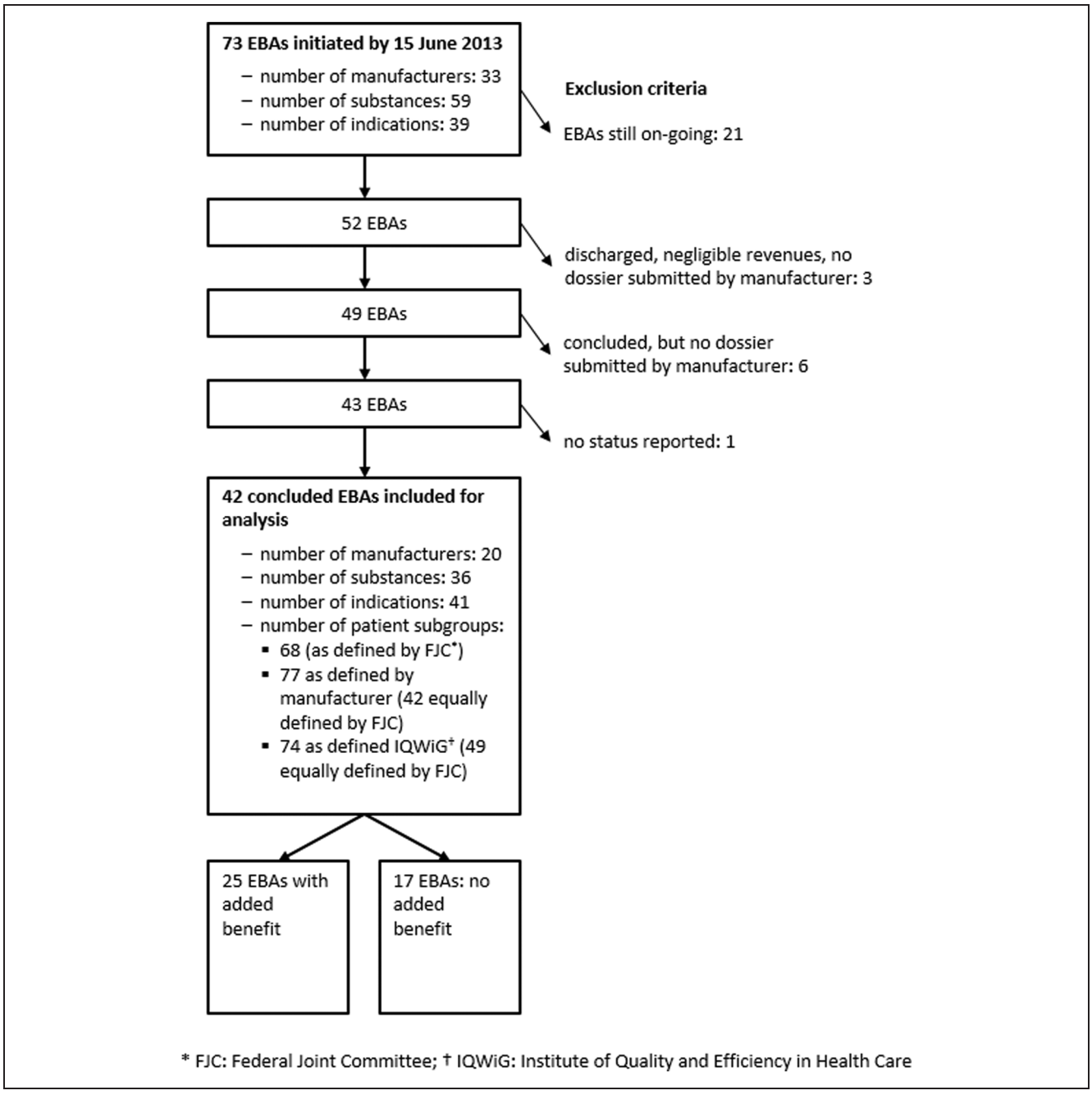

Figure 1 Status of all early benefit assessments (EBAs) initiated by 15 June 2013.

used by the FJC. The weighted kappa coefficient was 0.21 (95% confidence interval 0.107–0.31), which indicates low agreement between the 2 parties.

Agreement between the FJC and IQWiG was much higher, with a kappa coefficient of 0.64 (95% confidence interval 0.451–0.827) highlighting a substantial degree of agreement on the 59 patient subgroups rated by both institutions. In most cases (36 subgroups), both institutions decided that there was no evidence of added benefit. While the FJC's rating was generally lower than that of the manufacturer, that of IQWiG was even lower: In 9 subgroups (15.3%), the rating

Table 2 Decisions on Added Benefit by the Federal Joint Committee (FJC) and Manufacturers at Patient Subgroup Level

FJC Decision Manufacturer Decision	Less Benefit Than Comparator	No Added Benefit	Not Quantifiable Added Benefit	Minor Added Benefit	Considerable Added Benefit	Major Added Benefit	Sum
Less benefit than comparator	0	0	0	0	0	0	0
No added benefit	0	4	0	0	0	0	4
Not quantifiable added benefit	0	5	0	0	0	0	5
Minor added benefit	0	7	0	1	0	0	8
Considerable added benefit	0	7	1	7	0	0	15
Major added benefit	1	13	7	8	7	0	36
Sum	1	36	8	16	7	0	68

