

General Methods^a

Version 3.0 of 27.05.2008

Contact:

Institute for Quality and Efficiency in Health Care

Dillenburger Str. 27

51105 Cologne

Germany

Tel.: +49 (0)221 / 35685-0

Fax: +49 (0)221 / 35685-1

E-mail: methoden@iqwig.de

^a This translation is based on the German document “Allgemeine Methoden” (Version 3.0) of 27.05.2008. Please note: The translation is provided as a service by IQWiG to English-language readers. However, solely the German original text is absolutely authoritative and legally binding.

General comments:

The first draft of the Institute's methods paper was prepared in autumn 2004 and published for discussion on 1 November 2004. Following the receipt of comments and expert opinions, a round table was held in February 2005, including the contributors and some members of the Institute's Scientific Advisory Board. *Methods* (Version 1.0) of 1 March 2005 was subsequently published.

In 2006, the document was revised and two successive drafts were prepared for discussion; one internal draft of 27 April 2006, and a second draft of 28 September 2006, which was published on the IQWiG website. Taking the comments submitted on both drafts into account, *Methods* (Version 2.0) of 19 December 2006 was published.

The structure and content of Version 2.0 were revised under consideration of the Statutory Health Insurance (SHI) Act to Promote Competition.^b As new methods had to be developed, in particular for the assessment of the cost-benefit relation of SHI services, the chapter on health economics was extracted and published in a separate document (*Methods for Assessment of the Relation of Benefits to Costs*). In 2007, two general methods drafts were prepared for discussion; an internal draft of 5 July 2007, which was discussed in detail with the Institute's Scientific Advisory Board, and a further draft of 15 November 2007, which was published on the IQWiG website. Taking the comments submitted into account, *General Methods* (Version 3.0) of 27 May 2008 was published. This document is supplemented by a glossary, which is available on the IQWiG website (<http://www.iqwig.de/general-methods.428.en.html>).

The *General Methods* are usually reviewed annually with regard to any necessary revisions, unless errors in the document or relevant developments necessitate prior updating. Project-specific methods are defined on the basis of the version of the *General Methods* valid at that time. If changes are made to the *General Methods* during the course of the project, then it will be assessed whether project-specific procedures need to be modified accordingly.

^b GKV-Wettbewerbsstärkungsgesetz

Preamble

The *General Methods* explain the legal and scientific basis of the Institute for Quality and Efficiency in Health Care (IQWiG[°]). The tasks and structure of the Institute are described in this document, as are the scientific tools applied in the preparation of IQWiG's products. The methods for the assessment of the cost-benefit relation of SHI services are presented in the *Methods for Assessment of the Relation of Benefits to Costs*. Both methods papers therefore provide an important contribution towards transparency in the Institute's mode of operation.

The *General Methods* are primarily directed at researchers. In order to make the information on the Institute's mode of operation accessible to as many interested persons as possible, the authors have aimed to make the document comprehensible. However, as with any scientific text, the *General Methods* also assume a certain level of prior knowledge on the topic.

The *General Methods* can describe procedures only in a general manner. Which specific steps the Institute undertakes in the assessment of specific medical interventions depends, among other things, on the research question posed and the available scientific evidence. The *General Methods* should therefore be regarded as a kind of framework. How the assessment process is designed in individual cases is presented in detail for each specific project.

In order to further develop and improve its mode of operation, the Institute presents its *General Methods* for public discussion. This applies to the current valid version as well as to all drafts of future versions.

[°] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Table of contents

General comments	i
Preamble	ii
Table of contents	iii
List of abbreviations	vi
1 The Institute for Quality and Efficiency in Health Care	1
1.1 Legal responsibilities	1
1.2 Structure	3
1.3 Evidence-based medicine	6
1.3.1 <i>Practical evidence-based medicine</i>	7
1.3.2 <i>Strategies of evidence-based medicine</i>	7
1.3.3 <i>The relevance of evidence-based medicine for the Institute</i>	8
1.3.4 <i>Evidence-based decision-making in health care</i>	8
1.3.5 <i>The relevance of certainty of results</i>	9
1.3.6 <i>The connection between certainty of results and proximity to everyday conditions</i>	10
2 The Institute's products	13
2.1 Product-specific procedures	13
2.1.1 <i>Reports</i>	14
2.1.2 <i>Rapid reports</i>	19
2.1.3 <i>Health information</i>	21
2.1.4 <i>Working papers</i>	25
2.2 General aspects of the preparation of products	27
2.2.1 <i>Choice of external experts</i>	27
2.2.2 <i>Guarantee of scientific independence</i>	28
2.2.3 <i>Review of the Institute's products</i>	29
2.2.4 <i>Publication of the Institute's products</i>	30
3 Assessment of the benefit and harm of medical interventions	31
3.1 Patient-relevant medical benefit	31
3.1.1 <i>Definition of a patient-relevant medical benefit</i>	31
3.1.2 <i>Surrogates of a patient-relevant medical benefit</i>	33
3.1.3 <i>Assessment of the harm potential of medical interventions</i>	34
3.1.4 <i>Summarising assessment</i>	36
3.2 Specific aspects of the benefit assessment	38
3.2.1 <i>(Un)certainty of results</i>	38
3.2.2 <i>Dramatic effect</i>	39
3.2.3 <i>Study duration</i>	39
3.2.4 <i>Patient-reported outcomes</i>	40
3.2.5 <i>Benefit in small populations</i>	41
3.2.6 <i>Benefit in individual cases</i>	41
3.3 Benefit assessment of drugs	42
3.3.1 <i>Relevance of the drug approval status</i>	43
3.3.2 <i>Studies on the benefit assessment of drugs</i>	44
3.4 Non-drug therapeutic interventions	45

