

Benefit assessment in Germany: requirements & challenges presented for 6 topics

German Translation: Nutzenbewertung in Deutschland: Anforderungen und Herausforderungen beispielhaft für 6 Themen

Authors (in alphabetical order):

- Sarah Böhme, Sarah.Boehme@pfizer.com, Pfizer Deutschland GmbH
- Astrid Genet, Astrid.Genet@pfizer.com, Pfizer Deutschland GmbH
- Johanna Gillhaus, Johanna.Gillhaus@pfizer.com, Pfizer Deutschland GmbH
- Niclas Kürschner, Niclas.Kuerschner@pfizer.com, Pfizer Deutschland GmbH
- Friedhelm Leverkus, Friedhelm.Leverkus@pfizer.com, Pfizer Deutschland GmbH
- Julia Schiffner-Rohe, Julia.Schiffner-Rohe@pfizer.com, Pfizer Deutschland GmbH
- Sarah Schmitter, Sarah.Schmitter@pfizer.com, Pfizer Deutschland GmbH
- Fabian Volz, Fabian.Volz@pfizer.com, Pfizer Deutschland GmbH

The authors contribute to different parts of the manuscript:

- 1) Introduction: Fabian Volz, Sarah Schmitter, Friedhelm Leverkus
- 2) Impact of label on benefit assessment: Sarah Schmitter
- 3) Health related quality of life: Niclas Kürschner
- 4) Validation of surrogates: Johanna Gillhaus
- 5) Handling of missing values: Julia Schiffner-Rohe, Astrid Genet
- 6) Indirect comparisons and Network Meta-Analysis: Julia Schiffner-Rohe, Sarah Böhme
- 7) Health economic evaluation: Fabian Volz

Corresponding Author:

Sarah Schmitter
Senior Manager Health Technology
Assessment & Outcomes Research
Health & Value Germany

Tel: + 49 30 550055-53404

Mobil: + 49 172 210 7224

Sarah.Schmitter@pfizer.com

Pfizer Deutschland GmbH

Linkstraße 10

10785 Berlin

Guest-Editor

Friedhelm Leverkus
Director Health Technology Assessment &
Outcomes Research
Health & Value Germany

Tel: +49 30 550055-52357

Fax: +49 30 550054-52459

Friedhelm.Leverkus@pfizer.com

Pfizer Deutschland GmbH

Linkstraße 10

10785 Berlin

Introduction

Authors: Fabian Volz, Sarah Schmitter, Friedhelm Leverkus

In 2011 the AMNOG (Act on the Reform of the Market for Medicinal Products, Arzneimittelmarktneuordnungsgesetz) became effective and since then the pharmaceutical companies have to demonstrate an added benefit of a new pharmaceutical, a new combination or a new label versus the established therapies. The added benefit is the basis for price negotiations with the National Association of the Statutory Health Insurance Funds (Spitzenverband der Gesetzlichen Krankenkassen, GKV-SV). The central element is the benefit dossier, which has to be compiled and submitted by the pharmaceutical company. In this benefit dossier the pharmaceutical company has to demonstrate the added benefit of the new pharmaceutical versus the so called “zweckmäßige Vergleichstherapie” (ZVT, appropriate comparative therapy), which could be different to those in the pivotal studies. The “Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen” (IQWiG, Institute for Quality and Efficiency in Health Care) or the “Gemeinsamer Bundesausschuss” (G-BA, Federal Joint Committee) conducts the benefit assessment based on the submitted dossier and the G-BA decides afterwards on the added benefit. The criteria of the benefit assessment are stated in the general method paper of IQWiG, but are often not self-evidently. For that reason this series of articles are supposed to give an understanding of the complexity within the scope of preparation a benefit dossier and to discuss selected specific issues. In addition we would like to give deeper insights in the requirements of G-BA and IQWiG for our colleagues in the global clinical research departments. These requirements differ in parts from the requirements of the regulatory bodies, but are reasonable.

Basis for the articles are the rules of procedure by G-BA, the methods paper by IQWiG and the analyses of former benefit assessments. In the focus are statistical analyses of clinical trials, but also other relevant requirements and challenges, which are demanded by G-BA and IQWiG. The following topics will be discussed in separate papers:

1. Impact of label on benefit assessment

The paper discusses a twofold relevance of the wording of the label has on the outcome of the benefit assessment. The wording is relevant for the determination of the ZVT as well as for the consideration of available studies. Ideally the wording of the label totally corresponds to the inclusion criteria of the pivotal study and patients excluded from the study are also excluded by the wording of the label.

2. Health related quality of life

The paper describes the concept of health related quality of life including the point of view from regulatory bodies. Furthermore, the methodological requirements of health related quality of life data analyses are discussed to allow future acceptance of IQWiG and G-BA within the early benefit assessment.

3. Validation of surrogates

This paper gives an overview on the definition of surrogate endpoints, describes the requirements of the IQWiG for the validation of the oncological endpoint progression-free

survival as surrogate for the patient-relevant endpoint overall survival and the resulting consequences.

4. Handling of missing values

Missing values in clinical trials bear the risk of biased estimates. Despite numerous strategies to avoid missings as early as during planning and implementation of the study, missing values will remain inevitable – with according consequences for the assessment of added benefit. Last observation carried forward (LOCF) no longer is the one and only „gold standard“. Which strategies should be used?

5. Indirect comparisons

Often, no direct head to head studies are available for proof of added benefit. The German Code of social law foresees the method of indirect comparison in these cases. Which preconditions have to be fulfilled for these indirect comparisons to be accepted by the assessors? Which experiences have pharmaceutical entrepreneurs made? And what can be learnt from this, either for the implementation of indirect comparisons in dossiers, but also for future study planning?

6. Health economic evaluation

The conduct of health economic evaluations are still possible after the AMNOG became effective and this paper will clarify the current role of health economic evaluations and what are the essential requirements and which aspects should be considered.

Every single article follows the same structure: It starts with an introduction in the topic, then followed by a description of the demanded requirements by IQWiG and G-BA with a critical appraisal and concludes with a recommendation. These papers are supposed to clarify the requirements for the benefit assessment and to give assistance and recommendations for the preparation of the benefit dossiers, especially for communication with colleagues outside of Germany.

Impact of label on benefit assessment **(Auswirkung des Labels auf die Nutzenbewertung)**

Author: Sarah Schmitter

Abstract

In dieser Publikation geht es um die zweifache Bedeutung der Formulierung des Labels auf das Ergebnis der Nutzenbewertung. Die Formulierung zum einen für die Bestimmung der ZVT relevant und zum anderen die Berücksichtigung der verfügbaren Studien. Idealerweise korrespondiert die Formulierung des Labels gänzlich mit den Einschlusskriterien der pivotalen Studien. Patienten, die von der Studie ausgeschlossen wurden, sind ebenso bei der Formulierung des Labels ausgeschlossen.

Topic

The wording of the label has a twofold relevance on the outcome of the benefit assessment:

1. Determination of the zweckmäßige Vergleichstherapie (ZVT, appropriate comparative therapy)
2. Consideration of available studies

The G-BA determines the ZVT based on the criteria in its code of procedure (s. next paragraph). As a result of the application of these criteria the entire population according label is often divided into different subpopulations. In this case the pharmaceutical manufacturer needs to prove an added benefit of its drug compared to the ZVT in each subpopulation. The G-BA usually determines different ZVTs for each subpopulation. There are two examples of products in which the population is divided and the G-BA acknowledged no added benefit in one subpopulation because of missing clinical data:

- **Axitinib (Inlyta®, Pfizer Pharma GmbH, RCC)**

Axitinib is indicated for the treatment of adult patients with advanced renal cell carcinoma (RCC) after failure of prior treatment with Sunitinib or a cytokine in the European Union. In 2012 the G-BA determined the ZVT for Axitinib considering two subpopulations: patients pretreated with Sunitinib (ZVT: Everolimus) and patients pretreated with a Cytokine (ZVT: Sorafenib). Because of the results of the pivotal trial which compares Axitinib and Sorafenib the G-BA conclusively decided on an added benefit in this subpopulation. However the G-BA didn't accept the unadjusted indirect comparison of Axitinib and Everolimus and conclusively decided on no added benefit in this subpopulation [1].

- **Crizotinib (Xalkori®, Pfizer Pharma GmbH, pretreated ALK+ NSCLC)**

Crizotinib is indicated for the treatment of adults with previously treated anaplastic lymphoma kinase (ALK)-positive advanced non-small cell lung cancer (NSCLC) in Europe. The G-BA determined the ZVT for Crizotinib considering two subpopulations: patients with ECOG-PS 0, 1 where applicable 2 (ZVT: Docetaxel or Pemetrexed) and patients with ECOG PS 4, 3 where applicable 2 (ZVT: Best supportive care (BSC)). Only patients with ECOG PS 0-2 were included in the clinical study. Therefore no data was available for patients with ECOG PS 3 or 4. This was the reason why the G-BA conclusively decided on no added benefit in this subpopulation [2, 3].

The wording of the label determines the consideration of available studies. If the available studies don't sufficiently match the population according to the label, no added benefit is usually granted. Two examples for this case are the benefit assessments of Vandetanib (Caprelsa®, AstraZeneca GmbH) and Fingolimod (Gilenya®, Novartis Pharma GmbH):

- **Vandetanib (Caprelsa®, AstraZeneca GmbH, MTC)**

Vandetanib was primarily indicated for the treatment of aggressive and symptomatic medullary thyroid cancer (MTC) in patients with unresectable locally advanced or metastatic disease. However patients with unresectable locally advanced or metastatic MTC were included in the pivotal study, i.e. patients with a non-aggressive and asymptomatic MTC were also included. The G-BA conclusively decided on no added benefit because the percentage of patients with aggressive and symptomatic MTC was too low to draw conclusions based on the entire population (s. next paragraph) and the pharmaceutical manufacturer submitted no subgroup analysis for these patients [4].

- **Fingolimod (Gilenya®, Novartis Pharma GmbH, multiple sclerosis)**

Fingolimod was primarily indicated as single disease modifying therapy in highly active relapsing remitting multiple sclerosis for the specific adult patient groups. *(Patients with high disease activity despite treatment with at least one disease modifying therapy (for exceptions and information about washout periods see sections 4.4 and 5.1). These patients may be defined as those who have failed to respond to a full and adequate course (normally at least one year of treatment) of at least one disease modifying therapy. Patients should have had at least 1 relapse in the previous year while on therapy, and have at least 9 T2-hyperintense lesions in cranial MRI or at least 1 Gadolinium-enhancing lesion. A "non-responder" could also be defined as a patient with an unchanged or increased relapse rate or ongoing severe relapses, as compared to the previous year. OR Patients with rapidly evolving severe relapsing remitting multiple sclerosis defined by 2 or more disabling relapses in one year, and with 1 or more Gadolinium enhancing lesions on brain MRI or a significant increase in T2 lesion load as compared to a previous recent MRI.)* The inclusion criteria of the pivotal study don't correspond to the label. The EMA bridged outcomes of the study to a patient group not included in the study based on experiences with another active agent. Because of this approach there was no data from studies investigating Fingolimod in this specific patient group available. Therefore the G-BA conclusively decided on no added benefit for this patient group [5] [6].

Therefore the wording of the label can be the reason why the G-BA granted no added benefit in a subpopulation or the entire population.

Requirements by IQWiG and G-BA

First the rules for the determination of the ZVT will be described and second the criteria to consider available studies. According to § 6 in the 5th chapter of the code of procedure by G-BA (§ 6 VerfO) the following criteria in particular have to be considered when the ZVT is determined:

1. **Insofar as a pharmaceutical is considered as the comparator, the pharmaceutical must be authorised for the therapeutic label.**

The reason for this criterion is that the G-BA is not allowed to encourage off-label use of pharmaceuticals. However the interpretation of a specific label can differ between the G-BA and the pharmaceutical manufacturer. For example in the case of the Axitinib benefit assessment Pfizer held the opinion that Sorafenib is indicated after Sunitinib treatment. In contrast the G-BA stated that Sorafenib is not indicated in this case. This was one reason why Sorafenib was not determined as ZVT for patients pretreated with Sunitinib [1].

2. **Insofar as a non-pharmaceutical treatment is considered as the comparator, this must be deliverable within the framework of the statutory health insurance.**

The decisions by the G-BA apply to the statutory health insurance funds, persons insured by these funds, the responsible physicians and other service providers and are binding for these parties. Therefore the scope of the G-BA is the treatment (medical and non-medical) which is reimbursed by the statutory health insurance.

3. **Pharmaceuticals or non-pharmaceutical treatments are preferred as comparator, whose patient-relevant benefit has already been determined by the Federal Joint Committee.**

Treatments with an acknowledged benefit by G-BA are preferred as ZVT, if there are several possible comparators. The patient-relevant benefit could be determined based on a benefit assessment according § 139a SGB V and § 35b SGB V as well as on a benefit assessment according § 35a SGB V. The number of benefit assessments will increase over time with that more and more already assessed pharmaceuticals will be determined as ZVT.

4. **The comparator should belong to the appropriate therapies in the therapeutic indication according to the generally accepted state of medical knowledge.**

The “generally accepted state of medical knowledge” is operationalized by G-BA information from systematic reviews, meta-analyses, HTA-reports and evidence based systematic guidelines. The G-BA conducts a systematic literature research to identify relevant publications.

The G-BA needs to publically justify the determination of the ZVT. Therefore a document showing the way of proceeding is published on the internet at the time of the publication of the dossier [7].

These documents (e.g. [8, 9]) reveal how the G-BA proceeds by determining the ZVT:

1. **Summary of all pharmaceuticals approved within the indication with the ATC-code and the wording of the label**
2. **Systematic literature research regarding systematic reviews, meta-analyses, HTA-reports and evidence-based systematic guidelines within the indication**

Only documents published within the last 5 years are considered and synopsis of the relevant evidence to determine the ZVT. The search is conducted in the following sources Cochrane

Library, MEDLINE, Leitlinien.de (ÄZQ), AWMF, DAHTA, G-BA, GIN, IQWiG, NGC, TRIP. Additionally an internet search regarding current German and European guidelines (e.g. NICE, SIGN) is performed.

3. Assessment of the results regarding the relevance and quality of the methodology

4. Synoptic summary of the relevant evidence

The G-BA informs the pharmaceutical company about the ZVT in an advice meeting taking place within 8 weeks after the manufacturer's submission of a request for advice. In this request the manufacturer can explain in detail which is the appropriate ZVT from his point of view and the reasons for his opinion [7].