Table 3 Decisions on Added Benefit by the Federal Joint Committee (FJC) and the Institute for Quality and Efficiency in Health Care (IQWiG) at Patient Subgroup Level

FJC Decision IQWiG Decision	Less Benefit Than Comparator	No Added Benefit	Not Quantifiable Added Benefit	Minor Added Benefit	Considerable Added Benefit	Major Added Benefit	Sum
Less benefit than comparator	0	0	0	0	0	0	0
No added benefit	1	36	1	5	2	0	45
Not quantifiable added benefit	0	0	4	1	0	0	5
Minor added benefit	0	0	0	3	0	0	3
Considerable added benefit	0	0	0	2	4	0	6
Major added benefit	0	0	0	0	0	0	0
Sum	1	36	5	11	6	0	59[a]

a. Only patient subgroups rated by IQWiG;

given by the FJC was higher than that given by IQWiG. Only 3 subgroups (5.1%) were rated higher by IQWiG than by the FJC.

Evaluation of Endpoints by IQWiG and the FJC

Manufacturers presented evidence on different endpoint in their dossiers. Most frequently, they presented evidence on morbidity (89.7% of patient subgroups), followed by adverse events (85.3%), quality of life (72.1%), and mortality (70.6%).

In IQWiG's assessment and the FJC's appraisal of evidence on different endpoints, we identified 5 types of decisions: 1) the evidence was disregarded by IQWiG and/or the FJC, and no statement was made about added benefit; 2) the evidence was evaluated by IQWiG and/or the FJC but was considered insufficient; 3) the evidence was evaluated by IQWiG and/or the FJC and considered sufficient, but no decision on added benefit was reported; 4) the evidence was assessed by IQWiG and appraised by the FJC, but no evidence of added benefit was found by either party; and 5) the evidence was assessed by IQWiG and appraised by the FJC, and evidence for benefit was found by both parties (see Table 4).

Regarding the first type of decision, we observed differences between IQWiG and the FJC. IQWiG was more assiduous in its assessment of endpoints, failing to make a statement about added benefit for endpoints in only 1 patient subgroup (telaprevir in hepatitis C patients). The FJC, however, was much more likely than IQWiG to make no statement on added benefit at the endpoint level, particularly in the case of the endpoints mortality (16/48 patient subgroups in which mortality had been included as an endpoint) and quality of life (9/61 patient subgroups in which quality of life had been included as an endpoint).

Regarding the second type of decision, we observed that this type of decision was made most frequently in the case of patient subgroups that included information on adverse events (IQWiG, 34/56 patient subgroups; FJC, 25/58 patient subgroups) and least frequently in the case of patient subgroups that included information on mortality (IQWiG, 21/58 patient subgroups; FJC, 14/48 patient subgroups).

Table 4 Type of Evidence Submitted and Its Evaluation by the Federal Joint Committee (FJC) and Institute of Quality and Efficiency in Health Care (IQWiG) at Patient Subgroup Level

	Rating of Endpoint Categories by IQWiG						Rating of Endpoint Categories by FJC					
Endpoint Assessment	Patient Subgroups for Which Manufacturer Provided Data	No Statement about Added Benefit in Documents	Evidence Assessed, but Data Not Considered Sufficient	Evidence Assessed, but No Recommendation Made	No Added Benefit	Added Benefit	Patient Subgroups for Which Manufacturer Provided Data	No Statement about Added Benefit in Documents	Evidence Assessed, but Data Not Considered Sufficient	Evidence Assessed, but No Decision Taken	No Added Benefit	Added Benefit
Morbidity	57	1 (1.8%)	33 (57.9%)	4 (7%)	11 (19.3%)	8 (14%)	61	9 (14.8%)	25 (41%)	7 (11.5%)	5 (8.2%)	15 (24.6%)
Mortality	44	1 (2.3%)	21 (47.7%)	4 (9.1%)	12 (27.3%)	6 (13.6%)	48	16 (33.3%)	14 (29.2%)	4 (8.3%)	7 (14.6%)	7 (14.6%)
Adverse events	56	1 (1.8%)	34 (60.7%)	5 (8.9%)	7 (12.5%)	9 (16.1%)	58	9 (15.5%)	25 (43.1%)	7 (12.1%)	10 (17.2%)	7 (12.1%)
Quality of life	45	1 (2.2%)	26 (57.8%)	4 (8.9%)	13 (28.9%)	1 (2.2%)	49	12 (24.5%)	19 (38.8%)	6 (12.2%)	10 (20.4%)	2 (4.1%)

Note: Values are *n* (%).

Regarding the third type of decision, we found that, again, IQWiG appeared to assess endpoints more assiduously than did the FJC. Whereas this type of decision was made by IQWiG for 4 patient subgroups, it was made by the FJC for 6 patient subgroups on the average across all endpoint categories.