3.5	Diagnostic tests	47
3.6	Population-wide interventions	49
3.6.1	<i>Screening</i>	49
3.6.2	<i>Prevention</i>	50
4	Clinical practice guidelines and disease management programmes	51
4.1	Background of guideline appraisal	51
4.2	Aims of guideline appraisal	51
4.3	Methods of guideline appraisal	52
4.3.1	<i>Formal review</i>	54
4.3.2	<i>Comparison of guidelines and appraisal of content of key recommendations</i>	54
4.3.3	<i>Improvement in outcomes</i>	55
4.4	Presentation of quality assessment	56
4.5	Issue of recommendations on disease management programmes	56
5	Evidence-based health information for consumers and patients	57
5.1	Background and goals	57
5.2	Patient-centred communication	58
5.2.1	<i>Communication standards</i>	58
5.2.2	<i>Visual communication and multi-media</i>	60
5.2.3	<i>Accessibility</i>	61
5.3	Topic selection, research and evaluation of evidence	62
5.3.1	<i>Topic selection</i>	62
5.3.2	<i>Research</i>	64
5.3.3	<i>Evaluation of evidence</i>	66
5.3.4	<i>Updating</i>	67
5.4	Information products	68
5.4.1	<i>Core products</i>	68
5.4.2	<i>Supplementary items</i>	70
5.4.3	<i>Patient stories</i>	71
5.4.4	<i>Website</i>	72
5.5	Monitoring and evaluation	74
5.5.1	<i>Routine monitoring</i>	74
5.5.2	<i>Feedback, corrections and improvements</i>	75
5.5.3	<i>Evaluation</i>	76
6	General methodological aspects	78
6.1	Literature search	78
6.1.1	<i>General principles of topic-related literature searches</i>	78
6.1.2	<i>Search procedure for primary publications</i>	79
6.1.3	<i>Other data sources for literature searches</i>	80
6.1.4	<i>Selection of relevant publications</i>	81
6.1.5	<i>Documentation</i>	82
6.1.6	<i>Literature monitoring</i>	82
6.1.7	<i>Consideration of legal aspects of data protection/confidentiality</i>	83
6.2	Quality assessment of individual studies	83
6.2.1	<i>Relationship between study type and research question</i>	83
6.2.2	<i>Ranking of different study types/evidence levels</i>	84

6.2.3	<i>Aspects of the assessment of the bias potential of results</i>	85
6.2.4	<i>Interpretation of subgroup analyses</i>	88
6.2.5	<i>Assessment of prognosis studies</i>	89
6.2.6	<i>Assessment of the consistency of published data</i>	92
6.3	Consideration of secondary literature	93
6.3.1	<i>Classification of systematic reviews</i>	93
6.3.2	<i>Benefit assessment on the basis of systematic reviews and HTA reports</i>	94
6.3.3	<i>Consideration of published meta-analyses</i>	97
6.4	Specific statistical aspects	98
6.4.1	<i>Description of effects and risks</i>	98
6.4.2	<i>Evaluation of statistical significance</i>	98
6.4.3	<i>Evaluation of clinical relevance</i>	99
6.4.4	<i>Demonstration of a difference</i>	100
6.4.5	<i>Demonstration of equivalence</i>	101
6.4.6	<i>Adjustment principles and multi-factorial methods</i>	102
6.4.7	<i>Meta-analyses</i>	103
6.4.8	<i>Handling of unpublished or partially published data</i>	105
6.4.9	<i>Description of types of bias</i>	106
6.5	Qualitative methods	108
6.5.1	<i>Qualitative studies</i>	108
6.5.2	<i>Consultation techniques</i>	109
	Reference list	111

List of abbreviations

AGREE	Appraisal of Guidelines Research and Evaluation
CONSORT	Consolidated Standards of Reporting Trials
CPG	clinical practice guideline
DELBI	Deutsches Leitlinien-Bewertungs-Instrument (German Instrument for Methodological Guideline Appraisal)
EBM	evidence-based medicine
EU	European Union
G-BA	Gemeinsamer Bundesausschuss (Federal Joint Committee)
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
HTA	health technology assessment
IPD	individual patient data
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care)
ITT	intention to treat
PRO	patient-reported outcome
QoL	quality of life
RCT	randomised controlled trial
SGB	Sozialgesetzbuch (Social Code Book)
SHI	statutory health insurance

A chief cause of poverty in science is mostly imaginary wealth. The aim of science is not to open a door to infinite wisdom but to set a limit to infinite error.

Bertolt Brecht. Life of Galileo. Frankfurt: Suhrkamp. World premiere, first version, Zurich theatre, 1943.

1 The Institute for Quality and Efficiency in Health Care

1.1 Legal responsibilities

With the Health Care Reform of 2004 (Statutory Health Insurance [SHI] Modernisation Act^d), the legislator decided to establish a professionally independent Institute as a component of the German SHI [1]. The German Federal Joint Committee^e followed the legislator's recommendation in §139a (1) S. 2 SGB^f V and established a non-profit and non-government private law foundation in 2004, the Foundation for Quality and Efficiency in Health Care. The sole purpose of the Foundation is the maintenance of the Institute for Quality and Efficiency in Health Care (IQWiG^g).

The Institute addresses fundamental issues relating to the quality and efficiency of SHI services. Its specific responsibilities are outlined in detail in § 139a (3) SGB V:

- Search for, assessment and presentation of current scientific evidence on diagnostic and therapeutic procedures for specific diseases;
- Preparation of scientific reports and expert opinions on quality and efficiency issues of SHI services, taking age, gender, and personal circumstances into account;
- Appraisal of evidence-based clinical practice guidelines on the epidemiologically most important diseases;
- Issue of recommendations on disease management programmes.

The following two responsibilities were redefined in §139a (3) SGB V within the framework of the Health Care Reform 2007 (SHI Act to Promote Competition) [2]:

^d Gesetz zur Modernisierung der gesetzlichen Krankenversicherung = GKV-Modernisierungsgesetz – GMG. (Most parts of this law became effective on 01.01.2004.)

^e Gemeinsamer Bundesausschuss (G-BA). The Federal Joint Committee is the decision-making body of the self-government of the German health care system. More information on the Committee's responsibilities is provided at www.g-ba.de/institution/sys/english/

^f Sozialgesetzbuch (Social Code Book)

^g Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

- Assessment of the benefits and costs of drugs;
- Provision of easily understandable information for all citizens on the quality and efficiency of health care services, as well as on the diagnosis and treatment of diseases of high epidemiological relevance.