The IQWiG assesses the added benefit of a new pharmaceutical compared to the ZVT determined by G-BA on the basis of the submitted benefit dossier. It doesn't challenge the decision by the G-BA. However it examined whether the comparator selected by the pharmaceutical company corresponds to the determined ZVT [10].

The code of procedure by G-BA says about the considered studies in § 5 sec. 2 G-BAVerfO: „The proof of a therapeutic improvement is provided on the basis of a Summary of Product Characteristics (SmPC) and through assessment of clinical studies according to the international standards of evidence-based medicine. Clinical studies are preferred, in particular direct comparison studies with other pharmaceuticals [...] have to be taken into account.” [7] Besides the information that the assessment is done based on the methodology of evidence based medicine and that direct comparisons are preferred the G-BA does not make clear statements regarding the considered studies. The G-BA assigns the IQWiG with the assessment of the benefit dossier and the IQWiG follows its general methods published on its website. Therefore these methods have a huge impact on the outcome of the assessment. In section 3.3.1 Relevance of the drug approval status of IQWiG General Methods Draft of Version 5.0 it is said that only results of studies being conducted within the framework of the label of the assessed drug will be considered. “As a rule, study results are regarded to be “not applicable” if, for example, the age range or disease severity treated lay outside the approved label, if off-label combinations including other active ingredients were used, or if studies were conducted in patients with contra-indications for the intervention investigated.” [10] If a study is only partly conducted within the label, i.e. deviant population, intervention or comparator, the following criteria needs to be fulfilled (8.1.1 Criteria for study inclusion):

- ≥ 80% of the patients corresponded to the population according label
- ≥ 80% of patients receive the intervention according the SmPC
- ≥ 80% of patients receive the requested comparator intervention

Otherwise IQWiG doesn't consider the study in its assessment unless subgroup analysis for the patients correspond to the label or a rationale for the transferability of the data are provided [10]. In the case of Vandetanib (Caprelsa®, AstraZeneca GmbH, MTC) only 56 % of the patients in the study corresponded to the population according label and therefore the study results were not considered for the benefit

assessment [4]. In the benefit assessment of Ruxolitinib (Jakavi®, Novartis Pharma GmbH) in the indication “Polycythaemia vera (PV)” was the same problem. The pharmaceutical manufacturer provided sensitivity analyses to show the transferability. Due to these analyses IQWiG concluded that the results of the entire trial could be transferred to the patients corresponded the label. An added benefit was granted by G-BA [11, 12].

The benefit assessment and pricing process according AMNOG refers to pharmaceuticals being reimbursed by the statutory health insurance. This implies that the pharmaceuticals are assessed within their label. It is said in § 35a section 1 in the German Social Code Book V (SGB V, Sozialgesetzbuch V) that the basis of the assessment are the international standards of evidence based medicine. Therefore it is essential that the clinical trials must be conducted within the given label, otherwise it is not possible to achieve an added benefit based on the methodology of evidence based medicine.

Appraisal and scientific rationale for requirements

The meeting “IQWiG in dialogue 2013” with the title “Importance of approval status for the benefit assessment” emphasizes the relevance of the label in the German benefit assessment. The discussion and the statements by representatives of IQWiG, Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM, Federal Institute for Drugs and Medical Devices) and the industry show that everyone is aware of the problem, but no solution to satisfy everyone is available at the moment for the question how to deal with label deviating from the pivotal studies [13].

Implication

The wording of the label has impact of the determination of the ZVT by the G-BA as well as on the consideration of clinical trials as described above. If no sufficient clinical data is available for the comparison to the ZVT the result of the benefit assessment will be “no added benefit”. This is also the case if the label deviates from the population or application in the clinical trial because no sufficient data for the assessment is available. The ZVT has also impact on the negotiations of the reimbursement price because its price is one component for the determination of the reimbursement price.

Ideally the wording of the label totally corresponds to the inclusion criteria of the pivotal study and patients excluded from the study are also excluded by the wording of the label.

Furthermore the splitting of population into different subpopulations reduces the number of patients within a population. Subsequently the confidence intervals become broader in most cases. This could decrease the recommended extent of the added benefit by IQWiG being determined based on the confidence intervals. Furthermore if the randomization is not stratified according the subpopulation there will be the possibility that the results can be biased. This would result in a low reliability of the added benefit and so in a bad starting point in the negotiations of the reimbursement price.

The analyses of benefit assessments revealed that besides the wording of the label and the EPAR also the following SmPC sections could have an impact on the benefit assessment: 4.2 Posology and method of administration; 4.3 Contraindications; 4.4 Special warnings and precautions for use and probably 4.5 Interaction with other medicinal products and other forms of interaction. The reason is that IQWiG takes

the view that only studies in which all drugs (incl. comparator) are used according the SmPC can be considered in the benefit assessment.

Recently it is noticed that the G-BA sometimes granted an added benefit based on trials where the comparator is not approved in this specific indication (e.g. [14, 15]). This seems to be the case if the drug is approved in a late treatment line and the guidelines recommend drugs being not approved in this indication. Furthermore the statements of the medical societies have probably an impact.

Health related quality of life - HRQoL **(Gesundheitsbezogene Lebensqualität)**

Author: Niclas Kürschner

Abstract

Diese Publikation beschreibt das Konzept der gesundheitsbezogenen Lebensqualität inklusive des Standpunkts der Zulassungsbehörden. Weiterhin werden die methodischen Anforderungen an die Analyse von Lebensqualitäts-Daten diskutiert, um die zukünftige Akzeptanz durch das IQWiG und den G-BA innerhalb der frühen Nutzenbewertung zu ermöglichen.

Background

According to Walters [16], quality of life is a multidimensional concept, health related quality of life (HRQoL) considers the health related aspects and may include symptoms, physical functioning, general health, emotional functioning, social well-being, cognitive functioning and role functioning. This is a broad definition, especially one other term is rather frequently used within the discussion: Patient Reported Outcomes (PRO). According to Cappelleri et al. [17] PRO measures accommodate health questionnaires whose objective is to measure a patient's health from the patient's perspective rather than from a physician, caretaker or biological measure. HRQoL measurements are generally Patient Reported Outcomes. The U.S. Food and Drug Administration (FDA) published a "Guidance for Industry" in 2009 regarding PRO measures and their use in medical product development to support labeling claims [18]. A PRO is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. The outcome can be measured in absolute terms (e.g., severity of a symptom, sign, or state of a disease) or as a change from a previous measure. In the guidance by FDA multiple aspects are described: the conceptual framework of a PRO instrument; requirements regarding validity, reliability, ability to detect change; instrument modification; clinical trial design and statistical considerations (e.g. handling missing data, considerations for using multiple endpoints [18]. Also, the European Medicines Agency (EMA) published a reflection paper in 2005 regarding the regulatory guidance for the use of HRQoL measures in the evaluation of medicinal products. This short paper describes requirements for study design, statistical analysis (e.g. being adapted to address for multiplicity issues and using validated instruments), hypothesis and missing data [19]. Recent developments lead to an another EMA reflection paper on the use of PRO and HRQoL measures in oncology studies. This paper takes the increasing importance of PRO into account and describes the framework for drawing regulatory conclusions based on PRO more detailed [20]. It was followed by an adopted Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man - the use of patient-reported outcome (PRO) measures in oncology studies [21]. In addition to both regulatory authorities, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) offers a variety of initiatives regarding Patient Reported Outcomes and HRQoL [22]. Furthermore the International Society for Quality of Life Research (ISOQoL) committed itself to advance the scientific study of HRQoL and other patient-centered outcomes to identify effective interventions, enhance the quality of health care and promote the health of populations [23].

HRQoL gained importance through the benefit assessment in Germany. Within the benefit assessment according to AMNOG, the following outcomes are considered to be patient relevant: mortality, morbidity and HRQoL, as well as reduction in disease duration and adverse effects [24]. This emphasizes, that HRQoL has gained a high perception within the last years. For example, Josef Hecken, impartial chair of the G-BA mentioned on a health care committee meeting publicly, the importance of HRQoL, especially in oncology. He criticized based on completed benefit assessments in oncology, that the added benefit was only acknowledged because of a prolongation of overall survival without considering the quality of life. Regarding a long term perspective G-BA will increase the focus on HRQoL and adverse events data to evaluate these aspects in comparison to a prolonged e.g. overall survival [25], in addition missing HRQoL data could lead to negative consequences within the benefit assessment [26].

Requirements by IQWiG and G-BA

According to G-BA, the analysis of HRQoL demands the same methodological requirements regarding study design, data analyses and data assessment as for other patient relevant endpoints. HRQoL assessment demands psychometric validated instruments, preferably using a disease specific and a generic instrument. Generic instruments often lack of sensitivity to detect disease specific effects [27, 28]. As information on PROs is subjective due to their nature, open studies in this area are of limited validity. The size of the effect observed is an important decision criterion for the question as to whether an indication of a benefit of an intervention with regard to PROs can be inferred from open studies. Empirical evidence shows a high risk of bias for subjective outcomes in open studies, which should be considered in their interpretation. Especially in situations where blinding of physicians and patients is not possible, efforts are required to minimize and assess bias (e.g. blinded documentation and assessment of outcomes) [27].

Figure 1: Thresholds for determining the extent of an effect (source: [27])

| | | Outcome category | | |
|---|--------------|---------------------|---|---|
| | | All-cause mortality | Serious (or severe) symptoms (or late complications) and adverse events, as well as health-related quality of life ^a | Non-serious (or non-severe) symptoms (or late complications) and adverse events |
| Extent category | Major | 0.85 | 0.75 and risk $\geq 5\%$ ^b | Not applicable |
| | Considerable | 0.95 | 0.90 | 0.80 |
| | Minor | 1.00 | 1.00 | 0.90 |
| a: Precondition (as for all patient-reported outcomes): use of a validated or established instrument, as well as a validated or established response criterion. | | | | |
| b: Risk must be at least 5% for at least 1 of the 2 groups compared. | | | | |

Figure 1 shows the thresholds (upper confidence interval limits) for determining the extent of an effect, for example an added benefit in HRQoL (middle column). The precondition for determining the extent of added benefit for outcomes on HRQoL is that both the instruments applied and the response criteria must be validated or at least generally established. Two-sided 95 % confidence intervals, especially the upper confidence interval limit, for the effect measure (hazard ratio / relative risk) is required to determine the extent of the effect [27]. Responder analyses are used to determine the extent of added benefit in the case of continuous or quasi-continuous outcomes. For this purpose, a validated or established response criterion or cut-off value is required. On the basis of the responder analyses (2x2 tables) the relative risks are calculated directly from them [27].

EQ-5D & QALYs

Cost utility analyses, for example using EQ-5D to calculate QALYs, and are not accepted in Germany to display an added benefit, because QALYs (Quality Adjusted Life Years) do not measure the patients HRQoL. Instead, it is attempted to approximate the benefit function according to micro economic theory. These preferences can be used according to economic welfare theory to determine allocation decisions. Generic measurements of functions, e.g. through EQ-5D in clinical studies, are weighted using tariffs. Tariffs are an assessment of health conditions which were assessed through a sample of insured people [29]. QALYs are established as the main valuation technique for policy making or for reimbursement decision making in many countries. For example, they are recommended by bodies such as the National Institute for Health and Care Excellence (NICE). These cost-effectiveness analyses are not accepted in Germany. The QALY concept also remains to be discussed controversial, some experts claim, that the techniques used as a basis for QALY value calculation are flawed. In particular, the underlying assumptions of the multiattribute utility model do not correspond to behavior patterns observed in a real population [30].

Appraisal and scientific rationale for requirements

In only few cases receiving an added benefit due to HRQoL was achieved. The following examples from current assessments will be described to display the challenges for HRQoL in the benefit assessment according to AMNOG:

- Macitentan (Opsumit®, Actelion Pharmaceuticals, pulmonary hypertension): the generic instrument SF-36 was used in the clinical study and according to G-BA's early benefit assessment, SF-36 is not validated for pulmonary hypertension. In the clinical study, a statistical significant (but numeric low) improvement of the scores was shown, but individual responder analyses using a valid minimal important clinical difference (MCID) were not presented and hence no added benefit considering HRQoL was assessed [31].
- Pertuzumab (Perjeta®, Roche Pharma, HER2-positive metastatic or locally recurrent unresectable breast cancer): In the clinical study, FACT-B (Functional Assessment of Cancer Therapy – Breast Cancer) in version 4 was used to assess HRQoL. According to IQWiG's benefit assessment, only the third version of FACT-B is considered as valid, due to missing validation studies and a

changed subscale with unclear validity in version 4. In addition, the responder analyses were conducted using only a selection of scales which leads to high uncertainty [32].

- Nintedanib (Vargatef®, Boehringer Ingelheim, advanced/metastatic NSCLC): Both responder analyses and time-to deterioration for the EORTC-QLQ C30 and LC13 questionnaire using a MID of 10 points were presented. Although responder analyses have been accepted in previous benefit assessments in NSCLC [2], the G-BA finally considered only the results for time-to deterioration [33]. In the benefit assessment of Afatinib (Giotrif®, Boehringer Ingelheim, advanced/metastatic NSCLC) responder analyses weren't accepted either. IQWiG proposed time-to-improvement analyses to consider the different observational times in the treatment arms [34].
- Several examples exist in benefit assessments of oncological drugs [35, 36], where an added benefit was declined besides a prolongation of OS and under consideration of harm due to adverse events and HRQoL data. This emphasizes the increasing importance of HRQoL in the German benefit assessment.