For the remaining types of decisions (i.e., types 4 and 5), we found similar patterns of decision making between IQWiG and the FJC. Evidence of added benefit was found by both institutions in roughly the same proportion of patient subgroups that included the endpoints of mortality (IQWiG, 6/44; FJC, 7/48), adverse events (IQWiG, 9/56; FJC, 7/58), and quality of life (IQWiG, 1/45; FJC, 2/49). For morbidity, however, IQWiG appears to have been more strict than the FJC in finding evidence of added benefit in patient subgroups that included this endpoint (IQWiG, 8/57; FJC, 15/61).

Impact of the Different Characteristics of the Early Benefit Assessment Process on the Rating Decisions Made by the FJC

The association between the type of evidence submitted by the manufacturer and the rating decisions made by the FJC at the patient subgroup level (i.e., added benefit v. no added benefit) was significant for subgroups that included evidence on morbidity ($P = 0.009$) and adverse events ($P < 0.001$) but not for subgroups that included evidence on mortality ($P = 0.718$) or quality of life ($P = 0.783$). The rejection of a subgroup by the FJC was not associated with the rating decision of the FJC ($P = 0.506$), whereas the rejection of the comparator was ($P = 0.003$; see Table 5).

Some of the methods used to generate evidence were associated with the rating decision made by the FJC at the patient subgroup level. Manufacturers had used a systematic literature review to generate evidence in 78.4% of the subgroups for which the FJC ultimately made a negative rating decision compared with 45.2% of the subgroups for which the FJC ultimately made a positive rating decision ($P = 0.005$). Similarly, manufacturers had used indirect comparisons to generate evidence in 56.8% of the subgroups for which the FJC ultimately made a negative rating decision compared with 29.0% of the subgroups for which the FJC ultimately made a positive rating decision ($P = 0.019$). Finally, manufacturers had used sensitivity analyses to generate evidence in 83.3% of the subgroups for which the FJC ultimately made a negative rating decision compared with 58.1% of the subgroups for which the FJC ultimately made a positive rating decision ($P = 0.022$).

The various characteristics of the EBA process did not appear to have an impact on the rating decision made by the FJC at the patient subgroup level. Indeed, no clear pattern of association was observed for the average duration of the EBA process ($P = 0.486$), the number of participants in the official hearing ($P = 0.227$), the number of written statements submitted to the FJC ($P = 0.170$), the frequency of preliminary rating decisions ($P = 0.149$), or the use by manufacturers of consultations ($P = 0.149$).

When looking at variables related to the cost of a pharmaceutical, we found that patient subgroup sizes (mean of the differences, 8761; $P = 0.024$), annual treatment cost per patient (€6758; $P = 0.001$) and maximum possible budget impact (€85,749,209; $P = 0.002$) were significantly higher in the FJC's documentation than in the manufacturers' dossiers. In the FJC's documentation, the annual treatment costs per patient were highest for the orphan drug ivacaftor (€289,060) and lowest for apixaban (€85). The budget impact was highest for aflibercept (€3,441,854,850) and lowest for axitinib (€571,533).

There was a significant association between the rating decisions of the FJC and the annual treatment cost per patient (manufacturer dossier, $P = 0.003$; FJC's documentation, $P = 0.006$). Pharmaceuticals with a higher annual treatment cost per patient in either the manufacturer's dossier or the FJC's documentation were more likely to be rated by the FJC as having an added benefit. When maximum possible budget impacts (Figure 2) were graphed against the extent of benefit, no clear pattern could be identified (manufacturer dossier, $P = 0.347$; FJC's documentation, $P = 0.725$).

DISCUSSION

We analyzed 42 EBAs concluded in Germany between 1 January 2011 and 15 June 2013. These corresponded to 36 substances and 68 patient subgroups. By clarifying the roles of the parties in the EBA process, our findings help explain the decisions made by the FJC, highlight where the decision-making process can be improved, and can guide manufacturers in the development of future submissions.

The findings of our analysis of agreement between the different parties suggest that IQWiG and the FJC may follow an unarticulated strategy. Whereas both institutions frequently rated the added benefit of pharmaceuticals much lower than did manufacturers, IQWiG appears generally to have taken the hard line. From the manufacturer's perspective, it might be speculated that the role of IQWiG is to take a very strict approach based purely on the submitted evidence, whereas the role of the FJC is to step in and soften the conclusion reached by IQWiG. The aim of this strategy could be to make the final decision more palatable to manufacturers. It should be noted in this context that the FJC's decision-making criteria remain vague, with the only explicit criteria being health benefits and severity of disease.[18] Another explanation for the divergence in rating decisions between IQWiG and the Federal Joint Committee might be found in the additional information made available to the FJC during the official hearing.