The assessment of the benefits and costs of drugs is specified in §35b (1) SGB V (SHI Act to Promote Competition). According to this law, the assessment is to be performed by means of a comparison between a test drug and other drugs or treatment alternatives, taking into account the additional treatment benefit for patients in relation to the costs. The following criteria to determine the benefits for patients are specified: increase in life expectancy, improvement in health status and quality of life (QoL), and reduction in disease duration and adverse effects. The precise operationalisation of a “patient-relevant benefit” (considering the criteria above) is outlined in Section 3.1 (especially in 3.1.1) of this paper. In the economic evaluation, in particular the appropriateness and reasonableness of the reimbursement of services by the community of insured citizens should be given appropriate consideration.

The law also specifies that the assessment of the medical benefit of an intervention must be conducted following internationally recognised standards of evidence-based medicine (EBM), and that the economic evaluation must also be conducted following relevant international standards, in particular those relating to health economics.^h In Section 1.3 of this paper, the definition, development, and underlying concept of EBM are presented.

The modalities of IQWiG’s performance of tasks are specified in §139b SGB V. According to this law, either the Federal Joint Committee (Para. 1) or the Federal Ministry of Health (Para. 2) can act as contracting agencies for IQWiG commissions. In the case of commissioning by the Ministry, the Institute can reject a commission as unfounded, unless the Ministry provides separate funding. IQWiG forwards the results of the commissions (according to Paras. 1 and 2) to the Federal Joint Committee as recommendations that the Committee has to consider in the fulfilment of its legal responsibilities.

The Institute must ensure that external experts are involved in the work on commissions.ⁱ In order to secure the Institute’s professional independence, these experts are required to disclose all connections to associations and contract organisations, particularly in the pharmaceutical and

^h Acc. to §35b (1) and §139a (4) SGB V

ⁱ Acc. to §139b (3) S. 1 SGB V

medical devices industries, including details on the type and amount of any remuneration received.^j The same applies to all IQWiG staff members.^k

The legal obligation for transparency in the procedural steps is outlined in §139a (5) in conjunction with §35b (5) S. 6 SGB V, which were rewritten in the course of the preparation of the SHI Act to Promote Competition. In all key steps of the assessment procedure, the following persons and parties must be given the opportunity to submit comments on the Institute's work: medical, pharmaceutical, and health economic experts (from research and practice), drug manufacturers, relevant organisations representing the interests of patients and self-help groups for the chronically ill and disabled, as well as the Federal Government Commissioner for Patients' Affairs.^l Their comments must be considered in the assessment. The corresponding regulations are outlined in Section 2.1.1 of this paper in connection with the preparation of report plans (protocols) and preliminary reports.

A legal obligation for transparency also exists concerning the Institute's working methodology: the Institute must publish its working processes and results (including the basic principles) on a regular basis.^m This is also regulated in Section 2.1.1 and Sections 2.1.2 to 2.1.4 of this document by the mandatory publication of IQWiG products on its website.

The legal basis of the Institute's funding is outlined in §139c SGB V. It is funded by a specific levy ("system levy"), which originates to 50% from the inpatient sector (levy for every reimbursed hospital case) and to 50% from the outpatient sector (increase in fees for outpatient medical and dental SHI services). The Federal Joint Committee specifies the amount of this levy on an annual basis. The money from this "system levy" goes to both the Institute and the Federal Joint Committee.

1.2 Structure

The Institute is an establishment of the Foundation for Quality and Efficiency in Health Care, and is under scientifically independent management. The Foundation's bodies include a *Foundation Board*, of which 50% comprises members of the National Confederations of Regional Associations of SHI Funds and 50% comprises members of the organisations of service providers,ⁿ as well as a five-member *Board of Directors* appointed by the Foundation Board. One member of the Board of

^j Acc. to §139b (3) S. 2 SGB V

^k Acc. to §139a (6) SGB V

^l Acc. to §139a (5) SGB V

^m Acc. to §139a (4) S. 2 SGB V

ⁿ National Association of Statutory Health Insurance Physicians, German Federal Association of Sick Fund Dentists, German Hospital Federation

Directors is appointed by the Federal Ministry of Health. The third body of the Foundation is the Federal Joint Committee, which can decide on changes to the Charter by a two-thirds majority, according to §15 (1) of the Charter of the Foundation. The Senate Department of Justice, Berlin, is the Foundation's Supervisory Authority.

The 30-member *Board of Trustees* acts as the Institute's advisory committee. This Board includes members of the supporting organisations of the Federal Joint Committee and other members not represented in the Committee (i.e. relevant organisations of service providers and of employees and employers, as well as other relevant health care organisations). The *Scientific Advisory Board* is the Institute's second advisory committee, and is appointed by the Board of Directors after consultation with the Institute Management. It comprises up to 12 members.

The Institute is currently organised in the following departments:

- Drug Assessment
- Medical Biometry
- Non-drug Interventions
- Health Economics
- Quality of Health Care
- Communications
- Health Information
- Administration

The Institute's Steering Committee includes the Institute Management and the Department Heads, and advises the Institute Management. It deals with both internal and external tasks. The Steering Committee is responsible for ensuring compliance with the Institute's procedures and methods, which it develops and modifies together with other staff members. In addition, it lays down the requirements for the award of commissions to external experts (according to the principles of the Foundation's Board of Director's regarding the award of commissions).

The Institute Director, who is appointed by the Foundation's Board of Directors, represents the Institute internally and externally. The current Director is Professor Peter T. Sawicki. He is responsible for the fulfilment of the Institute's legal obligations, following the priorities set by the Federal Joint Committee. He is also responsible for observance of the Institute's Budget and Standing Orders.