Methodological requirements

- It is a prerequisite to perform analyses according to IQWiG's methods otherwise the results will not be accepted: hence, validated HRQoL measurement instruments and validated MID/MCID (minimal important (clinical) difference) are one of the keys to achieve an added benefit due to an improvement in HRQoL. The validity has to be proven for the current version of the instrument") [27]
- The requirements demand to conduct responder analyses using a valid response criteria, regarding the determination of added benefit according to IQWiG methodology [27]. The methodological discussion within the benefit assessment of Nintedanib and Afatinib described above, may result in time-to-event analyses to be considered as appropriate method for HRQoL analyses in case of different follow-up times between treatment arms or where survival times matter, e.g. in oncological assessments.
- This includes time-to-deterioration- and time-to-improvement-analyses composed of hazard ratios + 95%-CIs and Kaplan-Meier-Plots. Deterioration is mostly defined as a decrease from baseline by X points and improvement mostly as an increase from baseline by X points according to valid MID described above. Exceptions may occur considering the specific measurement or its scoring.
- **In case were MID/MCID ranges for are reported in the literature, e.g. for EORTC-QLQ C30 [37], the use of sensitivity analyses using the respective separate response criteria might be discussed**
- Measurement after end of study or after progression in oncological trials
 - No measurement at end of treatment (e.g. caused through toxicities) may lead to bias due to informative censoring. In that case, sensitivity analyses using complex statistical modelling (competing risk model, pattern mixture model) might be appropriate
 - In general, especially in oncological trials, a post-progression measurement and assessment of HRQoL data is highly recommended

- Missing values have to be taken seriously and also non-blinding may lead to downgrading (result certainty)
 - Unblinded measuring of subjective criteria such as PRO may lead to bias (blinding not always possible due to study design issues). Efforts are required to minimize and assess bias [29].
 - The G-BA does not accept as a general rule more than 20% of missing values [28].
- Disease specific measurements are required
 - Validity of the instrument must be proven within the disease area, which will often disqualify generic instruments such as the SF-36 because validated MID are not available for all indications [28].
- Validation study
 - In case that the used HRQoL measurements are not validated or a validation study to cite is not available, a validation study has to be conducted to allow acceptance within the IQWiG's benefit assessment. The criteria are not explicitly formulated in the IQWiG's methods, hence the validation study has to meet the highest scientific standards to allow its acceptance without flaws in the study design. Some of the following aspects, which are also described in [18], are essential in conducting a validation study:
 - Reliability: Internal consistency reliability will be estimated for each multi-item subscale with Cronbach's coefficient alpha. Cronbach's coefficient alpha is a measure of how closely correlated a set of items are with each other. A Cronbach's coefficient alpha of 0.70 or greater is generally considered acceptable for group comparisons.
 - Validity: Several types of validity must be evaluated
 - Construct-related: convergent and divergent validity, known-groups validity
 - Criterion-related: associations with criterion measures (valid measures of the same concept)
 - If applicable: Comparison of alternative methods of administration (paper vs. iPad)
 - Responsiveness: ability to detect change, evidence is required that the instrument can identify differences in scores over time in individuals or groups who have changed with respect to the measurement concept
 - Cross-cultural adaptation of research instruments has to be taken into consideration to assure validity of the instruments. Previous validation (studies) may not be valid in another language, time, culture or context, hence the instrument has to be adapted accordingly, a simple linguistic translation is not sufficient for other cultural settings [38].

Validation of surrogate endpoints for patient-relevant endpoints (Validierung von Surrogatparametern für patientenrelevante Endpunkte)

Author: Johanna Gillhaus

Abstract

In dieser Publikation wird die Definition von Surrogatendpunkten sowie die Anforderungen des IQWiG zur Validierung des onkologischen Endpunkts progressionsfreies Überleben als Surrogat für den patientenrelevanten Endpunkt Gesamtüberleben und deren Konsequenzen beschrieben.

Background

In the benefit assessment and pricing process according § 35a SGB V (AMNOG) Gemeinsamer Bundesausschuss (G-BA, Federal Joint Committee) decides on the extent and probability of the added benefit of a drug compared to the zweckmäßige Vergleichstherapie (ZVT, appropriate comparative therapy). The decision is based on the data presented in the benefit dossier submitted by the company and the assessment of the data by Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG, Institute for Quality and Efficiency in Health Care). On the basis of the G-BA decision companies have to negotiate the reimbursed price with the Spitzenverband der gesetzlichen Krankenkassen (GKV-SV, National Association of the Statutory Health Insurance Funds).

Topic

A clinical endpoint “reflects how a patient feels, functions, or survives” [39]; this is how the IQWiG defines patient relevance [27, 40]. Patient-relevant endpoints are the following according to §35b SGB V: increase in life expectancy, improvement in health status and quality of life, as well as reduction in disease duration and adverse effects.

However, often the patient-relevant endpoint might be difficult to measure due to complexity and/or duration of research. Reasons may be costs, time (i.e. long follow-up time, e.g. clinical progress, survival), the need to test multiple regimes, ethical considerations [41, 42]. A solution for these issues is the replacement of the true clinical endpoint by a surrogate. This is especially often the case in the field of oncology, but also true for other indications (see Table 2). Surrogate endpoints can be achieved in a “more quickly, less invasive and less costly manner” [43].

Table 2: Examples of possible surrogate endpoints (adopted by [39, 44])

| Disease | Surrogate endpoint | Clinical endpoint |
|------------------------|---------------------------|-----------------------|
| Cancer | Progression-free survival | Survival |
| Osteoporosis | Bone mineral density | Fracture |
| Cardiovascular disease | Ejection fraction | Myocardial infarction |
| Hypertension | Blood pressure | Survival |
| HIV infection | CD4 counts + viral load | Survival |

A surrogate is “a biomarker that is intended to substitute for a clinical endpoint” and is “expected to predict clinical benefit” [40]. It is a common misled that biomarkers automatically represent a valid surrogate [45]. The effect of treatment on a surrogate endpoint must be reasonable and relevant to the patient. However, sometimes “surrogates may not be a true predictor of the clinical outcome” [46] and/or most surrogate endpoints are unreliable in regard to patient relevance [27, 47, 48]. The ICH E8 states “A surrogate endpoint is an endpoint that is intended to relate to a clinically important outcome but does not in itself measure a clinical benefit. Surrogate endpoints may be used as primary endpoints when appropriate (when the surrogate is reasonably likely or well known to predict clinical outcome).” [49]

Regulatory authorities usually agree to use surrogate endpoints in oncological studies as they evaluate the effectiveness of a drug [50]. In the case of accelerated approval of anti-cancer drugs “the FDA may approve a drug based on evidence that the drug shrinks tumors, because tumor shrinkage is considered reasonably likely to predict a real clinical benefit” [51]. In both the regular and accelerated approval the FDA accept surrogates such as disease-free survival, objective response rate, complete response or progression-free survival. “Acceptable primary endpoints include cure rate, OS and PFS/DFS” in confirmatory phase III trials according to the EMA “Guideline on the evaluation of anticancer medicinal products in man”. This is due to the fact that the EMA considers “prolonged PFS/DFS as such, however, to be of benefit to the patient” [52].

Within the scope of benefit assessments the patient-relevant benefit or added benefit is assessed [27]. Only patient-relevant endpoints are acknowledged by the G-BA resp. the IQWiG. For this purpose it is absolutely necessary to validate a surrogate as a clinical, patient-relevant endpoint according IQWiG’s requirements; i.e. progression-free survival as a surrogate endpoint for overall survival.

Requirements by IQWiG and G-BA

Several methods for the validation of surrogate endpoints have been proposed [43, 53]. In 2011, the IQWiG published a rapid report (“Validity of surrogate endpoints in oncology”) giving a detailed statement concerning surrogate endpoints in the field of oncology by order of G-BA. The goal of this research was to present, to “summarize and to assess methodological approaches aiming to validate surrogate endpoints” [54]. Based on a systematic literature review, IQWiG detected the following different approaches: it involves (1) test-based methods, (2) causal model-based methods, (3) correlation-based models and (4) other approaches.

The test-based models are based on the Prentice’s criteria [55]. These criteria are restrictive and require significant treatment effects on surrogate and patient-relevant endpoints. The “assumptions of the Prentice’s criteria are “too stringent to be fulfilled in real situations” [39]. The causal model-based approaches are statistically challenging and require knowledge of specific causal paths of disease process [54]. The correlation-based models are the most used and topically in practice and research. They investigate correlation of treatment effects on surrogate and true endpoints on a study level as well on an individual level [39]. This last approach is recommended by the IQWiG.

Since correlation isn't equated with validation [44] it is not sufficient solely to show this relationship [54]. The ICH E9 states: "Relationships between clinical and surrogate variables for one product do not necessarily apply to a product with a different mode of action for treating the same disease." [46] The IQWiG agreed with this issue concerning that the validity of a surrogate must be both disease-specific and intervention-specific (see IQWiG Rapid Report [54], examples for breast cancer and colorectal cancer). In addition, analyses on heterogeneity should be available. A comprehensive data basis is required as both the surrogate and the clinical endpoint must have been analyzed in the included studies purposing validating the surrogate by the means of a meta-analysis [54].

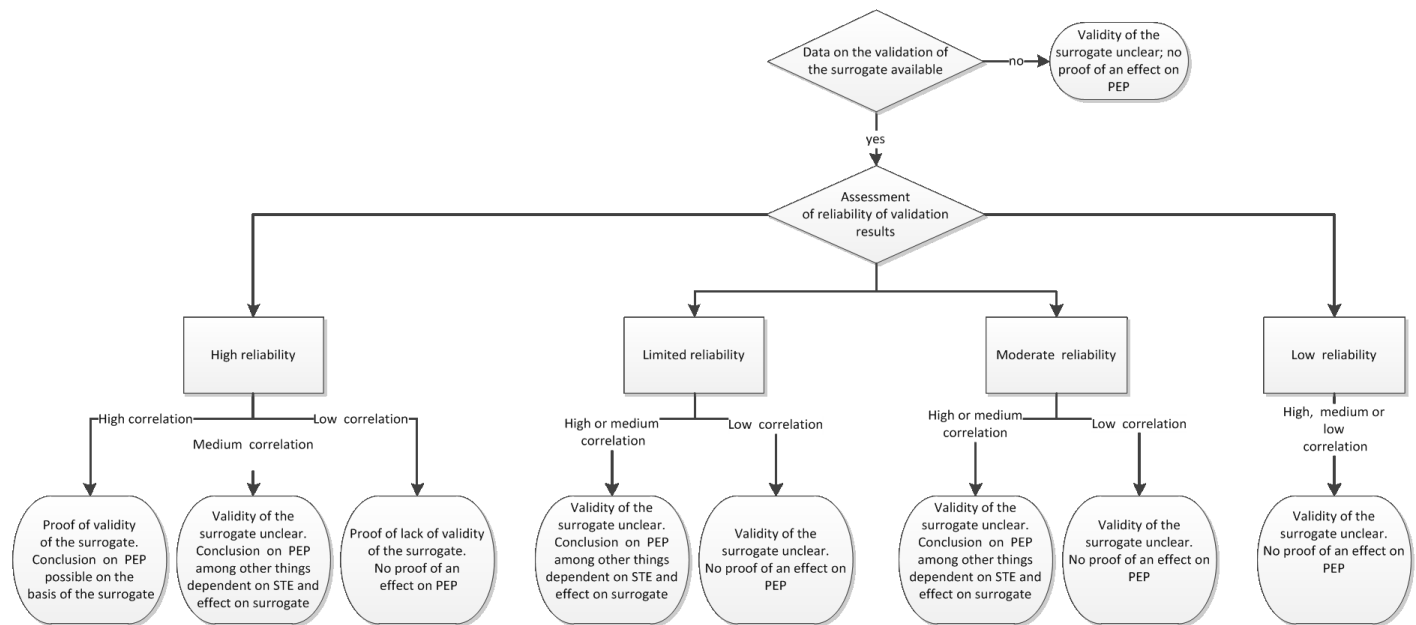
On the basis of this the IQWiG claimed a meta-analysis on the basis of RCTs including the surrogate and the patient-relevant endpoint, so that the correlation coefficients between treatment effects of surrogate and endpoint may be calculated. These calculated correlation coefficients are classified into three levels: high, medium and low correlation. A correlation is judged as high, if the lower confidence limit of the correlation coefficient is above 0.85, a correlation is judged as low if the lower confidence limit is below 0.7. If effects and confidence limits are in between these marked-out boundaries, then the correlation is judged as moderate. However, there is no consistent opinion on this classification among the scientific community [56-58]. Further, the IQWiG classifies the reliability of the validation study depending on the certainty of results of the study: high, limited, moderate and low reliability. If the validity of a surrogate is not clear, the concept of surrogate threshold effects (STE) may be applied according to IQWiG Rapid Report required. It is defined "as the minimum treatment effect on the surrogate necessary to predict a non-zero effect on the true endpoint" [45]. So, the "prediction of the treatment effect on the true endpoint from the effect on the surrogate is allowed" [45].

Described requests above are taken as the basis for a proposal of the IQWiG. It developed a two-level algorithm for handling surrogate endpoints in benefit assessments [54]:

7. Assessment of the validity of the surrogate
8. Conclusions about patient-relevant endpoints drawn from the effects on surrogates depending on the characterization of the validity of the surrogate

These considerations should be followed for a validation of a surrogate as a patient-relevant endpoint which is accepted by the IQWiG and the G-BA.

Figure 1: Impact (of the reliability of validation studies and the correlation of the effects on the surrogate and patient-relevant endpoint) on the validity of the surrogate (source: [59])



In the history of previous benefit assessments of anticancer drugs, in three benefit assessments case of validation of progression-free survival as surrogate for overall survival have been attempted: Dabrafenib (Tafinlar®, GlaxoSmithKline GmbH & Co. KG, unresectable melanoma or metastatic melanoma) [60], Pembrolizumab (KEYTRUDA®, MSD SHARP & DOHME GMBH, unresectable melanoma or metastatic melanoma) [61] and Palbociclib (Ibrance®, Pfizer Pharma GmbH, locally advanced or metastatic breast cancer) [62]. In all three cases, the IQWiG refused the acceptance of the surrogate validation due to several causes. Common reasons were that the IQWiG didn't fully accept the conducted literature researches (mainly due to technical reasons) and the selected studies for the meta-analysis (e.g. the different therapeutic schemes (mono vs. combination therapies) resp. substance classes (cytostatic therapies vs. targeted therapies) or definition of endpoints in studies) or the IQWiG criticized technical/methodological aspects. As a consequence of this it is vitally necessary to comply strictly the proposal of the IQWiG for the validation of a surrogate as a patient-relevant endpoint.

Appraisal and scientific rationale for requirements

In 2009, the Deutsches Institut für Medizinische Dokumentation und Information (DIMDI, German Institute of Medical Documentation and Information) published a report on the dealing of surrogate endpoints in international health technology agencies and assessments [63]. DIMDI conducted a systematic literature research aiming to find definitions of surrogate endpoints and validation methods. According to them, only the HTA-agencies of Australia (MSAC), Canada (CADTH) and New Zealand (PHARMAC) provide formal definitions for a surrogate endpoint which are very similar are to those to the German IQWiG Rapid Report. The validity of surrogate endpoints is topic in methods of several HTA-agencies stating that statistical, biological and pathophysiological association resp. pathway is necessary

and evidence from RCTs considering the surrogate and the clinical endpoint. All HTA-agencies prefer clinical endpoints instead of surrogate endpoints. Pursuant to the DIMDI report, there is no common gold standard concerning the validation of surrogate endpoints, however, meta-analytic approaches as recommended by the IQWiG are desired.