Our findings regarding the type of evidence submitted at the endpoint level suggest that manufacturers have encountered problems in gaining acceptance of their evidence base by IQWiG and the FJC, particularly in the case of quality-of-life data. Although it is difficult to do so at such an early stage in the product life cycle, manufacturers should strive to provide comprehensive data on quality-of-life outcomes, as these are also increasingly considered to be of relevance by social health insurers in Germany.[28] For morbidity measures, there would seem to be a clear opportunity for manufacturers to intervene by submitting statements or participating in the official hearing, as IQWIG appears to take a particularly strict approach to evaluating these measures.

Our findings also indicate that the rejection by IQWiG or the FJC of the manufacturer's choice of comparator is associated with a negative rating decision by the FJC. To avoid having the entire evidence base for what might be a truly beneficial pharmaceutical rejected outright by either institution, manufacturers should take greater care when choosing comparators. This issue may soon diminish in importance, however. Following a change in legislation in June 2013, manufacturers now have greater leeway in their choice of comparators.[29]

Divergent views on how patient subgroups should be defined may lead the different parties occasionally to choose different sizes for patient subgroups and assign different costs to these subgroups. The fact that the size of the patient subgroups and the annual treatment costs were generally lower in the manufacturers' dossiers than in the documentation produced by the FJC suggests that manufacturers have submitted lower values to reduce the risk of a negative rating decision by the FJC. In contrast, the FJC appears to err on the side of higher cost data estimates as a way to be more cautious about the true cost for social health insurers.

Table 5 Different Characteristics of the Early Benefit Assessment Process and the Rating Decisions Made by the Federal Joint Committee (FJC)

	All Patient Subgroups (n = 68)	No Added Benefit (n = 37 Patient Subgroups)	Added Benefit (n = 31 Patient Subgroups)	P Value
Type of evidence submitted				
Endpoint morbidity presented	89.7% (61)	97.3% (36)	80.7% (25)	0.009[a]
Endpoint mortality presented	70.6% (48)	73% (27)	67.7% (21)	0.718[a]
Endpoint adverse events presented	85.3% (58)	97.3% (36)	71% (22)	<0.001[a]
Endpoint quality of life presented	72.1% (49)	70.3% (26)	74.2% (23)	0.783[a]
Patient subgroup rejected	26.5% (18)	29.7% (11)	22.6% (7)	0.506[a]
Comparator rejected	30.9% (21)	46% (17)	12.9% (4)	0.003[a]
Methods for evidence generation				
Systematic literature search presented	63.2% (43)	78.4% (29)	45.2% (14)	0.005[a]
Indirect comparison presented	44.1% (30)	56.8% (21)	29% (9)	0.019[a]
Meta-analysis presented	47.1% (32)	54.1% (20)	38.7% (12)	0.207[a]
Sensitivity analysis presented	72.1% (49)	83.8% (31)	58.1% (18)	0.022[a]
Registry studies (number)	3.2 (1; 3)	3.2 (2; 3)	3.2 (1; 3)	0.991[b]
Randomized controlled trials (number)	2.3 (1; 3)	2.5 (1; 3)	2.1 (1; 3)	0.502[b]
Head-to-head trials (number)[c]	0.7 (0; 1)	0.6 (0; 1)	0.9 (0; 1)	0.425[b]
Other studies (nonrandomized, etc.) (number)	0 (0; 0)	0 (0; 0)	0 (0; 0)	—
Patients in intervention arms per study (number)[c]	521 (173; 518.8)	600.1 (173; 573.3)	412.6 (173; 425)	0.183[b]
Patients in intervention arms per study (normalized to patient subgroup size)[c]	2 (0; 0.9)	1.4 (0; 0.5)	2.8 (0.1; 0.9)	0.516[b]
Process				
Duration of decision process (days)[d]	178 (171; 176)	173 (171; 173)	182 (168; 180)	0.357[b]
Hearing participants (number)	15 (10; 20)	14 (9; 17)	16 (10; 20)	0.227[b]
Written statements (number)[c]	10 (6; 13)	9 (6; 13)	11 (9; 12)	0.170[b]
Preliminary decision	25% (17)	18.9% (7)	32.3% (10)	0.149[a]
Consultation requested[c]	83.3% (57)	89.2% (33)	75.9% (24)	0.149[a]
Cost/maximum possible budget impact				
Patient subgroup size (manufacturer)	155,221 (305; 74,821)	258,199 (974; 139,000)	32,312 (140; 7044)	0.124[b]
Patient subgroup size (FJC)	164,435 (1220; 121,925)	269,768 (5500; 221,400)	38,714 (260; 12,000)	0.111[b]
Annual therapy cost per patient in patient subgroup (manufacturer)[c]	€37,293 (€761; €46,629)	€10,107 (€557; €6008)	€67,109 (€13,157; €77,435)	0.003[b]
Annual therapy cost per patient in patient subgroup (FJC)	€44,051 (€1092; €61,427)	€15,972 (€737; €11,285)	€77,565 (€18,141; €111,609)	0.006[b]
Maximum possible budget impact (manufacturer)[c]	€199,700,616 (€7,975,768; €164,256,386)	€259,258,221 (€5,214,306; €232,160,168)	€134,379,373 (€11,589,531; €164,256,386)	0.347[b]
Maximum possible budget impact (FJC)	€284,456,152 (€24,570,082; €285,661,843)	€310,163,912 (€28,029,823; €289,153,130)	€253,772,697 (€24,570,082; €282,170,556)	0.725[b]

Note: Values are given as % (n) of patient subgroups or as mean (1st quartile; 3rd quartile).
a. Rao-Scott chi-square test.
b. t test from bivariate regressions.
c. Variable has missing values.
d. Excluding submissions before 1 August 2011.