Institute for Quality and Efficiency in Health Care

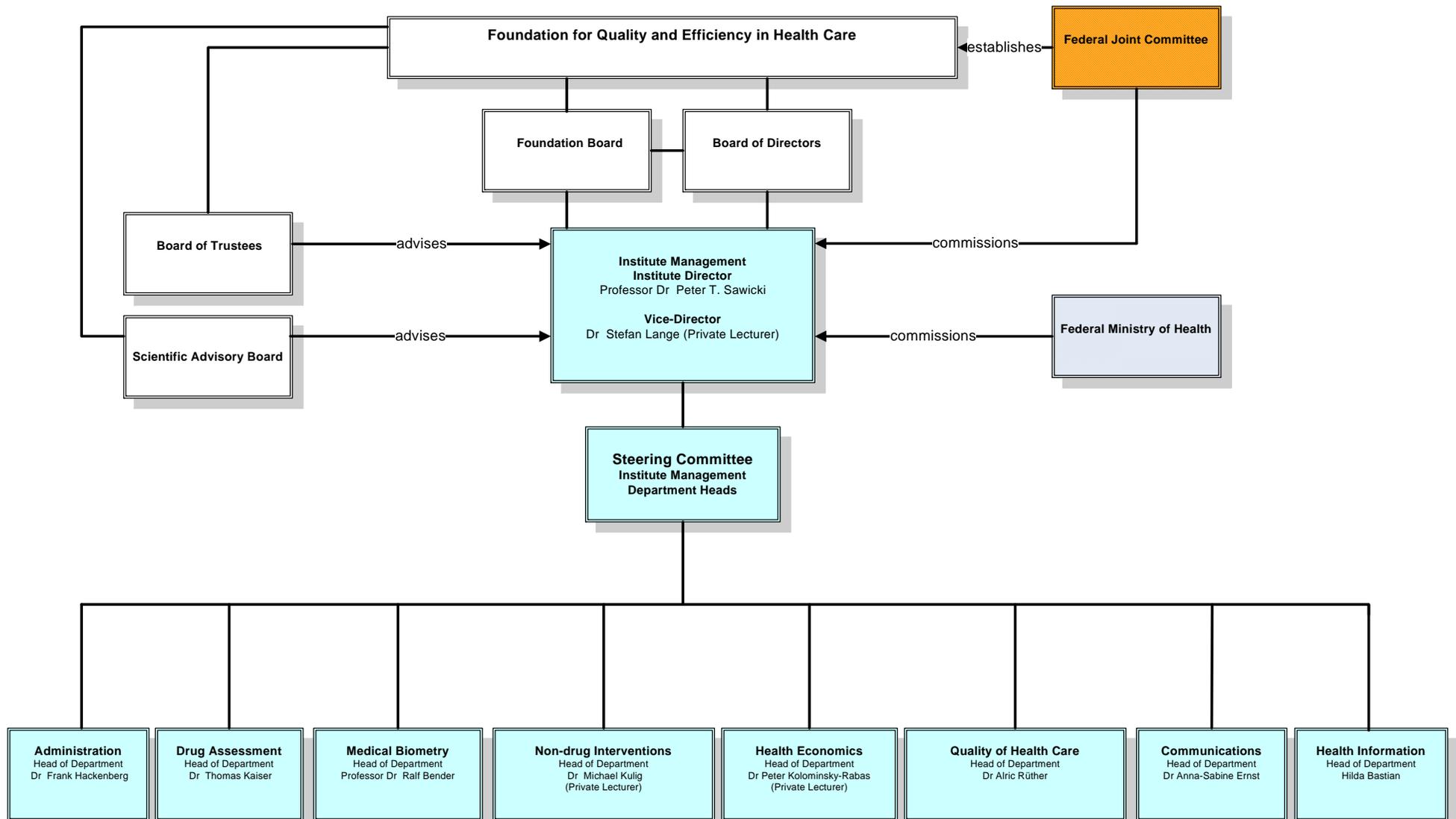


Figure 1: Organisation chart of the Institute for Quality and Efficiency in Health Care

1.3 Evidence-based medicine

EBM refers to patient health care that is not only based on opinions and consensus, but considers “evidence” – proof determined with the most objective scientific methods possible. EBM comprises tools and strategies designed to safeguard against false decisions and false expectations. In this context, a false decision can mean that beneficial interventions are not implemented in health care (or implemented with delay), or that useless or even harmful interventions are widely applied [14,123,165,168].

However, tools designed to prevent subjectively affected (and therefore often biased) assessments (see Chapter 6) were not first invented with the introduction of the term “EBM”, but originated decades ago. In Germany, Paul Martini described the main elements of a fair assessment of drug effectiveness as early as 1932 in his monograph *Methodology of Therapeutic Investigations* [258]. The method of randomly allocating study participants to comparison groups (randomisation) in order to assess the effectiveness and safety of medical interventions became the internationally accepted standard in the early 1960s [188]. Starting in the United States, in this period this type of study became the precondition for the approval of drugs (and some medicinal products) regulated by authorities and legislation [31]. About 20 years later, clinical epidemiologists attempted to establish this methodology in clinical practice [127]. Accompanied at times by serious controversy, this was not actually achieved until the 1990s, at the same time as the concept was defined as “EBM”. Since this time, primary studies, as well as the systematic search for and assessment of studies have formed the basis of the international scientific standard for health technology assessments (HTAs) [30].

The Institute is legally obliged to apply international EBM standards recognised by relevant experts.^o It is the task of this methods paper to describe the methods and strategies that define these international standards. EBM is not an inflexible concept. Which standard tool is to be applied, and when, depends on the question to be answered and the decision to be made. Despite the use of standards, decisions have to be made repeatedly in the search for, and the processing and assessment of studies for which no international standards are (as yet) available. EBM also includes the freedom to make one’s own specifications in such situations. However, this freedom is closely linked to the obligation to define such specifications preferably a priori, and to explain assessments in a transparent manner, so that the rationale is comprehensible. This chapter explains that in the

^o Acc. to §35b (1) SGB V

implementation of EBM and the definition of specifications, an institution such as IQWiG is in a different situation from clinicians who are seeking support for a decision on treatment.

1.3.1 Practical evidence-based medicine

The EBM concept is a strategy for physicians who, from a range of possible treatment alternatives, seek the most promising option suited best to the needs of their patients, and who aim to present prospects of success without exaggeration. This implementation of EBM for individual patients in daily clinical practice was defined by David Sackett et al. as follows: “EBM is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of the individual patient. It means integrating individual clinical expertise with the best available external clinical evidence from systematic research” (1996) [93].

However, the “best available evidence” is often incomplete or not reliable. EBM also supports physicians and patients in recognising such uncertainty; they can then discuss how to deal with this. Especially in uncertain situations, the personal preferences of patients are relevant, who decide which option to choose. Apart from being based on evidence, decisions are also ideally based on the clinical condition and circumstances of the individual patient, as well as on his or her preferences and actions [180]. At the same time, the description of identified gaps in knowledge creates the precondition for medical research targeted towards patients’ needs.