Recommendations

As written in 'Requirements by IQWiG and G-BA' the requirements stated in the rapid report ("Validity of surrogate endpoints in oncology") should be complied. The IQWiG claims:

1. Systematic literature research (according to the IQWiG criteria!; see IQWiG rapid report [54] and IQWiG methods [27])
2. Disease- and intervention-specific meta-analyses
3. Calculation of correlation coefficients (and surrogate threshold effects)
4. Sensitivity analyses

A stringently followed approach is vital for a successfully accepted validation of PFS for OS by the IQWiG and G-BA. This includes a transparent presentation of the methodological strategy as it is required especially by the IQWiG.

Handling of missing values in early benefit dossiers **(Handhabung von fehlenden Werten in der frühen Nutzenbewertung)**

Author: Julia Schiffner-Rohe, Astrid Genet

Zusammenfassung

Fehlende Werte in klinischen Studien tragen das Risiko von verzerrten Schätzern. Trotz zahlreicher Strategien bereits während der Planung und Durchführung einer Studie, um fehlende Werte zu reduzieren, sind fehlende Werte unvermeidbar – mit entsprechenden Konsequenzen für die Bewertung eines Zusatznutzens im Rahmen der frühen Nutzenbewertung. Das Fortschreiben des letzten verfügbaren Wertes (*Last observation carried forward -LOCF*) ist nicht mehr der “Goldene Standard”. Welche Strategien sollten also genutzt werden?

Abstract

Missing values in clinical trials bear the risk of biased estimates. Despite numerous strategies to avoid missings as early as during planning and implementation of the study, missing values will remain inevitable – with according consequences for the assessment of added benefit. Last observation carried forward (LOCF) no longer is the one and only „gold standard“. Which strategies should be used?

Topic

“It should be the aim of those conducting clinical trials to achieve complete capture of all data from all patients, including those who discontinue from treatment. Whilst it is unavoidable that some data are missing from all confirmatory clinical trials, it should be noted that just ignoring missing data is not an acceptable option when planning, conducting or interpreting the analysis of a confirmatory clinical trial. [...] Interpretation of the results of a trial is always problematic when the proportion of missing values is substantial. When this occurs, the uncertainty of the likely treatment effect can become such that it is not possible to conclude that evidence of efficacy has been established. [...] Ignoring missing data in the analysis violates the strict ITT principle which requires measurement and analysis of all patient outcomes regardless of protocol adherence.”[64]. While missing values jeopardize validity of results per se, this is even more the case when missing values occur in a different frequency in therapy groups to be compared. Further, “It should be noted that the strategy employed to handle missing values might in itself be a source of bias” [65]. While analysts should use a conservative approach – not systematically in favor of the test substance, there is no clear “one size fits all”-method for handling of missing values. Often enough, it cannot even be predicted whether derived results tend to be conservative or anti-conservative.

From the IQWiG perspective, benefit assessments have to be performed and presented using data on an ITT analysis. Whether ITT has been addressed correctly is a fundamental criterion during assessment.

Same priority is given to the aim of gaining unbiased results for benefit assessment. This makes it clear that handling of missing values is a critical topic in assessments. Some measures can be taken during planning and implementation of clinical trials to keep the amount of missing values as low as possible. These measures are described elsewhere and are not within the scope of this paper. This document focuses on how to handle missing values in the analysis phase of individual clinical trials within the frame of the benefit assessment.

Requirements by IQWiG and G-BA

According to IQWiG, the “probability of added benefit” is down-graded if based on potentially biased evidence. As a consequence, missing values directly impact the “probability of added benefit” if the possibility of bias due to missing cases cannot be ruled out. On the other hand, IQWiG is very vague on its requirements on how to handle missing values. There is no clear official recommendation on the right way to handle missing values, only a few statements can be found.

Statements from the IQWiG Methods paper [27] for handling of missing values

The methods paper by IQWiG [27] does not give precise recommendations regarding the way to handle missing values for the benefit assessment and mostly states that no general rule applies. When missing values do not result from a random mechanism (i.e. missing completely at random, MCAR,) they may lead to biased results. In this case reasons for missing values and potential consequences on the results are to be discussed individually. Statistical imputation strategies should then be applied to adjust for bias. Different approaches are possible and none them is generally accepted. According to the EMA, IQWiG recommends to perform sensitivity analysis using different methods and to mirror the results.

The special case of “escape strategies”

A special case is the situation where “missing” values derive from escape strategies in trial design: Patients on control arm may switch to treatment arm at pre-defined criteria (e.g. progression of disease). For trials especially in the oncological field, IQWiG has developed following recommendations for analysis [66]:

- Perform first analysis/data cut BEFORE patients can switch treatment
- Use methods of survival analysis (if adequate)
- Compare crude rates under consideration of direction of bias

Examples from IQWiG assessments

To compensate the lack of clear requirements for the general case, this section inventories IQWiG comments and recommendations regarding missing values found in IQWiG assessments of submitted benefit dossiers. Noteworthy, to date, IQWiG only refers to handling of missing values for individual trials.

- Missing values at baseline: when baseline is missing for some patients, randomization might be broken regarding this criterion and this could have an effect on disease progression. Biased results cannot be excluded [67]. The level of evidence can be down-graded.
- Proportion of missing values occurring during study: Especially when the proportion of missing values is high, IQWiG requires a discussion on the reason for missing values [68]. In general, IQWiG differentiates between a common (balanced) rate of missing values in both treatment arms and consequences deriving from unbalanced rate of missing values between treatment arms.

- Balanced proportion between study arms: a large proportion of missing values ($\geq 30\%$ in both study arms) invalidates the endpoint. The effect size is expected to be strongly biased by missing information and an objective interpretation of the results is not possible [11, 69]. A smaller proportion of missing values ($\geq 10\%$ in both arms) can still down-grade the level of evidence [70].

An important question related to those thresholds is how to compute the percentage of missing values. While the numerator has to be the patient count at the time of evaluation (for instance end of study), the denominator could be either the number of randomized patients (Full analysis set) or the number of patients at baseline. IQWiG assessment [11] helped us determine that IQWiG considers the number of randomized patients in its estimation.

- Unbalanced proportion between study arms:
 - For randomized controlled trials (RCTs), IQWiG introduced a benchmark for unbalanced rates of missing values between both arms. In case this difference was reached, results were to be subject to sensitivity analysis and evidence graded downward if results were inconsistent between sensitivity analysis and original results. For the assessment reports A10-01 and A15-04 [69, 71] this benchmark was set to 10% difference. In an earlier assessment [72], this benchmark was even set to 5% difference with studies to be excluded from evidence syntheses if endpoint data was missing for more than 10% of patients.
 - Observational studies (Cohort or cross-sectional) normally are not relevant for early benefit assessments, unless used to demonstrate a so-called “dramatic effect” (see IQWiG Methods [27], p. 48f for definition). For observational studies to be included in evidence synthesis, as a rule of thumb, IQWiG states a benchmark of no more than 30% missing information on endpoint data (safety or efficacy). Furthermore, studies should not differ in missing data rate of more than 15 %-points in treatment arms [73-75]. In certain cases, it could be derived from this rule (e.g. [72]).
- Imputation of missing values: for its assessment, IQWiG requires unbiased estimates of the effect size, which cannot be obtained when values are missing. IQWiG tends to prefer pre-specified analyses to be considered the main approach. However, when pre-specified approaches could be considered biased, it is recommended to implement other imputation strategies as sensitivity analyses. The following gives an overview of methods used to date within the framework of the benefit assessment, together with their chance of acceptance by IQWiG.

- LOCF: IQWiG clearly criticizes underlying evidence on clinical trials with documented/used LOCF only. In diseases with progression over time (e.g. Alzheimer Demenz – [76]); Rheumatoide Arthritis [69]), effect estimates are presumed to be biased.
- Artificial coding:
 - IQWiG itself proposed the implementation of “non-response” (binary coding) respectively “zero change” (continuous coding) for missing values in situations of naturally decreasing values due to disease deterioration in its assessment reports for Alzheimer [76] and biologics in RA [69] and considered this approach as “most restrictive”.
 - In the case of particularly relevant endpoints, when IQWiG doubts the robustness of the effect measured, it might itself perform additional analysis based on the aggregated data available [77]. The analyses performed are based on the percentage of responder measured on non-missing values on the control or active arm. This probability of success is then reported on the missing values, the rest being considered non-responder. The effect is re-evaluated using updated responder rates. Variants of the method can be implemented as different scenarios where: the reference responder rate can come from either the active or comparator arm; be applied to both or solely one of the arms while non-responder analysis are applied to the other arm, etc. Censoring of patients in time-to-event analyses: patients with missing values at the beginning of the study can be considered as “censored” [78].
- Observed cases only: IQWiG clearly criticized this approach as resulting in (potentially) biased results which might turn to uninterpretable [79]. This is especially the case when missing values do not derive from a random mechanism (note, unless not specified, IQWiG means missing completely at random). Furthermore, IQWiG interprets a „clearly different rate in dropout in groups to be compared“ as a hint for missing data NOT deriving from a MCAR-process.
- Sensitivity analysis: To control for potential consequences on decision making, IQWiG requires (or performs) sensitivity analyses. Depending on whether the effects are robust between sensitivity analyses and pre-specified analysis, IQWiG might down-grade the level of evidence. Consistent results do not affect the level of evidence granted by the underlying evidence. When an effect is not consistent between approaches, it is always hard to decide which scenario is the most realistic. The data for the given endpoint are then seen as highly biased and inconsistent results lead to downgrading of level of evidence, –i.e. the “probability of added benefit”. (see benefit dossier assessments [11, 80-82].

Critical appraisal and scientific rationale for requirements

Little, D'agostino [83] give a good overview of handling of missing values during analysis. This document just gives a short overview, for details, the reader should refer to the original paper or further relevant literature, e.g. [84-86]. In general, the authors distinguish four approaches for handling of missing values:

1) complete case analysis, 2) single imputation methods, 3) estimation equation methods and 4) methods based on a statistical model. A description and discussion of the above mentioned methods is above the framework of this paper and can be found in the relevant literature (eg [86]). “The properties of these methods depend on the mechanisms leading to missing data” [83], where following mechanisms of missing values can be distinguished: (a) Missing completely at random(MCAR): missing values are neither dependent on observed nor on non-observed data (b)Missing at random (MAR): missing values may depend on data observed, but not on unobserved data; (c) Not missing at random (MNAR): if the missing mechanism is neither MCAR nor MAR, it is called MNAR or informative missing (IM). The main problem with this categorization is the fact that the underlying “pattern” cannot be tested and therefore assumptions are always a point of criticism. However, although “in practice trial data are rarely MCAR, [...], that does not mean that methods valid under MAR are of little use; quite the contrary” [84].

A report from the U.S. National Academy of Sciences research arm, the National Research Council (NRC) [83] gives a very detailed insight into “The Prevention and Treatment of Missing Data in Clinical Trials”. Besides focusing on strategies in prevention of missing values during study conduct, the book gives guidance on drawing inference in incomplete data and introduces principles on sensitivity data as well as giving some recommendations.

The final ICH Concept Paper E9(R1) also describes handling of missing values as one of the factors to be considered in the definition of study estimands for clinical trials. The authors further emphasize the necessity of a “targeted range of thoughtfully constructed sensitivity analyses [that] can help to investigate and understand the robustness of estimates” [87]. Little, D'agostino [83]” give an interpretation aid for the results from these sensitivity analyses: “If the treatment effect [of the sensitivity analyses] is qualitatively maintained for the range of offsets that are considered as clinically plausible, then the findings are considered to be robust “

The methodological framework by IQWiG is also in line with guidance by regulatory agencies. EMA Guideline on Missing Data in Confirmatory Clinical Trials [65] clearly state that “Complete case analysis cannot be recommended as the primary analysis in a confirmatory trial”. Methods for handling of missing values can only be “accepted if it is considered that an important bias in favor of the experimental treatment can be reasonably excluded and if it can be verified that the associated confidence interval does not underestimate the variability of the estimated treatment effect to an important extent. In the event that the proportion of missing data is non-negligible, it is likely that no single method will provide a comprehensive solution to the missing data problem”, clearly supporting the requirement by IQWiG to perform sensitivity analyses to assess robustness of results.

For evaluation of potential bias due to missing of incomplete data on individual study basis, the American Agency for Healthcare Research and Quality (AHRQ) refers to the methods as described in the Cochrane handbook [88]. According to the methodology by AHRQ, information on study bias will then further be incorporated in grade of evidence or meta-analysis.

The AHRQ gives guidance on handling of missing values of individual trials on the meta-analysis level: “Individuals missing from a study due to withdrawal and other reasons create an issue at the study level more than at the meta-analysis level. While missing individuals will also affect the results of meta-analysis, it is very difficult to deal with at the meta-analysis level without access to individual patient data. Nevertheless, three methods have been proposed to account for missing patient data: reweighting by completion rate, incorporation of the completion rate into a Bayesian random-effects model, and inference based on a Bayesian shared-parameter model (including the completion rate). Missing study-level characteristics will not affect the overall meta-analysis but can affect or even prevent subgroup analysis and meta-regression. Bayesian methods have been suggested to account for missing study-level data during meta-regression, but these issues are complex and do not specifically pertain to continuous data. No particular methods are recommended, and investigators may try the methods outlined above for exploratory purposes” [89].

The National Institute for Health and Care Excellence (NICE) uses the GRADE approach for quality assessment of evidence on endpoint level. The grade of recommendation is then considered using the methodology checklist on randomized controlled trials reflecting – besides others – “attrition bias”, a source of bias referring to the loss of participants during the course of a study [90]. The checklist does however do not mention explicit thresholds to when “important or systematic differences” should be considered. NICE does not recommend or propose missing data implementation on study or meta-analysis level.