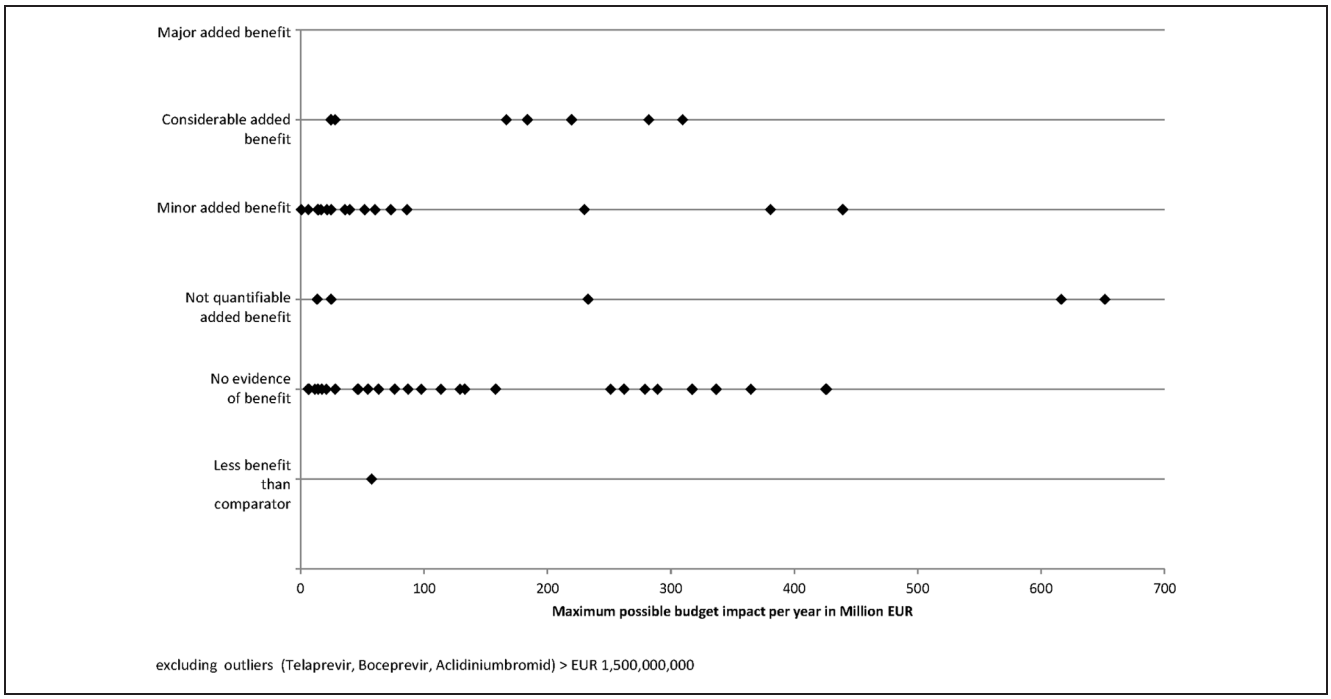

Figure 2 Maximum possible budget impact and rating of added benefit per patient subgroup (n = *68) by the Federal Joint Committee (FJC).*

Further, there was no pattern between rating decisions of the FJC and the maximum possible budget impact (as defined by the FJC). This finding suggests that clinical benefit is the main decision-making criterion used by the FJC. Concerns voiced by some manufacturers that the FJC might be exceeding its statutory authority and taking nonclinical evidence into account therefore appear to be unfounded. This being said, our measure of maximum budget impact is only a rough proxy for real budget impact, as it captures direct medical costs only (i.e., the costs of drugs, procedures, and diagnostics).[30] Moreover, the market penetration of a pharmaceutical is unlikely to reach the total indicated population.