EBM is based on a critical approach [231]. The importance of scepticism is underlined by the fact that over the past few decades, several insufficiently tested but widely applied therapies have been assessed with EBM methods, and that these assessments have shown that a hasty, overoptimistic approach to a new intervention can have dangerous consequences for patients [104,304]. It is the Institute’s task to assess objectively with what certainty the benefit of medical interventions has been demonstrated, in order to counter inappropriate judgements.

1.3.2 Strategies of evidence-based medicine

A characteristic standard element of EBM is the structured and systematic approach to seeking a response to a medical question:

1. The medical question must be formulated precisely. Medicine (nearly) always deals with the choice between at least two alternatives. This can refer to treatments, diagnostic tests or complex changes in life style. From this, the following question is always inferred: Is Option A better than Option B? In this context, the decision not to undergo treatment can also be an option that should be seriously considered. However, it should be stressed that this option (e.g. “watchful waiting”) is not the same as “doing nothing”.

2. It must be defined how the treatment benefit (or diagnosis or lifestyle change) should be measured. A standard element of EBM is the issue of patient-relevant outcomes: Can life expectancy be increased? Do symptoms and QoL improve?
3. In EBM it is explicitly noted that in medicine, only statements on probability or only conclusions about groups of patients are usually possible with regard to the benefit of treatment, diagnostic procedures, or life style changes. Benefit is demonstrated by showing that an intervention increases the probability of a beneficial outcome and/or reduces the risk of a non-beneficial outcome. In order to demonstrate the benefit of an intervention, studies in sufficiently large groups of suitable patients are necessary. International researchers have developed a range of rules and tools for the planning, conduct, and analysis of such studies. The most important aim is to minimise (or, if this is impossible, at least document) factors that can distort the results of a comparison. The effects of such confounding factors are referred to as “bias”. The rules and tools that are internationally accepted as the prevailing standard, and are under continuous development, are the methodological basis of EBM and the Institute’s work.
4. A further key EBM strategy is to identify all studies of appropriate quality on a question and, in this way, to summarise the evidence available. In this context, if large differences are shown between the results of individual studies (heterogeneity), an attempt should be made to explain them. The findings of these summaries and assessments are referred to as systematic reviews; the statistical analyses are referred to as meta-analyses.

1.3.3 The relevance of evidence-based medicine for the Institute

The Institute’s main task is to provide the most reliable answer possible to the question specified by the contracting agency as to whether evidence is available of the benefits or harms of an intervention. The aim is to present sufficiently reliable proof that “Treatment A” is better for patients than “Alternative B” for a specific disease. In short: What is the benefit of A compared with B?

The Institute’s commissions are purposely targeted, not towards the treatment of individual patients with their potential specific characteristics, but towards indicating for which patient groups proof of a benefit of an intervention is available.

1.3.4 Evidence-based decision-making in health care

Most reports prepared by the Institute are to serve the Federal Joint Committee as a basis for future decisions, which are binding for all SHI-insured persons. Other reports are, for example, to serve as

information for this Committee. The type of decision made by institutions such as the Federal Joint Committee has an effect on the use of EBM methods.

1.3.5 The relevance of certainty of results

A specific characteristic of EBM is that it allows an assessment with regard to what extent the available evidence is reliable. Decisions made by the Federal Joint Committee must be based on highly reliable scientific evidence, as they have binding consequences for all SHI-insured persons (e.g. exclusion of services from reimbursement).

This also affects the Institute's work: the assessment of the certainty of results plays a key role in the Institute's reports. Numerous details regarding how studies are planned, conducted, analysed, and published have an impact on how reliable the available results are. It is an international EBM standard to test and assess these aspects critically. However, how the necessary certainty of results can be achieved in order to answer a question also depends on the disease and on the effect size of an intervention: if two runners pass the finishing line of a fair race with a great distance between them, one does not need a stopwatch to recognise the winner. For example, the benefit of a new therapy that results in the cure of a disease that had previously always been fatal can be proven in a relatively small number of surviving patients. In this case, the judgement is also ultimately based on a comparison, but in interventions with such dramatic effects, the comparison between historical and current patients may already provide sufficient certainty. However, therapies that show such dramatic benefits are rare in modern medicine.

It is particularly common in chronic diseases that differences between two therapy alternatives may be easily confounded by a fluctuant course of disease. In these cases, precise methods are required in order to be able to recognise therapy effects under such fluctuations.

It can be assumed that the Institute will be specifically commissioned to compare such interventions where it is not immediately recognisable which alternative will be more beneficial. However, the smaller the expected differences between two alternatives are, the more precise and reliable the studies must be in order to be sufficiently certain that an observed effect is not caused by chance or measurement errors (a world record over 100 metres can no longer be measured with a sandglass).

In addition, when small differences are shown, their clinical relevance must also be assessed. The following requirements for precision and reliability determine the Institute's mode of operation.

1. It is an international EBM standard to specify for every question the type of study that minimises the risk of unjustifiably discriminating against one of the alternatives.

2. If it emerges that studies of the required quality and precision are generally lacking, it is the core task of the Institute to describe the circumstances and conclude that on the basis of the “currently best available” evidence, it is not possible to make reliable recommendations.
3. The Institute’s assessments on the benefits and harms of interventions are therefore normally based only on studies with sufficient certainty of results. This ensures that the decisions made by the Federal Joint Committee, which are based on the Institute’s recommendations, are supported by a sound scientific foundation. Moreover, an assessment that includes a literature search for studies with insufficient certainty of results would be costly and time consuming.
4. In addition to considering scientific evidence, it is also the Federal Joint Committee’s responsibility to consider other aspects in its decisions, such as the efficiency of interventions and the needs and values of the population [155]. In an uncertain scientific situation, such aspects take on an increased importance.