Conclusion and further recommendations

The assessment of bias in individual trials is fundamental to regulatory agencies and HTA bodies. All bodies taken into consideration in this paper consider missing data as a potential source of bias. None of them however gives clear recommendations on how to handle them. The position adopted by the IQWiG in the assessments referenced in this paper gives some indication regarding their expectations:

- only a reasonable amount of missing values (less than 30% in both arms) is acceptable and does not invalidate the endpoint;
- the proportion of missing values should be balanced between the groups;
- when possible, give a rationale for the presence of missing values, especially when the proportion of missing values is large. It might help get an insight on the nature of the missing data mechanism;
- sensitivity analyses, ideally representing best and worst cases scenarios, should be proactively performed and presented, in order to assess the robustness of the estimates and therefore the impact of missing values on the treatment effect;
- imputation strategies should be realistic and the assumptions made should be compatible with the characteristics of the condition at hand (degenerative, chronic, etc.).
 - Single imputation methods, such as LOCF (last observation carried forward) or BOCF (baseline observation carried forward) not only make possibly inadequate assumptions on disease progress but they also result in an underestimation of standard errors. They will therefore not be sufficient for submission in early benefit dossiers in future.

- Using Individual Patient Data (IPD) increases the power of the study in comparison to artificial recoding on aggregated databases and should be used wherever possible.

Those expectations should be considered when dealing with missing values within the frame of the German benefit assessment.

Indirect Comparisons and Network Meta-Analysis **(Indirekte Vergleiche und Netzwerk-Metaanalysen)**

Authors: Julia Schiffner-Rohe, Sarah Böhme

Abstract

Häufig sind keine direkt vergleichenden („head-to-head“-) Studien für den Beleg für einen Zusatznutzen verfügbar. Selbst im Falle der Verfügbarkeit direkter Evidenz, gibt es Situationen, in denen für einen solchen Vergleich nicht ausreichend Studien vorliegen.

Das fünfte deutsche Sozialgesetzbuch (SGB V) sieht in diesem Fall die Verwendung von indirekten Vergleichen für die frühe Nutzenbewertung vor. Welche Voraussetzungen müssen für die indirekten Vergleiche erfüllt sein um von den Gutachtern der Nutzendossiers der Hersteller akzeptiert zu werden? Welche Erfahrungen haben pharmazeutische Unternehmen gemacht? Und was kann man daraus lernen; einerseits für die Implementierung von indirekten Vergleichen in Nutzendossiers, andererseits für die zukünftige Studienplanung?

Topic

The choice of appropriate comparator therapy is fundamental for the design of the early benefit dossier. Experience has demonstrated that dossiers using another ZVT than the one determined by the G-BA will not be assessed and therefore will not be able to demonstrate an added benefit of the new drug. Therefore it is essential to use the ZVT as defined by the G-BA for the (early) benefit assessment.

However, in many cases choice of ZVT by G-BA does not reflect the comparator for the marketing authorization (MA)-studies of the new substance to be launched – in general or in subpopulations as defined by the G-BA for assessment of added benefit.

Typical situations, where this occurs are:

- Study on new substance is designed versus Placebo
- ZVT represents an “old-generation” drug with much evidence available, whereas MA-studies use “new generation” product as comparator
- ZVT was not launched at the time of study design/conduct (e.g. benefit assessment Axitinib (Inlyta®, Pfizer Pharma GmbH, renal cell carcinoma) [91])
- Comparator for MA-studies is registered/launched in the US-market, but not in Europe/Germany (i.e. cannot be defined as ZVT for benefit assessment)
- Comparator is registered for different treatment regimens (e.g. dose, treatment duration) in the US and Europe
- Comparator was chosen after discussions with MA-institutions (FDA, EMA) and does not clearly reflect current guidance (e.g. benefit dossier microbial collagenase clostridium histolyticum (Xiapex®, Pfizer Pharma GmbH, Dupuytren’s contracture) [92])

Basis for reimbursement price negotiations and subsequently reimbursed price of new substances are the extent (major, considerable, minor, not quantifiable) and probability (proof, indication, hint) of added benefit in comparison to ZVT. Added benefit can only be *proven* by at least two (or alternatively one giant) randomized active controlled trials (RCTs) to ZVT of high quality. Still, the extent of added benefit for the new substance in comparison to ZVT can be estimated by using indirect comparisons (IC). The term “indirect comparison” can be understood as both, a simple indirect comparison of two interventions as well as the combination of direct and indirect evidence [27]. Techniques combining direct and indirect evidence are often also referred to as “network meta-analysis” [93] or “mixed treatment comparisons” [94].

However, ICs bear a high potential of bias. This is reflected in the assessment of probability of added benefit derived based on indirect comparisons. Adjusted indirect comparisons can reach an *indication* at maximum, as they have a lower certainty of results than network meta-analyses [95] (see *Derivation of “probability of added benefit” from meta-analyses*) and direct comparisons. In case of demonstrating dramatic effects by the use of unadjusted indirect comparisons a *hint* can be reached as utmost of probability [81] (see *Unadjusted indirect comparisons*).

Requirements by IQWiG and G-BA

Benefit dossiers as submitted by the manufacturer can be assessed by IQWiG or, in case of orphan drugs, by G-BA and IQWiG. For orphan drugs, which do not exceed a yearly turnover of 50 million euros, the extent of added benefit is regularly assessed by G-BA whereas information provided by the pharmaceutical companies on the number of affected patients and the cost of treatment are assessed by IQWiG. In general, the requirements are the same although far more difficult to fulfill for orphan drugs¹. The following describes the requirements according to the methods paper by IQWiG [27].

General requirements

Even if it may seem trivial, it is noteworthy mentioning that a general requirement for the acceptance of indirect comparisons and network meta-analysis is the fact that the comparison clearly addresses the research question (defined by the PICO criteria). This includes (but not exclusively) using evidence on the population of interest, interventions (including dosing, treatment duration etc.) of interest, all patient-relevant outcomes. “Of interest” is defined as required by the G-BA and might differ in subgroups.

A fundamental requirement is an absolutely transparent and comprehensive description of the approach – including rationale for the choice of model, choice of assumptions and technical details. If this description is not provided or not comprehensive enough to potentially reproduce results, this will inevitably lead to a rejection of the method and results.

¹ Orphan drug in the context of benefit dossier in Germany is defined as market authorization as orphan drug according EC-legislation No. 141/2000 of the European Parliaments and a turnover of sales not exceeding 50 Mio Euro within one calendar year (SGB Z, §35a (8))

Requirements to (conventional) meta-analyses:

To derive in valid estimates, the following requirements on indirect comparisons and network meta-analyses need to be fulfilled:

- Study pool must derive from systematic review on available literature. Databases to be used as default are: MEDLINE, EMBASE databases and Cochrane database. Optionally, a search can be conducted in additional subject-specific databases (e.g., CINAHL, PsycINFO, etc.).
- Similarity:
Assessment of similarity of underlying studies should be based with regards to content. Criteria include PICO, but should also address criteria such as study design or setting. Sole analysis of covariates is not sufficient, although it can supplementary be performed in certain situations (such as sensitivity analyses). Analyses of covariates do strongly depend on the choice of covariables assessed and included in the model and therefore underlie a preselection which needs to be scientifically justified.
- Choice of model to be used for network meta-analysis:
If information is available that the effects of the individual studies are homogeneous, a meta-analysis assuming a fixed effect is sufficient. However, such information will often not be available, so that in order to evaluate studies in their totality, an assumption of random effects might be useful. Moreover, it should be noted that the confidence intervals calculated from a fixed-effect model may show a substantially lower coverage probability with regard to the expected overall effect, even if minor heterogeneity exists when compared with confidence intervals from a random-effects model. IQWiG therefore primarily uses random-effects models [27]. According to the draft version of IQWiG General Methods 5.0 the approach proposed by Knapp-Hartung should be used [10].²
In its draft version for General Methods 5.0 IQWiG added a section on meta-analyses in case of only a few studies (< 5). In this situation random-effect models are not recommended, but rather fixed-effect models or qualitative evaluations should be used. Depending on the context, alternative approaches (e.g. Bayesian methods or methods from the field of generalized linear models) may also be considered.

² At time of finalization of this manuscript only a draft version was available to the authors.

- Heterogeneity:

Heterogeneity needs to be statistically tested within the benefit dossier. Due to the low statistical power of tests for heterogeneity, tests are to be performed on a significance level of 0,1 to 0,2. Furthermore, the extent of heterogeneity needs to be quantified. IQWiG favorably uses I^2 -measure and categorizes heterogeneity into “might not be important” (I^2 : 0 to 40%), “moderate” (I^2 : 30 to 60 %), “substantial” (I^2 : 50 - 90 %) and “considerable” (I^2 : 75 to 100 %).

If heterogeneity is detected, this needs to be analyzed in more detail. Methods to be used are meta-analyses on subgroups as well as meta-regression.

In case heterogeneity is too high – even after reduction to subgroups -, study data should not be statistically pooled. Although the definition of “high heterogeneity” depends on the context, it is recommended to refrain from a meta-analysis if the test on heterogeneity is significant at a level of 0,2. According to the draft version of IQWiG General Methods 5.0 this significant level will be reduced to 0,05 [10].

Nevertheless, if single studies show a clear and rectified effect, joint effects deriving from a random effect model may lead to a conclusion of added benefit. For details, see **“Fehler! Verweisquelle konnte nicht gefunden werden.”**

- Consistency:

Sufficient consistency is required in the effects estimated from direct and indirect evidence. Consistency cannot be examined in case of simple indirect comparisons where no direct evidence is available. Although fundamental for indirect comparisons to lead to valid results, there are currently no accepted techniques to check consistency. Therefore, it is necessary within the benefit dossier to describe the methods applied in a thorough and transparent manner.

- Presentation of results from meta-analyses: prediction interval

Results of meta-analysis need to be presented using forest plots.

To present effect estimates from meta-analyses with random effects, IQWiG has developed prediction intervals [96]. The localization of the prediction interval is basis for the presence of “rectified effect” and as a consequence basis for the probability of added benefit. Prediction intervals are critically discussed for meta-analyses with less than four studies. IQWiG has therefore defined following procedure to follow for benefit dossiers [27]:

Table 1: Decision procedure for rectified effects in early benefit assessments with small number of studies (adopted from [27])

| No of studies | Prediction interval (PI) | criteria | direction |
|---------------|---|---|----------------------|
| >= 4 | PI does NOT cover the zero effect | -/- | Clearly rectified |
| | PI not presented / covers the zero effect | Effect estimates of >=2 studies point into same direction and for these the following applies: overall weight of studies is >= 80% AND >=2 studies show significant results AND >= 50% of study weight* derives from statistically significant results | Moderately rectified |
| 3 | n.a. | All studies show significant results | Clearly rectified |
| | | two studies show significant results | Moderately rectified |
| 2 | n.a. | Both studies show significant results | Clearly rectified |

*inverse variance from of meta-analysis, sample size otherwise

- Derivation of “probability of added benefit” from meta-analyses:
Depending on the quality of evidence, the number of underlying studies and the heterogeneity of results, IQWiG derives the following probability of added benefit [27]:

Table 2: Decision procedure for certainty of evidence in early benefit assessments (adopted from [27])

| | | No of studies | | | | |
|-----------------------------------|----------|----------------------------------|---|-----------------------|----------------------|---------------|
| | | 1 | >= 2 | | | |
| | | | Homogeneous results | Heterogenous results* | | |
| | | Statistically significant effect | Meta-analysis statistically significant | Clearly rectified | Moderately rectified | Not rectified |
| Qualitative certainty of evidence | High | Indication | Proof | Proof | Indication | -/- |
| | Moderate | Hint | Indication | Indication | Hint | -/- |
| | Low | -/- | Hint | Hint | -/- | -/- |

*assuming the underlying studies all show the same quality of evidence

Available methods for indirect comparisons:

Indirect comparisons (IC) can derive from the frequentist or Bayesian methodology. The remainder of this paper will use IC, independent of the methodology used.

IQWiG clearly requires adjusted indirect comparisons; unadjusted indirect comparisons are in general not accepted as they do not consider randomization within the trials [27, 95]). The only exception is the presence of “dramatic effects” where findings clearly can be rejected to be result of systematic bias (see Unadjusted indirect comparisons).

1. Adjusted indirect comparisons

IQWiG favours the use of indirect comparisons by the Bucher-method [97]. An accepted way to integrate direct (H2H) and indirect evidence is the use of network meta-analysis (mixed treatment comparisons, MTC) [94].

The utmost “probability” of added benefit to be reached by adjusted indirect comparisons is an *indication*.

Besides the usual requirements for meta-analysis (for example see [98]), the requirement of a connected network has to be fulfilled to gain validity of estimates on added benefit from adjusted indirect comparisons. If the network is not connected or the bridging comparator is not similar enough to be used as bridge (violation of similarity assumption), only unadjusted indirect comparisons are feasible.

2. Unadjusted indirect comparisons

As a rule, unadjusted indirect comparisons are not accepted for estimation of added benefit. Unadjusted ICs break randomization and only provide a descriptive contrasting of study results. This is independent of whether results are purely contrasted narratively or efforts were taken to make results more similar (e.g. using matching adjusted indirect comparison (MAIC) described by Signorovitch [99] and Parks [100] or simulated treatment comparison (STC) [101]).

Nevertheless, if adjusted ICs are not possible, several applicants of benefit dossiers have tried to submit unadjusted indirect comparisons.

Although this approach formally could be accepted (due to the high potential of bias, unadjusted ICs can reach a *hint* of added benefit as utmost of probability [81], only one applicant was successful with his approach because he was able to demonstrate a dramatic effect. All other unadjusted ICs were rejected.

An effect is considered to be dramatic, if the hypothesis of equal effects could be rejected at a 1%-level and the estimated relative risk is above 10 (or below 0,1). A 10-fold increase (resp. reduction) of risk often reflects a (quasi-)deterministic disease progression. Additionally, the risk for the event of interest needs to be at least 5% in one of the comparator groups to reflect the relevance of the event on population level [27].

Even if an applicant might be able to demonstrate a dramatic effect derived from unadjusted IC it does not necessarily imply that the unadjusted IC will be accepted. Furthermore at least the following requirements need to be fulfilled:

- Underlying studies (on the new substance) are adequate in general to answer the research question (e.g. correct population, intervention, outcomes, study duration)
- Sufficient data available on the appropriate comparator therapy (ZVT)

Appraisal and scientific rationale for requirements

Indirect treatment comparisons are known to be sensitive to bias. Therefore, potential sources of bias have to be assessed and discussed prior to performance of IC and finally the interpretation of results (see e.g. ISPOR 2014 [102]). As a consequence, IQWiG and G-BA have defined criteria when IC might be accepted for estimation of added benefit but prior to interpretation clearly assess whether requirements for a valid interpretation are fulfilled. An assessment of all indirect comparisons assessed by IQWiG until March 2017 demonstrated that there are several requirements that have to be fulfilled for valid indirect comparison (and meta-analyses in general). A formal checklist for the evaluation of indirect comparisons and network meta-analyses can be found in Kiefer et al. [103].