Although it was possible for the FJC to commission cost-effectiveness analyses before AMNOG, this was not done on a regular basis. Studies on pharmaceutical regulation in Germany have focused primarily on the development of methods for assessing pharmaceuticals or have described individual decisions.[9,31] Kreis and Busse,[9] for example, describe the controversial decision of the FJC on oral antidiabetics (glinide class) in early 2011. Both IQWiG and the FJC concluded that the glinides should be excluded from reimbursement due to a lack of evidence on long-term benefits. The German Federal Ministry of Health challenged the decision, however, and demanded a clear justification from the FJC. During the discussion that accompanied the development and passage of the AMNOG, law makers found fault with the vagueness of the criteria being used by IQWiG and the FJC, particularly that of ''expediency.''[9] The documentation of the EBAs examined in this study is comprehensive and largely consistent, indicating that assessments of added benefit have become more structured since the AMNOG came into force. This is an encouraging development, as the lack of a systematic approach to such assessment and the vagueness of the decision-making criteria in Germany had long been a concern.[31]

The large amount of data now compiled and submitted to regulators shortly after a pharmaceutical has been launched represents a major advance in technology assessment in Germany, as does the consistent enforcement of the EBA process. Germany has moved away from a system consisting primarily of reference pricing applied to off-patent pharmaceuticals to one in which coverage decisions for patented drugs are made based on the extensive assessment and appraisal of clinical benefit.[32]

From a procedural perspective, the EBA process seems to be accountable to the various stakeholders, especially the manufacturers. As suggested by our data, process characteristics do not influence the

rating decisions made by the FJC at the patient subgroup level. Thus, the concerns voiced by manufacturers about process uncertainties shortly after ANMOG came into effect would seem unfounded.

The EBA process in Germany has been discussed internationally as a model of technology assessment, as it yields quick assessment and appraisal decisions soon after market entry while using data submitted by the manufacturer.[12,15,16] The present study is one of the first studies to analyze differences in the roles of institutions that are involved in the separate stages of assessment and appraisal but draw conclusions using the same evidence. As suggested by our findings, the main decision-making criterion of the FJC is clinical benefit. Internationally, the French Haute Autorité de Santé (HAS), which also appraises clinical benefit, is the institution most similar to the German FJC. One study has shown that all substances evaluated by HAS between 2005 and 2010 have entered subsequent price negotiations.[33] In countries that base their decisions on cost-effectiveness information, such as the UK[34] and Australia,[35] coverage decisions are frequently made for specific patient subgroups, particularly those that show favorable cost-effectiveness. Although IQWiG and the FJC evaluate patient subgroup sizes, we did not find any significant associations between patient subgroup size and the rating decisions made by the FJC at patient subgroup level.

Our results can guide manufacturers in developing future submissions, especially with regard to those characteristics of the EBA process that manufacturers themselves can influence. In terms of evidence generation, manufacturers should be aware that certain methods, such as systematic searches and indirect comparisons, have tended not to lead to a favorable rating decision made by FJC, at least not without triggering an audit. In many cases, the manufacturer provided evidence based on systematic searches or indirect comparisons for all patient subgroups, but the FJC tended to accept the evidence of added benefit only for a small proportion of these subgroups. In 7 EBAs with mixed results in the rating of added benefit by the FJC where systematic searches were presented, 12 subgroups were rated ''no added benefit'' whereas 9 subgroups were rated ''added benefit.''

Another issue that manufacturers may want to consider is the methods they use to generate evidence. For example, in 9 EBAs in which manufacturers used a systematic review to generate evidence, IQWiG rejected the manufacturer's choice of comparator (8 EBAs) or the study period (1 EBA) and would therefore not consider any of the submitted evidence by the manufacturer. Disease characteristics may be another reason why patient subgroups in which added benefit was found were less likely to have had systematic searches presented as evidence. For example, if a treatment has dramatically beneficial effects or no alternative treatment is available, less evidence may be needed to show an added benefit.

Our findings also highlight several areas in which the decision-making framework itself could be improved. While the EBA process and the related documentation are very transparent, the decision-making criteria are not. For example, the way in which the FJC translates quantitative study results into qualitative categories of added benefit at the patient subgroup level is not explicit. Similarly, the criteria used by the FJC at the endpoint level are not explicit, and the FJC's reporting of its conclusions at this level is inconsistent and lacking in detail.

Finally, it remains unclear which decision-making criteria are used by the FJC beyond clinical benefit. The findings of our analysis of agreement between IQWiG and the FJC suggest that other aspects do indeed play a role, as the FJC softened the conclusions drawn in IQWiG's assessments in the case of several patient subgroups that showed no evidence of added benefit (Table 4). Currently, it is unclear which criteria drive the differences between IQWiG and the FJC in their evaluation of the same substance.

Limitations

Our analysis has several important methodological limitations. Although the extent of a pharmaceutical's added benefit is determined at the patient subgroup level, we had to conduct our analyses from the perspective of one stakeholder, as definitions of patient subgroups can differ among the parties involved in the EBA process. We selected the FJC's perspective. Our imputation of data for the subgroups that had not been defined by the manufacturers (10 subgroups/5 substances) or IQWiG (2 subgroups/1 substance) may therefore have biased our findings. We believe, however, that this bias of unknown direction is small because, in our sample, IQWiG or the FJC most often split or merged subgroups. This means that the manufacturers' data and rating decision applies to the newly defined subgroups just as it did to the original ones.

In general, standardizing data using a structured list of variables can lead to bias, not least through the omission of information. This is particularly problematic if the goal of the research is to capture

variation in the presentation of evidence and the rating decisions made at the endpoint level. Although we analyzed whether IQWiG or the FJC had evaluated data at the endpoint level, we did not check whether the endpoint level itself was clinically relevant. Strictly speaking, we can therefore only claim to have measured the relationship between the evidence available at the patient subgroup level and the rating decisions made by the FJC at the patient subgroup level.