1.3.6 The connection between certainty of results and proximity to everyday conditions

The Institute’s focus on the assessment of the certainty of results is often criticised. One argument is that studies with a high certainty of results (especially randomised controlled trials; RCTs) may possess internal validity, but often do not represent patient health care under everyday conditions, and are therefore not transferable to this setting, i.e. do not possess external validity. This criticism is then often connected to the requirement to consider additional study types, in order to incorporate everyday conditions more successfully. However, this criticism conflates levels of arguments that should be clearly separated.

The following aspects should to be taken into account:

1. The basic precondition for a benefit assessment is the demonstration of causality, and an indispensable requirement for the demonstration of causality is a comparative experiment in which an effect can be ascribed to a single determining factor. This means that considerable efforts need to be made in clinical trials, as there are numerous confounding factors that feign or mask effects. One of the strongest influences arises from the unfair allocation of study participants to comparison groups. There is no doubt that randomisation is currently the best available tool to minimise this type of bias. Random allocation of participants to groups ensures that there are no systematic differences between groups, neither regarding known factors (e.g. age, gender, disease severity), nor unknown factors. If randomisation is ensured, the basic precondition for the demonstration of causality is given. However,

randomisation alone does not guarantee a high certainty of results. To achieve this, unbiased assessment, summarisation and publication of results, for example, are also required.

2. One of the key findings of EBM is that study types other than RCTs are usually not suited to prove causality. It is possible to investigate patients under everyday conditions in other study types; however, most of these studies cannot answer with sufficient certainty the relevant question as to whether a difference is caused by the intervention. Non-randomised studies always provide potentially biased results, even if only minor selection bias existed in the choice of participants. As a matter of principle, structural equality of groups cannot be assumed in these studies. The use of non-randomised studies as proof of the causality of an intervention therefore requires particular justification.
3. For example, it is correct that RCTs often exclude patients with common concomitant diseases. However, this is not a consequence of randomisation, but is due to other factors (e.g. study inclusion and exclusion criteria). In addition, patients in RCTs are often cared for more intensively than in everyday practice. However, these are intentional decisions made by those persons who wish to answer a specific question in a study. There is also a selection of participants in non-randomised studies through design characteristics, so that external validity cannot be automatically assumed in this study type any more than in an RCT.
4. Even if patient groups in an RCT differ from those in everyday health care, this does not mean the external validity of study results must be questioned. The decisive issue is in fact whether it is to be expected that a therapy effect determined in a population varies in a different population.
5. It depends on the individual case what effects the intensity of care provided in a study could have. For example, it is conceivable that a benefit of an intervention actually exists only if patients are cared for by specially qualified physicians, as under everyday conditions an increased rate of complications may otherwise occur. However, it is also possible that intensified care may reduce differences between groups. For example, differences in treatment compliance may be smaller in studies where patients are provided with intensified care as a matter of principle.
6. However, the initiator of a clinical trial is responsible for the specification of study conditions. The initiator can define research questions and outcomes as relevant for investigation in the study. If, for example, a drug manufacturer regards treatment compliance to be an important aspect of the benefit of a product, it would be logical to design studies that can measure this aspect with the greatest possible certainty of results and proximity to everyday conditions, and at the same time demonstrate relevance for patients.

The above remarks show that certainty of results and proximity to everyday conditions (or internal and external validity) have no fixed relationship. Great certainty of results and proximity to everyday conditions do not exclude one another, but only require the intelligent combination of study type, design, and conduct.

Even if criticism of the lack of proximity to everyday practice is actually justified for many studies, nothing would be gained by dispensing with great certainty of results in favour of improved proximity to everyday practice, because one would thereby be attempting to compensate one deficit by another, more fundamental deficit.

Studies that combine proximity to everyday conditions and a high certainty of results are both desirable and feasible. RCTs are quite feasible that specify neither requirements for patients beyond everyday health care nor fixed study visits. Such studies are being discussed at an international level (“real world trials”, “practical trials” or “pragmatic trials”) [138,140,151,257,361].

2 The Institute's products

According to its legal remit, the Institute generates a variety of products in the form of scientific reports and easily understandable health information for consumers and patients. The following chapter describes procedures and general methods applied in the preparation of the Institute's products. The individual products and product-specific procedures are presented in Section 2.1, and further aspects independent of products are described in Section 2.2.

The general methodology applied in the preparation of products arises from the Institute's task of systematically compiling and assessing information on a research question (information enabling interpretable statements on the benefit or harm of a medical intervention). It is not the Institute's task to *prove* the potential benefit or harm (or the lack of a benefit or harm) of a medical intervention.

Theoretically, the assessment of a method or intervention is possible at any time. The time point of the assessment is mainly specified by the time point the Institute is commissioned (insofar as a commission is required). There are no general specifications laying down that the preparation of a systematic review by the Institute should be conducted at the earliest after a certain period following the approval or establishment of a method or intervention. If, in the case of an early assessment of an intervention, great uncertainty of results is noted due to the lack of long-term studies, this is described according to the Institute's general working methods.

2.1 Product-specific procedures

The product range of the Institute includes four main products:

- Detailed **reports** (especially benefit and cost-benefit assessments);
- **Rapid reports**;
- **Health information** (easily understandable information for consumers and patients);
- **Working papers** (on relevant health care developments and on the Institute's methodological work).

The preparation of reports and rapid reports is conducted solely on the basis of the award of individual commissions through the Federal Joint Committee or Ministry of Health. The primary aim of reports is to provide recommendations for decisions on directives of the Federal Joint Committee. The aim of a rapid report is to provide information at short notice on relevant current topics, as well as on research questions not targeted at supporting such decisions on directives.

Health information can be prepared either on the basis of an individual commission, or as a result of the commissioning of a benefit assessment (easily understandable report version), or within the framework of IQWiG's general legal remit to provide health information.

Working papers are prepared on the Institute's own initiative; specific commissioning by the Federal Joint Committee or Ministry of Health is not required. This takes place either on the basis of the general commission (see Section 2.1.4), with the aim of providing information on relevant health care developments, or on the basis of the legal remit to develop the Institute's methods. The Institute's *General Methods* are not to be understood as a working paper in this sense, and are subjected to a separate preparation and updating procedure (see "General comments", page i).