From 185 early benefit dossiers assessed by IQWiG, 58 (31%) substances were submitted using indirect comparisons to demonstrate added benefit (*As at March, 2017*). Only four adjusted and one unadjusted indirect comparisons were assessed as valid to estimate the added benefit of the launched substance. The unadjusted indirect comparison (historic comparison) could demonstrate a dramatic effect and was therefore accepted by IQWiG. According to the assessment appraisals as well as Kiefer et al. [103], following aspects are part of the assessment process for an adjusted indirect comparison:

- Study populations:
 - o Is the study population within the market authorization (MA) (e.g. patients excluded from MA may not be used for the analyses)?
 - o Does the study population cover the research question (e.g. subpopulations for different ZVTs: the study population may only cover evidence on the subpopulation)?
 - o Are the study populations comparable between trials (within treatments (test and ZVT) and between treatments (test trials versus ZVT trials))?

- Is the study population for trials on test and ZVT comparable in baseline characteristics (e.g. disease severity, treatment line, previous treatment)?
- Treatments:
 - Are treatments (test and ZVT) applied in the dosage and dosing scheme (e.g. frequency) according to the MA?
- Study design:
 - Is the trial duration comparable and adequate in trials for test and ZVT?
 - Were concomitant medication / background medication comparable in trials for test and ZVT?
 - Were trials other than RCTs used for the indirect comparison?
- Common comparator ("bridging substance"):
 - Is the bridging substance comparable in trials for test and ZVT (e.g. completely inactive in both settings)? *(Noteworthy: it seems to play no role whether the bridging comparator is licensed/used within the licensed dose)*
- Study pool:
 - Were all relevant (at least the 3 required) databases used for the literature research?
 - Was an appropriate literature research performed to find all relevant evidence?
 - Was the literature research performed adequately and according to the IQWiG methodology?
 - In case a group of ZVTs (e.g. group of β -Interferon- α) is defined: Was all evidence on the ZVT (and not only one dosage, application form etc.) used for the indirect comparison?
- Methodology of indirect comparison:
 - Was a random effect model used? Or alternatively: was the use of a fixed effect model adequately verified by demonstrating lack of heterogeneity?
 - Was heterogeneity of effect measures between trials tested and discussed appropriately?
 - In case of MTC: were similarity, consistency and heterogeneity addressed and adequately discussed?
 - Was the analysis performed for EVERY patient-relevant endpoint (i.e. avoidance of reporting bias)?
 - Were all data correctly transferred from the publications to the meta-analysis software?
- Presentation of methodology and results
 - Is the methodology clearly described?
 - Are the assumptions for the choice of model evidence based or otherwise clearly described in a scientifically rational manner?
 - Are assumptions (including technical details, e.g. prior distributions in Bayes approaches) justified?
 - Are the results presented in a transparent manner and potential sources of bias discussed adequately?

Recommendation

- All efforts should be done to use the appropriate comparator (ZVT) as defined from G-BA at least in one of the registration trials. Proof of added benefit can only be gained from two direct head to head comparisons of low potential for bias between the new substance and ZVT as defined by the G-BA.
- If there is a decision **not** to use the ZVT (as required by G-BA) in a Phase III study as comparator, but to launch the drug in Germany, the study design should make IC possible.
This means the study design should adapt to the design of the comparator trials (e.g. trial duration, definition of endpoints, background medication, dosing of treatments to allow for bridging, allowance of rescue medication).
In the event a new substance gets marketing authorization in the timeframe of the study planning and may become a potential ZVT for the Pfizer substance in the benefit assessment, it should not only be considered to use the ZVT - as defined for the benefit assessment of the comparator - for the H2H-design, but also design the studies according to those by the comparator in case the G-BA changes ZVT.
- Studies to be used for the benefit dossier must cover the population from market authorization not only for the new substance but also for studies on ZVT. IQWiG has defined an 80%-rule, implicating that at least 80% of the (described) population must meet the MA. This needs to be considered not only when planning the phase 3 trials (where subgroup analyses are possible), but also for the available data on studies on ZVT. If no published data are available for the relevant population on ZVT, this should be reflected in the decision on study comparator.
- Studies to be used for the benefit dossier must reflect the dose regime from market authorization - not only for the new substance but also for studies on ZVT.
- Refrain from uncontrolled (e.g. single arm) studies if possible: if adjustment is not feasible, unadjusted indirect comparisons will be rejected unless dramatic effect. Approaches such as simulated treatment comparison (STC) or propensity score matching (PMS) based approaches were not accepted to date and will not be accepted unless clear demonstration of dramatic effects (see above for definition according to IQWiG).
- Only one historic comparison has been accepted to date (as at March 2017) due to clear demonstration of a dramatic effect.

Health economic evaluation of pharmaceuticals (Kosten-Nutzen-Bewertung von Arzneimitteln)

Author: Fabian Volz

Abstract

In dieser Publikation werden die Voraussetzungen und Anforderungen einer Kosten-Nutzen-Bewertung von Arzneimitteln in Deutschland dargestellt, insbesondere in Bezug auf die Nutzenbewertung im Rahmen des Arzneimittelmarktneuordnungsgesetzes (AMNOG).

Topic

Currently health economic evaluation of pharmaceuticals is common and often used to support reimbursement decisions in countries worldwide. In the framework of the AMNOG process it is possible in only two cases:

First in case of G-BA is deciding that a new launched intervention has “no added benefit” compared to the ZVT (6 months after initiating the benefit assessment) the pharmaceutical company may file an application for a health economic evaluation with the G-BA.

In the second case both the pharmaceutical company and the GKV-SV can apply for a health economic evaluation after the ruling of the arbitration board (15 months after initiating the benefit assessment).

The aim of the health economic evaluation is to determine an adequate reimbursement price for the concerning pharmaceutical, but caveat that this determined price will be just an orientation for pricing and is not definite. If the pharmaceutical company requests a health economic evaluation in case of no added benefit, it has to bear all costs [104].

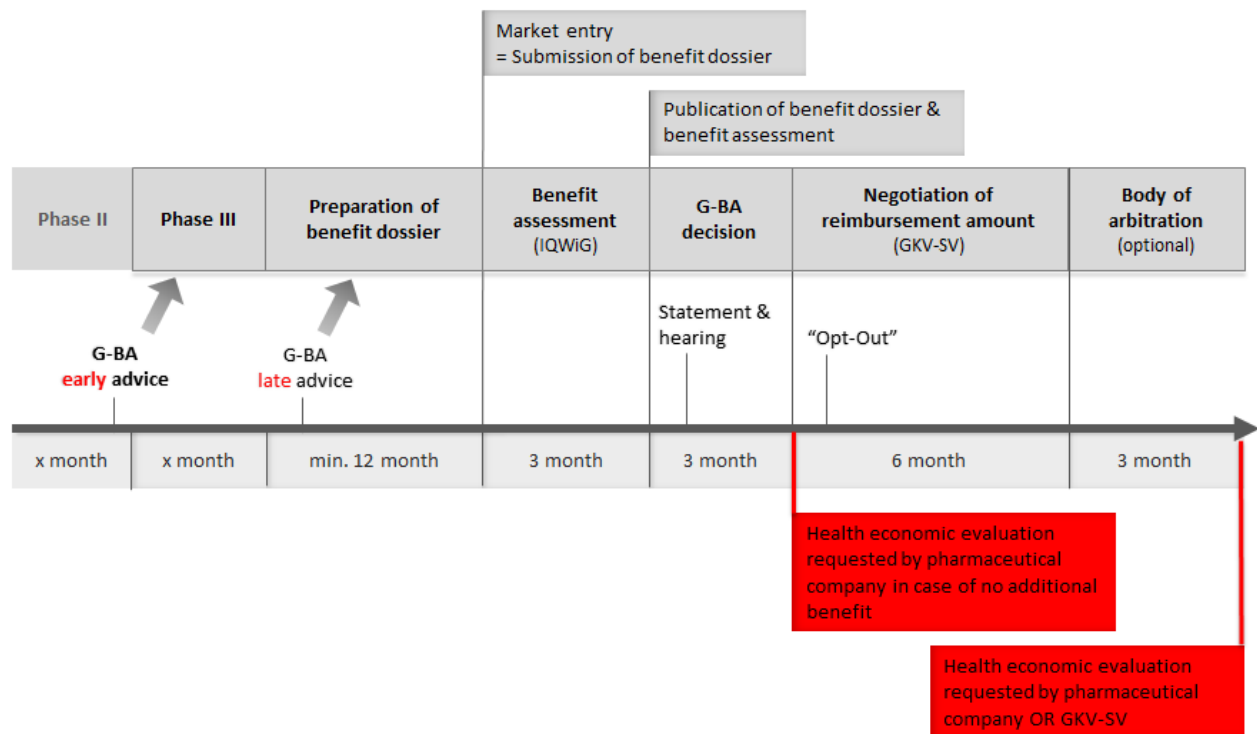
Before a health economic evaluation will be conducted a scoping workshop is held and relevant elements of the evaluation are defined like the comparator, patient (sub-)population(s), the perspective (it should regularly be the Statutory Health Insurance (SHI) insurees’ perspective), the time horizon, kind of benefits and costs to consider and the measure of overall benefit. Afterwards the G-BA requests the pharmaceutical company to submit a dossier containing the health economic evaluation within 3 months. The pharmaceutical company has to comply with the dossier template by the G-BA, which is based on the IQWiG’s general methods [27]. The basis of the health economic evaluation is clinical studies and real-world data studies agreed with the G-BA or approved by the G-BA on request [104]. Also piggyback-studies are possible, but only when these health economic studies match exactly the requirements determined by the G-BA and represent the real life situation of health care in Germany [105]. Health economic models are the central of a health economic evaluation and are used to extrapolate the benefit and costs beyond the given study duration [27].

In case of an incomplete or not sufficient dossier submitted by the pharmaceutical company, the IQWiG will undertake the economic evaluation itself. There is neither an obligation for the G-BA nor the IQWiG

to consider a delayed dossier or any delayed additional information. The IQWiG prepares a preliminary assessment within 12 months and after a formal hearing, in which the pharmaceutical company can submit his written and oral statement, the final assessment is sent to G-BA. As of February 2017 no health economic evaluation has been commissioned by G-BA and undertaken by IQWiG.

In conclusion the impact of the health economic evaluation for statutory decision making in Germany is actually non-existing at this time.

Figure 2: Process of health economic evaluation according to AMNOG (source: own figure)



Requirements by G-BA and IQWiG

The following core points of a health economic evaluation are listed in the template for the health economic evaluation by the G-BA and are mandatory for the pharmaceutical company [105]:

- Indication
- Patient (sub-)population
- Pharmaceutical for the health economic evaluation
- Zweckmäßige Vergleichstherapie (ZVT, appropriate comparative therapy)
- Study perspective
- Time horizon
- Benefit (endpoints)
- Measure of overall benefit

- Costs
- Approach for the health economic evaluation
- Base year for the health economic evaluation

In the following the stated core points of a health economic evaluation will be specified and their requirements analyzed and discussed.

Indication:

Indication has to be equal according the previous conducted benefit assessment.

Patient population/groups:

Patient population and subpopulations have to be equal according to the previous conducted benefit assessment.

Study perspective:

Demanded is the perspective of the SHI plus co-payments of patients as well, but no indirect costs like productivity loss. Other perspectives could be commissioned by the G-BA, e.g. the societal perspective with additional costs to be considered.

Comparator:

All approved pharmaceutical and non-pharmaceutical products which are reimbursable by the SHI with the same therapeutic indication could be potential comparators. The G-BA determines the relevant comparators and these could be more than in the previous conducted benefit assessment. According the general methods by IQWiG all health care relevant comparators in the concerning indication should be incorporated [27].

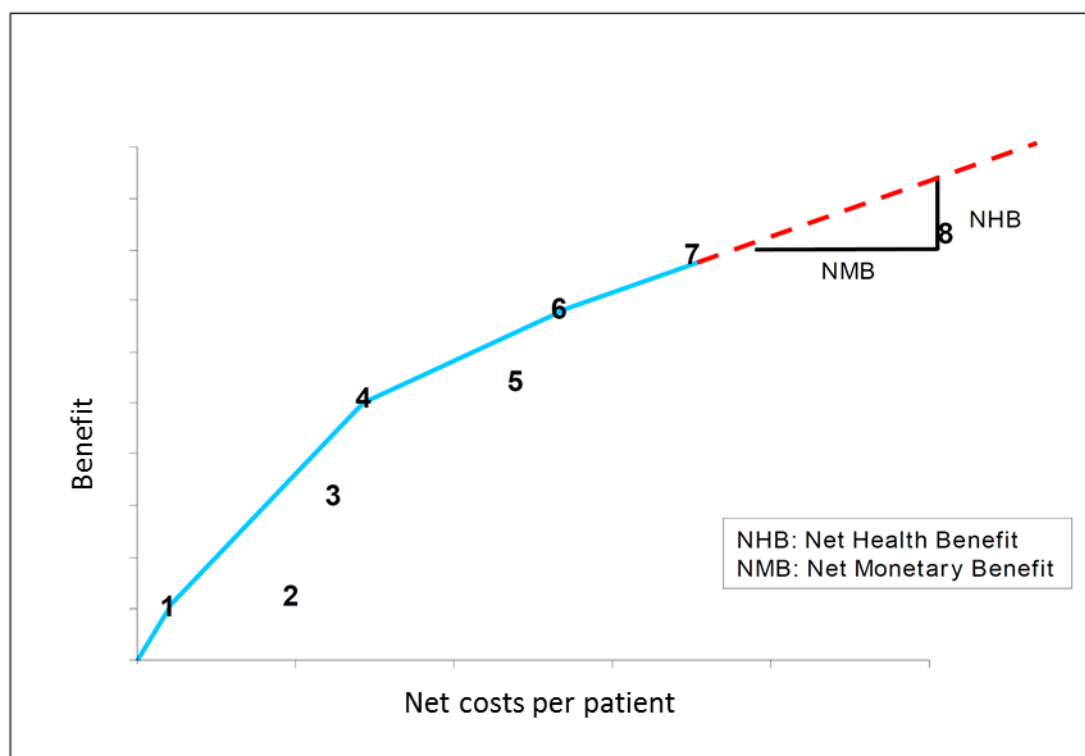
Analysis technique/approach:

Only a cost-effectiveness analysis with the efficiency frontier approach will be accepted. For a detailed guidance of the construction of the efficiency frontier see IQWiG's general methods [27].