On a case-by-case basis, we aggregated decisions for each of the 4 endpoint categories (if there was more than 1 type of endpoint in a given patient subgroup: e.g., data on 2 adverse events). In such cases, we captured the highest rating of added benefit. As this definition may lead to bias toward a more favorable rating of endpoint categories, our findings may reflect an overly optimistic picture of rating decisions made by IQWiG or the FJC.

The relative youth of AMNOG means that we were unable to perform multivariate analysis. We nevertheless felt that it was important to conduct an early, first analysis of the different parties' rating decisions to help increase the transparency of the EBA process and potentially reduce the transaction costs of all stakeholders. Furthermore, our small sample size did not allow us to stratify decisions for statistical testing according to disease-related characteristics, such as indication or orphan drug status. Also, it would have been interesting to analyze the rating decision of the different parties according to disease severity or some measure of innovation. Finding objective measures of these, particularly for innovation, is challenging, however, and would have gone beyond the bounds of this early analysis.

Last, the initial rating decisions of IQWiG and the FJC will not be representative of the rating decision made by both institutions over the long run. Manufacturers will learn from the choices they made in presenting evidence in their initial submissions. Over time, they will adjust to the criteria used by IQWiG and the FJC. The time lag between the late-phase clinical trials used to generate evidence and market launch will always result in some uncertainty in terms of the treatment standards that are used as comparators.

CONCLUSIONS

The EBA process introduced in Germany as part of the AMNOG in January 2011 aims to reduce the expenditure of statutory health insurers and promote innovation in pharmaceutical development. It is a relatively young piece of legislation, however, and has led to concerns by manufacturers about process uncertainty. The findings of our study suggest that the 2 institutions involved in the assessment and appraisal stages of the EBA process may follow different roles based on an unarticulated strategy. In its rating of any given pharmaceutical's added benefit over a comparator treatment, IQWiG appears take a harder line than either the manufacturer or the FJC. The FJC subsequently takes a range of explicit and nonexplicit criteria into account to soften this decision, leading to an outcome that is presumably more acceptable to the manufacturer.

While the EBA process and related requirements for documentation are transparent, the decision-making criteria used by IQWiG and the FJC are not. The findings of our study suggest that from the manufacturer's perspective, the main challenges when submitting dossiers are to choose the appropriate evidence at the level of individual endpoints and to choose the appropriate comparator. In cases where manufacturers succeeded in meeting these 2 challenges, however, the FJC appears to have based its rating decisions on clinical evidence. Concerns voiced by some manufacturers that the FJC might be exceeding its statutory authority and taking nonclinical evidence into account appear to be unfounded. The fact that the majority of EBAs in our sample resulted in favorable rating decisions by the FJC and therefore entered the stage of price negotiations should strengthen confidence in this new technology assessment process over the long run.

Note

Presented at the 5th annual meeting of the German Health Economics Association, Essen, Germany, March 2013; and the 9th World Congress on Health Economics (iHEA), Sydney, Australia, July 2013.

ACKNOWLEDGMENTS

The authors thank Christine Carstens and Christian Mathes for their support with data abstraction.

REFERENCES

1. Stargardt T, Busse R, Dauben H-P. Germany - Pharma Profile - Pharmaceutical Pricing and Reimbursement Information. WHO Regional Office for Europe [Internet]. 2008 [cited 2012 Nov 19];

Available from: http://whocc.goeg.at/Literaturliste/Dokumente/CountryInformationReports/Germany_PPRI_2008.pdf

2. Huttin C. Experiences with reference pricing. In: Dukes M, Haaijer-Ruskamp F, de Joncheere C, Rietveld A, eds. Drugs and Money—Prices, Affordability and Cost Containment. Amsterdam: IOS Press; 2003. Chapter 8.

3. Gerber A, Stock S, Dintsios C-M. Reflections on the changing face of German pharmaceutical policy: how far is Germany from value-based pricing? Pharmacoeconomics. 2011;29(7):549–53.

4. Henschke C, Sundmacher L, Busse R. Structural changes in the German pharmaceutical market: price setting mechanisms based on the early benefit evaluation. Health Policy. 2013;109(3):263–9.

5. Schlette S, Hess R. Early benefit assessment for pharmaceuticals in Germany: lessons for policymakers. Issue Brief (Commonw Fund). 2013;29:1–9.

6. Federal Joint Committee. Chapter 5: Benefit assessment of pharmaceuticals according to s. 35a SGB V. Available from: URL: http://www.english.g-ba.de/downloads/17-98-3042/Chapter5-Rules-of-Procedure-G-BA.pdf

7. Velasco-Garrido M, Zentner A, Busse R. Health systems, health policy and health technology assessment. In: Velasco-Garrido M, Kristensen F, Nielsen C, Busse R, eds. Health Technology Assessment and Health Policy-Making in Europe— Current Status, Challenges and Potential. Copenhagen: World Health Organization; 2008. p 53–78.