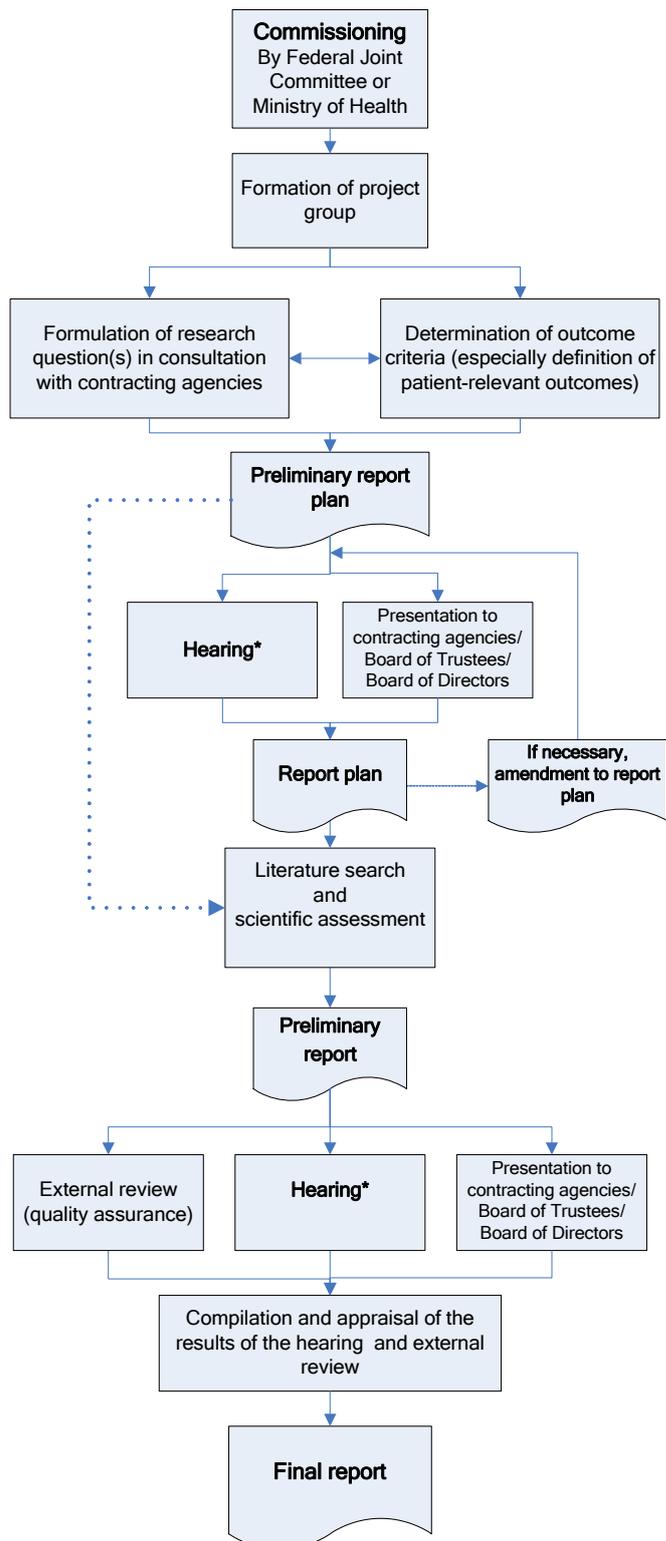
An overview of the Institute's various products is shown in Table 1. Product-specific procedures are described in the subsequent Sections 2.1.1 to 2.1.4.

Table 1: Overview of the Institute's products

Product	Objective	Procedure described in	Commissioned by
Report	Recommendations for decisions on directives of the Federal Joint Committee	Section 2.1.1	Federal Joint Committee, Ministry of Health
Rapid report	Information at short notice on current topics and on research questions not targeted at supporting decisions on directives of the Federal Joint Committee	Section 2.1.2	Federal Joint Committee, Ministry of Health
Health information	Easily understandable information for consumers and patients; wide scope of topics	Section 2.1.3	Federal Joint Committee or Ministry of Health / own initiative of the Institute
Working paper	Information on relevant health care developments or methodological aspects	Section 2.1.4	Own initiative of the Institute

2.1.1 Reports

The procedure of report production is presented in Figure 2. All working steps made fall under the Institute's responsibility and involve external expertise where appropriate. If necessary, the Institute's Scientific Advisory Board is also involved. The internal quality assurance process is not outlined in this flow chart.



* The hearing is conducted by obtaining written comments. In addition, an optional oral scientific debate may be held to discuss any unclear aspects of the written comments.

Figure 2: Production procedure for an IQWiG report

After **commissioning** by the Federal Joint Committee or Ministry of Health, the Institute's internal project group is formed under management of the department concerned. A project manager is then appointed. The composition of the project group is not fixed at this point, as changes may be necessary due to the subsequent steps taken. The research question is worded in consultation with the contracting agency's responsible committees, involving external professional expertise or individual affected persons or parties if appropriate. As a rule, relevant patient organisations are involved, especially in the definition of patient-relevant outcomes. Subsequently, the report plan is prepared.

The **report plan**, comparable to the study protocol of a clinical trial, contains the precise scientific research question, including the outcome criteria (e.g. patient-relevant outcomes), as well as the inclusion and exclusion criteria of the information to be used in the benefit assessment. This plan also includes a description of the project-specific methodology applied in the literature search and in the assessment of the information retrieved. This plan is prepared under the responsibility of the IQWiG project group, usually involving external experts. After completion of the internal quality assurance process and approval by the Institute's Steering Committee, the preliminary report plan is then forwarded to the contracting agency (also to examine its completeness in respect of the commission originally awarded). It is also forwarded to the Foundation's Board of Trustees and Board of Directors. The preliminary report plan is then published on the Institute's website (usually 5 working days later), in order to provide the opportunity to submit comments.