The analysis of efficiency frontier has to be conducted separately by scenario and patient (sub-) population. For every patient-relevant endpoint considered in the health economic evaluation a separate efficiency frontier has to be constructed (detailed guidance in IQWiG's general methods). Afterwards all results for every single endpoint will be combined in one measure of overall benefit for every single patient population. This measure for the overall benefit will be the same for every scenario. The efficiency frontier is composed by the other therapeutic options in the indication. On basis of the constructed efficiency frontier an added benefit adjusted price for the evaluated pharmaceutical will be calculated. This price has to be calculated for every considered scenario and for every patient (sub-)population. The threshold for this price is the last segment of the efficiency frontier or the extrapolation of the last segment, but caveat this threshold is just a piece of information for the decision maker how much the system is willing to pay for an added benefit.

The following figure shows an example of the efficiency frontier:

Figure 3: Efficiency frontier (source: [105])



In this fictive example the number 8 is the evaluated pharmaceutical and has a higher benefit and higher net costs per patient compared to all other interventions in this indication. The added benefit adjusted price could be calculated with the help of the net monetary benefit and the net health benefit respectively [105].

Time horizon:

The time horizon will be determined by the G-BA on the nature of disease. The horizon could be equal to the study duration or even life-long depending on nature of disease.

Benefit (endpoints):

The patient relevant endpoints will be determined by the G-BA as well. The selected endpoints will be the same like in the previous conducted benefit assessment or even less. For every comparator the benefit for each endpoint has to be assessed. The endpoints and effect estimates has to be operationalized from the benefit assessment. A transformation of the effect measures might be necessary to ensure requirement of the approximate cardinal scaling of the benefit axis [105].

Measure of overall benefit:

The measure of overall benefit depends on the results of the benefit assessment. Either the measure or the approach how to construct the measure of overall benefit is be determined by the G-BA. The use of

Quality-Adjusted Life Years (QALY) is conceivable, but keep in mind that the acceptance in Germany is limited. A deviation from this decision is not possible [105].

For the presentation of results the incremental cost effectiveness ratio (ICER) for the measure of overall benefit has to be calculated [105].

Costing methods:

Two options for the determination of costs are possible: event-based or performance-based depending on the health economic model, e.g. event-based in a markov-model. Preferred are the incurred costs per year. Essential is the presentation of the yearly therapy costs of the treatment with the all in the health economic evaluation included treatment options (both the evaluated pharmaceutical and all comparators) [105].

In the basis perspective, i.e. the perspective of the SHI insured persons, all co-payments of the insurants have to be considered in the evaluation. Outpatient costs should orientate on the actual doctor's fee scale (Einheitlicher Bewertungsmaßstab, EBM) and inpatient costs on diagnosis related groups (DRG). An adjustment on inflation has to be conducted at a basis year. For the construction of the efficiency frontier the net costs per patients should be used, i.e. cost-offsets adjusted costs [105].

In addition a budget-impact-analysis has to be conducted. The perspective has to be the perspective of the payer, i.e. the SHI. This means no co-payments by the insured persons will be considered, but all costs has to be included which the SHI has to bear, e.g. transfer payments. Time horizon has to be one and three years. Two scenarios have to be presented:

- Reference scenario: actual combination of health technologies (without the evaluated pharmaceutical)
- Prognosis scenario: predicted combination of health technologies [105]

Modelling techniques:

The choice of the modelling technique (influence diagram and model concept) has to be explained. Usually decision trees, markov models or discrete event simulation will be used. Other methods like agent based simulation could be used, but rarely in practice. The model concepts have to be internally and externally validated on the basis of literature and clinical experts [105].

Discounting:

If the time horizon is more than one year the costs and benefits has to be discounted. The discounting has to be undertaken with 5% for costs and benefits, which is a deviation from the IQWiG's methods (3%) [27]. In sensitivity analyses the discount rates should be between 0 and 10 % [105].

Sensitivity analysis:

Usually there are two kinds of uncertainty: There is uncertainty because of variability in certain parameters or uncertainty because of assumptions in the model. Therefore deterministic sensitivity analyses and probabilistic sensitivity analyses (monte carlo simulation) have to be conducted [105].

Implications

The benefit of a health economic evaluation for the pharmaceutical industry is highly questionable by this point, because of the strict specification by the G-BA in advance and the short time frame for the conduct of the health economic evaluation particularly regard to the vast extent. In addition the determined price is not binding and should be seen as an additional piece of information for the pricing. Keep in mind that no health economic evaluation of pharmaceuticals in the context of AMNOG has been conducted so far.

References

1. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Axitinib [online]. 21.03.2013 [last accessed: 21.02.2017]. URL: https://www.g-ba.de/downloads/40-268-2240/2013-03-21_AM-RL-XII_Axitinib_TrG.pdf.
2. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Crizotinib [online]. 02.05.2013 [last accessed: 21.02.2017]. URL: https://www.g-ba.de/downloads/40-268-2301/2013-05-02_AM-RL-XII_Crizotinib_TrG.pdf.
3. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Crizotinib [online]. 15.12.2016 [last accessed: 28.02.2017]. URL: https://www.g-ba.de/downloads/40-268-4106/2016-12-15_AM-RL-XII_Crizotinib_D-240_TrG.pdf.
4. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Vandetanib [online]. 06.09.2012 [last accessed: 21.02.2017]. URL: https://www.g-ba.de/downloads/40-268-2039/2012-09-06_AM-RL-XII_Vandetanib_TrG.pdf.
5. European Medicines Agency. Assessment report Gilenya; Procedure No. EMEA/H/C/2202; Doc.Ref.: EMA/108602/2011 [online]. 17.02.2011 [last accessed: 21.02.2017]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002202/WC500104529.pdf.
6. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Fingolimod [online]. 29.03.2012 [last accessed: 24.02.2017]. URL: https://www.g-ba.de/downloads/40-268-1908/2012-29-03_AM-RL-XII_Fingolimod_TrG.pdf.
7. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses; zuletzt geändert am 20. Oktober 2016 [online]. 24.01.2017 [last accessed: 24.02.2017]. URL: https://www.g-ba.de/downloads/62-492-1331/VerfO_2016-10-20_iK-2017-01-20.pdf.
8. Gemeinsamer Bundesausschuss. Recherche und Synopse der Evidenz zur Bestimmung der zweckmäßigen Vergleichstherapie nach § 35a SGB V - Vorgang: 2016-B-059 Cabozantinib; Abteilung Fachberatung Medizin [online]. 07.06.2016 [last accessed: 24.02.2017]. URL: https://www.g-ba.de/downloads/92-975-1732/2016-11-01_Informationen-zVT_Cabozantinib_D-263.pdf.
9. Gemeinsamer Bundesausschuss. Kriterien zur Bestimmung der zweckmäßigen Vergleichstherapie und Recherche und Synopse der Evidenz zur Bestimmung der zweckmäßigen Vergleichstherapie nach § 35a SGB V - Vorgang: Afibercept [online]. April 2015 [last accessed: 24.02.2017]. URL: https://www.g-ba.de/downloads/92-975-1242/2016-02-26_Informationen-zVT_Afibercept-D-197.pdf.
10. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Entwurf für Version 5.0; [online]. 07.12.2016 [last accessed: 09.12.2016]. URL: https://www.iqwig.de/download/IQWiG_Methoden_Entwurf-fuer-Version-5_final.pdf.
11. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ruxolitinib – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A15-13 [online]. 13.07.2015 [last accessed:

- 24.02.2017]. (IQWiG-Berichte – Nr. 314). URL: https://www.g-ba.de/downloads/92-975-833/2015-07-13_A15-13_Ruxolitinib_Nutzenbewertung-35a-SGB-V.pdf.
12. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Ruxolitinib (neues Anwendungsgebiet) [online]. 15.10.2015 [last accessed: 24.02.2017]. URL: https://www.g-ba.de/downloads/39-261-2356/2015-10-15_AM-RL-XII_Ruxolitinib-nAWG_2015-04-15-D-161_BAnz.pdf.
 13. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Importance of approval status for the benefit assessment [online]. 21.06.2013 [last accessed: 24.02.2017]. URL: <https://www.iqwig.de/en/events/iqwig-in-dialogue/iqwig-in-dialogue-2013.3367.html>.
 14. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Pomalidomid [online]. 17.03.2016 [last accessed: 28.02.2017]. URL: https://www.g-ba.de/downloads/40-268-3670/2016-03-17_AM-RL-XII_Pomalidomid_D-193_TrG.pdf.
 15. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Ramucirumab [online]. 20.10.2016 [last accessed: 28.02.2017]. URL: https://www.g-ba.de/downloads/40-268-4022/2016-10-20_AM-RL-XII_Ramucirumab_Magenkarzinom_D-224_TrG.pdf.
 16. Walters SJ. Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation: A Practical Guide to Analysis and Interpretation. Chichester: John Wiley & Sons, Ltd; 2009.
 17. Cappelleri JC, Zou KH, Bushmakina AG, Alvir J, Alemayehu D, Symonds T. Patient-Reported Outcomes: Measurement, Implementation and Interpretation. Boca Raton: Taylor & Francis Group; 2014. 1-331 p.
 18. Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims [online]. December 2009 [last accessed: 24.02.2017]. URL: <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>.
 19. European Medicines Agency. Reflection Paper On The Regulatory Guidance For The Use Of Health-related Quality Of Life (HRQL) Measures In The Evaluation Of Medicinal Products; Doc. Ref. EMEA/CHMP/EWP/139391/2004 [online]. 27.07.2005 [last accessed: 24.02.2017]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC50003637.pdf.
 20. Oncology Working Party. Reflection Paper on the use of Patient Reported Outcome 4 (PRO) measures in oncology studies. London: European Medicine Agency, 2014 EMA/CHMP/292464/2014.
 21. European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: The use of patient-reported outcome (PRO) measures in oncology studies; EMA/CHMP/292464/2014 [online]. 01.04.2016 [last accessed: 24.02.2017]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2016/04/WC500205159.pdf.
 22. International Society for Pharmacoeconomics and Outcomes Research. ISPOR Patient Reported Outcomes and Quality of Life Initiatives 2015. URL: http://www.ispor.org/research_initiatives/qol_initiatives.asp.
 23. International Society for Quality of Life Research. Health-Related Quality of Life Research Milwaukee 2015. URL: <http://www.isoqol.org/>.

24. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung - AMNutzenV), (2010).
25. Ausschuss für Gesundheit. Hecken lobt Arzneimittelgesetz 19.03.2014/2014. URL: https://www.bundestag.de/presse/hib/2014_03/01/262758.
26. Staack F. GBA-Chef will Vorfahrt für Lebensqualität [online]. Ärzte Zeitung; 03.06.2016 [last accessed: 24.02.2017]. URL: http://www.aerztezeitung.de/politik_gesellschaft/arzneimittelpolitik/nutzenbewertung/article/912741/nutzenbewertung-gba-chef-will-vorfahrt-lebensqualitaet.html.
27. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 4.2; [online]. 22.04.2015 [last accessed: 24.02.2017]. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4-2.pdf.
28. Klakow-Franck R. Die Bedeutung von Lebensqualität für die Arbeit des Gemeinsamen Bundesausschusses. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen. 2014;2-3:151-6.
29. Leverkus F. Methodik für die Auswertung von Lebensqualitätsdaten im Rahmen der Nutzenbewertung von Arzneimitteln. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen. 2014;108:111-9.
30. Duru G, Auray JP, Beresniak A, Lamure M, Paine A, Nicoloyannis N. Limitations of the methods used for calculating quality-adjusted life-year values. Pharmacoeconomics. 2002;20(7):463-73.
31. Gemeinsamer Bundesausschuss. Nutzenbewertungsverfahren zum Wirkstoff Macitentan [online]. 2014 [last accessed: 04.04.2017]. URL: <https://www.g-ba.de/informationen/nutzenbewertung/101/#tab/nutzenbewertung>.
32. Gemeinsamer Bundesausschuss. Nutzenbewertungsverfahren zum Wirkstoff Pertuzumab [online]. 2013 [last accessed: 04.04.2017]. URL: <https://www.g-ba.de/informationen/nutzenbewertung/65/#tab/nutzenbewertung>.
33. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Nintedanib [online]. 18.06.2015 [last accessed: 04.04.2017]. URL: https://www.g-ba.de/downloads/40-268-3248/2015-06-18_AM-RL-XII_Nintedanib_2015-01-01-D-147_TrG.pdf.
34. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Afatinib – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A15-17 [online]. 13.08.2015 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 318). URL: https://www.g-ba.de/downloads/92-975-853/2015-08-13_A15-17_Afatinib_Nutzenbewertung-35a-SGB-V.pdf.
35. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Regorafenib [online]. 17.03.2016 [last accessed: 24.02.2017]. URL: https://www.g-ba.de/downloads/40-268-3667/2016-03-17_AM-RL-XII_Regorafenib_2015-10-01-D-189_TrG.pdf.
36. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Ramucirumab (neues Anwendungsgebiet) [online]. 01.09.2016 [last accessed: 24.02.2017]. URL: https://www.g-ba.de/downloads/40-268-3954/2016-09-01_AM-RL-XII_Ramucirumab_D-216_TrG.pdf.
37. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1998;16(1):139-44.