8. General Methods for the Assessment of the Relation of Benefits to Costs. Cologne: Institute for Quality and Efficiency in Health Care; 2009.

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

10. Nink K, Hörn H, Pieper K, Böhler Y-B. Ein Jahr frühe Nutzenbewertung—Das IQWiG zieht Bilanz. GGW. 2012;12(4):18–24.

11. Morgan SG, McMahon M, Mitton C, et al. Centralized drug review processes in Australia, Canada, New Zealand, and the United Kingdom. Health Aff (Millwood). 2006;25(2):337–47.

12. Stafinski T, Menon D, Davis C, McCabe C. Role of centralized review processes for making reimbursement decisions on new health technologies in Europe. Clin Outcomes Res. 2011;3:117–86.

13. Götte D. Consequences of the German AMNOG for the identification of study objectives to demonstrate clinical efficacy and cost effectiveness of innovative drugs [in German]. Dtsch Med Wochenschr. 2012;137(6):274–80.

14. Gemeinsamer Bundesausschuss. Frühe Nutzenbewertung (nach § 35a SGB V). Available from: URL: http://www.g-ba.de/informationen/nutzenbewertung/

15. Fischer KE. A systematic review of coverage decision-making on health technologies—evidence from the real world. Health Policy. 2012;107(2–3):218–30.

16. Stafinski T, Menon D, Philippon DJ, McCabe CJ. Health technology funding decision-making processes around the world: the same, yet different. Pharmacoeconomics. 2011;29:475–95.

17. Federal Joint Committee. Verfahrensordnung des Gemeinsamen Bundesausschusses [code of procedure of the Federal Joint Committee] [Internet]. Jan 19, 2012. Available from: https://www.g-ba.de/informationen/richtlinien/42/

18. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach $ 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung - AM- NutzenV) [Internet]. 2010 p. 2324–9. Available from: http://www.gesetze-im-internet.de/bundesrecht/am-nutzenv/gesamt.pdf

19. § 35a SGB V: Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen [Internet]. Jan 1, 2011. Available from: http://www.gesetze-im-internet.de/sgb_5/__35a.html

20. Mason AR, Drummond MF. Public funding of new cancer drugs: is NICE getting nastier? Eur J Cancer. 2009;45:1188–92.

21. Chim L, Kelly PJ, Salkeld G, Stockler MR. Are cancer drugs less likely to be recommended for listing by the Pharmaceutical Benefits Advisory Committee in Australia? Pharmacoeconomics. 2010;28:463–75.

22. Blankart CR, Stargardt T, Schreyögg J. Availability of and access to orphan drugs: an international comparison of pharmaceutical treatments for pulmonary arterial hypertension, Fabry disease, hereditary angioedema and chronic myeloid leukaemia. Pharmacoeconomics. 2011;29:63–82.

23. German Institute of Medical Documentation and Information. International Classification of Diseases-10, German Modification. Available from: URL: http://www.dimdi.de/static/de/klassi/icd-10-gm/kodesuche/onlinefassungen/htmlgm2013/index.htm

24. WHO Collaborating Centre for Drug Statistics Methodology. ATC/DDD Index 2014. Available from: URL: http://www.whocc.no/atc_ddd_index/

25. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46.

26. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. Biometrics. 1992;48(2):577–85.

27. Aparasu RR. Research methods for pharmaceutical practice and policy. London (UK), Chicago (IL): Pharmaceutical Press; 2011.

28. Perleth M, Gibis B, Göhlen B. A short history of health technology assessment in Germany. Int J Technol Assess Health Care. 2009;25:112–9.

29. Deutscher Bundestag (17. Wahlperiode). Beschlussempfehlung und Bericht des Ausschusses für Gesundheit (14. Ausschuss)/Drucksache 17/13770. Available from: URL: http://dip21.bundestag.de/dip21/btd/17/137/1713770.pdf

30. Mauskopf JA, Sullivan SD, Annemans L, et al. Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices—budget impact analysis. Value Health. 2007;10(5):336–47.

31. Fricke F-U, Dauben HP. Health technology assessment: a perspective from Germany. Value Health. 2009;12 Suppl 2:S20–7.

32. Drummond M, Jonsson B, Rutten F, Stargardt T. Reimbursement of pharmaceuticals: reference pricing versus health technology assessment. Eur J Health Econ. 2011;12:263–71.

33. Bending M, Hutton J, McGrath C. A comparison of pharmaceutical reimbursement agencies' processes and methods in France and Scotland. Int J Technol Assess Health Care. 2012;28:187–94.

34. O'Neill P, Devlin NJ. An analysis of NICE's ''restricted'' (or ''optimized'') decisions. Pharmacoeconomics. 2010;28(11):987–93.

35. Harris AH, Hill SR, Chin G, Li JJ, Walkom E. The role of value for money in public insurance coverage decisions for drugs in Australia: a retrospective analysis 1994–2004. Med Decis Making. 2008;28:713–22.