38. Gjersing L, Caplehorn J, Clausen T. Cross-cultural adaptation of research instruments: language, setting, time and statistical considerations. *BMC medical research methodology*. 2010;10(13).
39. Molenberghs G, Burzykowski T, Alonso A, Buyse M. A perspective on surrogate endpoints in controlled clinical trials. *Statistical methods in medical research*. 2004;13(3):177-206.
40. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*. 2001;69(3):89-95.
41. Miksad RA, Zietemann V, Gothe R, Schwarzer R, Conrads-Frank A, Schnell-Inderst P, et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. *International journal of technology assessment in health care*. 2008;24(4):371-83.
42. Molenberghs G, Buyse M, Burzykowski T. Introduction. In: Burzykowski T, Molenberghs G, Buyse M, editors. *The Evaluation of Surrogate Endpoints*. New York, NY: Springer; 2005. p. 1-5.
43. Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. *Statistical methods in medical research*. 2008;17(5):477-86.
44. Fleming TR, DeMets DL. Surrogate End Points in Clinical Trials: Are We Being Misled? *Annals of Internal Medicine*. 1996;125(7):605-13.
45. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical statistics*. 2006;5(3):173-86.
46. ICH Expert Working Group. ICH Harmonised Tripartite Guideline: Statistical Principles For Clinical Trials E9; Step 4 Version [online]. 05.02.1998 [last accessed: 22.05.2017]. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf.
47. Gøtzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *International journal of technology assessment in health care*. 1996;12(2):238-46.
48. Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstetrics and Gynecology*. 2005;105(5):1114-8.
49. European Medicines Agency. ICH Topic E 8: Note For Guidance On General Considerations For Clinical Trials; CPMP/ICH/291/95 [online]. March 1998 [last accessed: 22.05.2017]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002877.pdf.
50. Food and Drug Administration. Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics [online]. May 2007 [last accessed: 22.05.2017]. URL: <http://www.fda.gov/downloads/Drugs/Guidances/ucm071590.pdf>.
51. U.S. Food and Drug Administration. Accelerated Approval [online]. 15.09.2014 [last accessed: 29.05.2017]. URL: <https://www.fda.gov/ForPatients/Approvals/Fast/ucm405447.htm>.
52. European Medicines Agency. Guideline on the evaluation of anticancer medicinal products in man; EMA/CHMP/205/95/Rev.4 [online]. 13.12.2012 [last accessed: 22.05.2017]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137128.pdf.
53. Molenberghs G, Buyse M, Burzykowski T. The History of Surrogate Endpoint Validation. In: Burzykowski T, Molenberghs G, Buyse M, editors. *The Evaluation of Surrogate Endpoints*. New York, NY: Springer; 2005. p. 67-82.
54. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Aussagekraft von Surrogatendpunkten in der Onkologie; Rapid Report; Auftrag: A10-05 [online]. 21.11.2011 (IQWiG-Berichte – Jahr: 2011 Nr. 80). URL: https://www.iqwig.de/download/A10-05_Rapid_Report_Version_1-1_Surrogatendpunkte_in_der_Onkologie.pdf.
55. Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in medicine*. 1989;8(4):431-40.
56. Burzykowski T, Molenberghs G, Buyse M. *The Evaluation of Surrogate Endpoints*. New York, NY: Springer; 2005.

57. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Statistical methods in medical research*. 2008;17(3):303-40.
58. Piedbois P, Buyse M. Endpoints and surrogate endpoints in colorectal cancer: a review of recent developments. *Current opinion in oncology*. 2008;20:466-71.
59. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Validity of surrogate endpoints in oncology; Executive Summary; Commission No. A10-05 [online]. 21.11.2011 [last accessed: 29.05.2017]. URL: https://www.iqwig.de/download/A10-05_Executive_Summary_v1-1_Surrogate_endpoints_in_oncology.pdf.
60. Gemeinsamer Bundesausschuss. Nutzenbewertungsverfahren zum Wirkstoff Dabrafenib [online]. 16.06.2016 [last accessed: 29.05.2017]. URL: <https://www.g-ba.de/informationen/nutzenbewertung/80/>.
61. Gemeinsamer Bundesausschuss. Nutzenbewertungsverfahren zum Wirkstoff Pembrolizumab [online]. 04.02.2016 [last accessed: 29.05.2017]. URL: <https://www.g-ba.de/informationen/nutzenbewertung/184/>.
62. Gemeinsamer Bundesausschuss. Nutzenbewertungsverfahren zum Wirkstoff Palbociclib [online]. 18.05.2017 [last accessed: 29.05.2017]. URL: <https://www.g-ba.de/informationen/nutzenbewertung/269/>.
63. Mangiapane S, Velasco Garrido M. Surrogatendpunkte als Parameter der Nutzenbewertung; Schriftenreihe Health Technology Assessment, Bd. 91; Deutsches Institut für Medizinische Dokumentation und Information (DIMDI), editor [online]. 2009 [last accessed: 22.05.2017]. URL: https://portal.dimdi.de/de/hta/hta_berichte/hta250_bericht_de.pdf.
64. European Medicines Agency. Scientific Guidelines [online]. 2017 [last accessed: 04.04.2017]. URL: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000043.jsp&mid=WC0b01ac05800240cb.
65. European Medicines Agency. Guideline on Missing Data in Confirmatory Clinical Trials; EMA/CPMP/EWP/1776/99 Rev. 1 [online]. 02.07.2010 [last accessed: 04.04.2017]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.
66. Potthast R, Beckmann L, Vervölgyi V. Herausforderungen bei der Bewertung onkologischer Studien [Poster]. IQWiG Herbst-Symposium; 28.-29.11.2014; Köln, Deutschland. URL: https://www.iqwig.de/download/IQWiG-Symposium_2014_Poster_Herausforderungen_bei_der_Bewertung_onkologischer_Studien.pdf.
67. Gemeinsamer Bundesausschuss. Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V - Bewertung von Arzneimitteln für seltene Leiden nach § 35a Absatz 1 Satz 10 i.V.m. 5. Kapitel § 12 Nr. 1 Satz 2 VerfO Wirkstoff: Brentuximab Vedotin [online]. 01.11.2016 [last accessed: 04.04.2017]. URL: https://www.g-ba.de/downloads/92-975-1627/2016-08-01_Nutzenbewertung-G-BA_Brentuximab-Vedotin_nAWG_D-253.pdf.
68. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Nivolumab (nicht plattenepitheliales NSCLC) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A16-25 [online]. 28.07.2016 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 416). URL: https://www.iqwig.de/download/A16-25_Nivolumab_Nutzenbewertung-35a-SGB-V.pdf.
69. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 2013 28.06.2013. Report No.: Contract No.: A10-01.

70. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Telaprevir - Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A11-25 [online]. 12.01.2012 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 115). URL: https://www.iqwig.de/download/A11-25_Telaprevir_Nutzenbewertung_35a_SGB_V.PDF.
71. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ombitasvir/Paritaprevir/Ritonavir - Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A15-04 [online]. 29.04.2015 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 297). URL: https://www.iqwig.de/download/A15-04_Ombitasvir-Paritaprevir-Ritonavir_Nutzenbewertung-35a-SGB-V.pdf.
72. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Osteodensitometrie bei primärer und sekundärer Osteoporose; Abschlussbericht; Auftrag D07-01; Version 1.1 [online]. 11.10.2010 [last accessed: 04.04.2017]. (IQWiG-Berichte – Jahr: 2010 Nr. 73). URL: https://www.iqwig.de/download/D07-01_Abschlussbericht_Osteodensitometrie_bei_primaerer_und_sekundaerer_Osteoporose.pdf.
73. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Osteodensitometrie bei primärer und sekundärer Osteoporose; Berichtsplan; Version 2.0; Auftrag: D07-01 [online]. 29.09.2008 [last accessed: 04.04.2017]. URL: https://www.iqwig.de/download/D07-01_Berichtsplan_V_2_0_Osteodensitometrie_bei_primaerer_und_sekundaerer_Osteoporose.pdf.
74. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Positronenemissionstomographie (PET und PET/CT) bei rezidivierendem kolorektalen Karzinom; Abschlussbericht; Auftrag: D06-01C [online]. 28.08.2012 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 140). URL: https://www.iqwig.de/download/D06-01C_Abschlussbericht_PET-und-PET-CT-bei-rezidivierendem-kolorektalem-Karzinom.pdf.
75. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Positronenemissionstomographie (PET und PET/CT) bei rezidivierendem kolorektalen Karzinom; Berichtsplan; Version 1.0; Auftrag: D06-01C [online]. 17.12.2009 [last accessed: 04.04.2017]. URL: https://www.iqwig.de/download/D06-01C_Berichtsplan_PET_und_PET_CT_bei_rezidivierendem_kolorektalen_Karzinom.pdf.
76. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Cholinesterasehemmer bei Alzheimer Demenz; Ergänzungsauftrag: Rivastigmin-Pflaster und Galantamin; Abschlussbericht; Auftrag: A09-05 [online]. 03.02.2012 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 118). URL: https://www.iqwig.de/download/A09-05_Abschlussbericht_Cholinesterasehemmer_Ergaenzungsauftrag_Rivastigmin_Pflaster_Galantamin.pdf.
77. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Daclatasvir - Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A14-31 [online]. 27.11.2014 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 261). URL: https://www.iqwig.de/download/A14-31_Daclatasvir_Nutzenbewertung-35a-SGB-V.pdf.
78. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Nivolumab (Nierenzellkarzinom) - Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A16-24 [online]. 28.07.2016 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 415). URL: https://www.iqwig.de/download/A16-24_Nivolumab_Nutzenbewertung-35a-SGB-V.pdf.
79. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ipilimumab (neues Anwendungsgebiet) - Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A13-44 [online]. 13.03.2014 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 209). URL: https://www.iqwig.de/download/A13-44_Ipilimumab-neues-Anwendungsgebiet_Nutzenbewertung-35a-SGB-V.pdf.

80. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Tafamidis Meglumin - Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A11-31 [online]. 13.03.2012 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 122). URL: https://www.iqwig.de/download/A11-31_Tafamidis_Meglumin_Nutzenbewertung_35a_SGB_V.pdf.
81. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Sofosbuvir - Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A14-05 [online]. 29.04.2014 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 219). URL: https://www.iqwig.de/download/A14-05_Sofosbuvir_Nutzenbewertung-35a-SGB-V.pdf.
82. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Propranolol - Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag: A14-29 [online]. 26.11.2014 [last accessed: 04.04.2017]. (IQWiG-Berichte – Nr. 260). URL: https://www.iqwig.de/download/A14-29_Propranolol_Nutzenbewertung-35a-SGB-V.pdf.
83. Little RJ, D'agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. New England Journal of Medicine. 2012;367(14):1355-60.
84. Carpenter JR, Kenward MG. Missing data in randomised controlled trials - a practical guide. 21.11.2007.
85. Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. American Journal of Epidemiology. 2004;160(1):34-45.
86. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. New York: Cambridge University Press; 2006.
87. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Final Concept Paper E9(R1): Addendum to Statistical Principles for Clinical Trials on Choosing Appropriate Estimands and Defining Sensitivity Analyses in Clinical Trials [online]. Oktober 2014 [last accessed: 04.04.2017]. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9_R1_Final_Concept_Paper_October_23_2014.pdf.
88. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions; Version 5.1.0. März 2011 [last accessed: 04.04.2017]. URL: www.handbook.cochrane.org.
89. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews; AHRQ Publication No. 10(14)-EHC063-EF [online]. Januar 2014 [last accessed: 04.04.2017]. URL: <https://www.effectivehealthcare.ahrq.gov/ehc/products/60/318/CER-Methods-Guide-140109.pdf>.
90. National Institute for Health and Care Excellence. The guidelines manual: appendices B–I - Appendix C: Methodology checklist: randomised controlled trials [online]. November 2012 [last accessed: 04.04.2017]. URL: <https://www.nice.org.uk/process/pmg6/resources/the-guidelines-manual-appendices-bi-1967364/chapter/appendix-c-methodology-checklist-randomised-controlled-trials>.
91. Gemeinsamer Bundesausschuss. Nutzenbewertungsverfahren zum Wirkstoff Axitinib [online]. 21.03.2013 [last accessed: 29.05.2017]. URL: <https://www.g-ba.de/informationen/nutzenbewertung/40/>.
92. Gemeinsamer Bundesausschuss. Nutzenbewertungsverfahren zum Wirkstoff Mikrobielle Collagenase aus Clostridium histolyticum [online]. 19.04.2012 [last accessed: 29.05.2017]. URL: <https://www.g-ba.de/informationen/nutzenbewertung/11/>.
93. Lumley T. Network meta-analysis for indirect treatment comparisons. Statistics in medicine. 2002;21(16):2313-24.

94. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in medicine*. 2004;23(20):3105-24.
95. Bender R, Schwenke C, Schmoor C, Hauschke D. Gemeinsame Stellungnahme von IQWiG, GMDs und IBS-DR: Stellenwert von Ergebnissen aus indirekten Vergleichen [online]. 02.02.2012 [last accessed: 04.05.2017]. URL: https://www.iqwig.de/download/12-03-07_Gemeinsame_Stellungnahme_IQWiG_GMDs_IBS-DR_zum_indirekten_Vergleich.pdf.
96. Guddat C, Grouven U, Bender R, Skipka G. A note on the graphical presentation of prediction intervals in random-effects meta-analyses. *Systematic Reviews*. 2012;1(34):5.
97. Bucher HC, Guyatt G, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of clinical epidemiology*. 1997;50(6):683-91.
98. Borenstein M, Hedges LV, Higgins J, Rothstein HR. *Introduction to Meta-Analysis*: John Wiley & Sons, Ltd; 2009.
99. Signorovitch J, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, et al. Comparative Effectiveness Without Head-to-Head Trials. *PharmacoEconomics*. 2010;28(10):935-45.
100. Parks D, Lin X, Menius AJ, Lee K. Matching-Adjusted Indirect Comparison Using Bootstrapping for Comparative Effectiveness: Individual Patient Data vs. Published Aggregated Data [Poster]. Joint Statistical Meetings; 01.08.2012; San Diego, California. URL: <http://www.amstat.org/meetings/jsm/2012/onlineprogram/AbstractDetails.cfm?abstractid=305200>.
101. Caro J, Ishak KJ. No Head-to-Head Trial? Simulate the Missing Arms. *PharmacoEconomics*. 2010;28(10):957-67.
102. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2014;17(2):157-73.
103. Kiefer C, Sturtz S, Bender R. Indirect Comparisons and Network Meta-Analyses. *Deutsches Arzteblatt international*. 2015;112(47):803-8.
104. Gemeinsamer Bundesausschuss. Die Kosten-Nutzen-Bewertung von Arzneimitteln nach § 35b SGB V [online]. 28.10.2014 [last accessed: 28.02.2017]. URL: <https://www.g-ba.de/institution/themenschwerpunkte/arzneimittel/kosten-nutzen/>.
105. Gemeinsamer Bundesausschuss. Dokumentvorlage: Dossier zur Kosten-Nutzen-Bewertung gemäß § 35b SGB V; Modul K4 [online]. 14.11.2013 [last accessed: 28.02.2017]. URL: https://www.g-ba.de/downloads/17-98-3669/2013-11-14f_AnIVIII_Dossievorlage_Modul_K4.pdf.