For a period of at least 4 weeks, the public is given the **opportunity to submit written comments (hearing procedure)**. This includes medical, pharmaceutical, and health economic experts from research and practice, professional representatives of pharmacists, drug manufacturers, the relevant organisations responsible for representing the interests of patients and self-help groups of chronically ill and disabled persons, as well as the Federal Government Commissioner for Patients' Affairs. This enables an open and independent reviewing procedure for the report plan. The opportunity to provide comments refers especially to the project-specific methodological approach to answer the research question, including the specification of patient-relevant outcomes. At the same time, the opportunity is also provided to submit any type of document of appropriate quality (especially unpublished data), which is suited to answer the research question of the report, according to the person submitting comments. If the search strategy defined in the report plan is restricted to RCTs, for example, non-randomised studies may nevertheless be submitted within the framework of the submission of comments. However, in such cases, appropriate justification of the validity of the causal interpretation of the effects described in these studies is also required. The research question itself is usually specified by the commission, and is not an object of the commenting procedure. Optionally, an oral scientific debate including persons submitting

comments may be held. This debate serves the potentially necessary clarification of aspects of the written comments and aims at improving the scientific quality of the report plan. In order to avoid undue delay of the Institute's work, the comments must fulfil certain formal requirements.

Further information on the commenting procedure for the preliminary report plan and amendments to the report plan are published on the IQWiG website in the relevant guideline (http://www.iqwig.de/download/Guideline_on_the_submission_of_comments_on_report_plans_preliminary_version_Amendments_to_report_plans_new.pdf). The conditions noted in the current version of this guideline apply.

After the analysis of the comments, the revised report plan is published together with the results of the hearing (i.e. written comments submitted; meeting minutes of the [optional] oral scientific debate; appraisal of comments). This report plan is the basis for the preparation of the preliminary report. If further relevant methodological changes are required in the course of the preparation of the preliminary report, these are usually presented in one or several amendments to the report plan. An opportunity to submit comments is usually also provided after publication of an amendment, following the conditions outlined above.

The results of the literature search and the scientific assessment are presented in the **preliminary report**. In order to avoid undue delay of the Institute's work, the literature search and assessment start before completion of the commenting procedure for the report plan on the basis of the criteria formulated in the preliminary report plan. However, the outcome of the commenting procedure is explicitly not anticipated, as these criteria may be modified on grounds of the comments on the preliminary report plan, which may lead to a supplementation and/or modification of the literature search and assessment.

The preliminary report includes the preliminary recommendation to the Federal Joint Committee. It is produced under the responsibility of the IQWiG project group, usually with the involvement of external experts. After completion of the internal quality assurance process and approval by the Institute's Steering Committee, the preliminary report is then forwarded to the contracting agency (also to examine its completeness in respect of the commission that was originally awarded). It is also forwarded to the Foundation's Board of Trustees and Board of Directors. An additional step in the quality assurance of the preliminary report is the conduct of a review by one or several external experts with recognised methodological and/or topic-related competence.

The preliminary report is published on the Institute's website (usually 5 working days after delivery to the contracting agency) in order to provide the public **the opportunity to submit comments (hearing procedure)** for a period of at least 4 weeks. This includes medical, pharmaceutical, and health economic experts from research and practice, professional representatives of pharmacists,

drug manufacturers, relevant organisations responsible for representing the interests of patients and self-help groups of chronically ill and disabled persons, as well as the Federal Government Commissioner for Patients' Affairs. This enables an open and independent reviewing procedure for the preliminary report. The topics addressed in the commenting procedure refer in particular to the results of the literature search and assessment presented in the preliminary report. At the same time, the opportunity is also provided to submit any type of document of appropriate quality (especially unpublished data), which is suited to answer the research question of the report, according to the persons submitting comments. If the search strategy defined in the report plan is restricted to RCTs, for example, non-randomised studies may nevertheless be submitted within the framework of the commenting procedure. However, in such cases, appropriate justification of the validity of the causal interpretation of the effects described in these studies is also required. Optionally, an oral scientific debate with those submitting comments may be held. This debate serves the potentially necessary clarification of aspects of the written comments and aims at improving the scientific quality of the final report. In order to avoid inappropriate delay of the Institute's work, the comments must fulfil certain formal requirements. Further information on the commenting procedure for the preliminary report is published on the IQWiG website in the relevant guideline (http://www.iqwig.de/download/Guideline_on_the_submission_of_comments_on_preliminary_reports_new.pdf). The conditions noted in the current version of this guideline apply.

The **final report**, which is based upon the preliminary report and contains the assessment of the scientific findings (considering the results of the hearing), represents the concluding product of the work on the commission. It is produced under the responsibility of the IQWiG project group, usually involving external experts. After completion of the internal quality assurance process and approval by the Institute's Steering Committee, the final report is initially forwarded to the contracting agency and subsequently (usually 4 weeks later) forwarded to the Foundation's Board of Trustees and Board of Directors (together with the documentation of the written comments, the meeting minutes of the [optional] scientific debate, and the appraisal of the comments). These documents (final report and documentation/appraisal of comments) are then published on the IQWiG website (usually a further 4 weeks later). If comments are received on final reports that contain substantial evidence not considered, or if the Institute receives information on such evidence from other sources, the contracting agency will be sent well-founded information as to whether, in the Institute's opinion, a new commission on the topic is necessary (report update) or not. The contracting agency then decides on the renewed commissioning of the Institute. The general methodological and procedural requirements for the Institute's products apply in the updating process.

2.1.2 *Rapid reports*

The production procedure for a **rapid report** is presented in Figure 3. All steps are performed under the responsibility of IQWiG, involving external experts where appropriate. If necessary, the Institute's Scientific Advisory Board is also involved. The internal quality assurance process is not presented in this flow chart.

The primary aim of a **rapid report** is to provide information at short notice on relevant health care developments, including new technologies, as well as on research questions not targeted at supporting decisions on directives of the Federal Joint Committee. A shorter preparation period is usually required. The production procedure for a rapid report differs from that of a full report in two main points:

1. No report plan or preliminary report is produced, and no hearing is performed. Working documents (project outline, preliminary rapid report) are not published.
2. The assessment is usually conducted on the basis of published information.

After **commissioning** by the Federal Joint Committee or Ministry of Health, the Institute's internal project group is formed under management of the department concerned. A project manager is then appointed. The composition of the project group is not fixed at this point, as changes may be necessary due to the subsequent steps taken. The research question is worded in consultation with the contracting agency's responsible committees, involving external professional expertise or individual affected persons or parties, if appropriate. As a rule, relevant patient organisations are involved, especially in the definition of patient-relevant outcomes. The project outline is subsequently prepared.

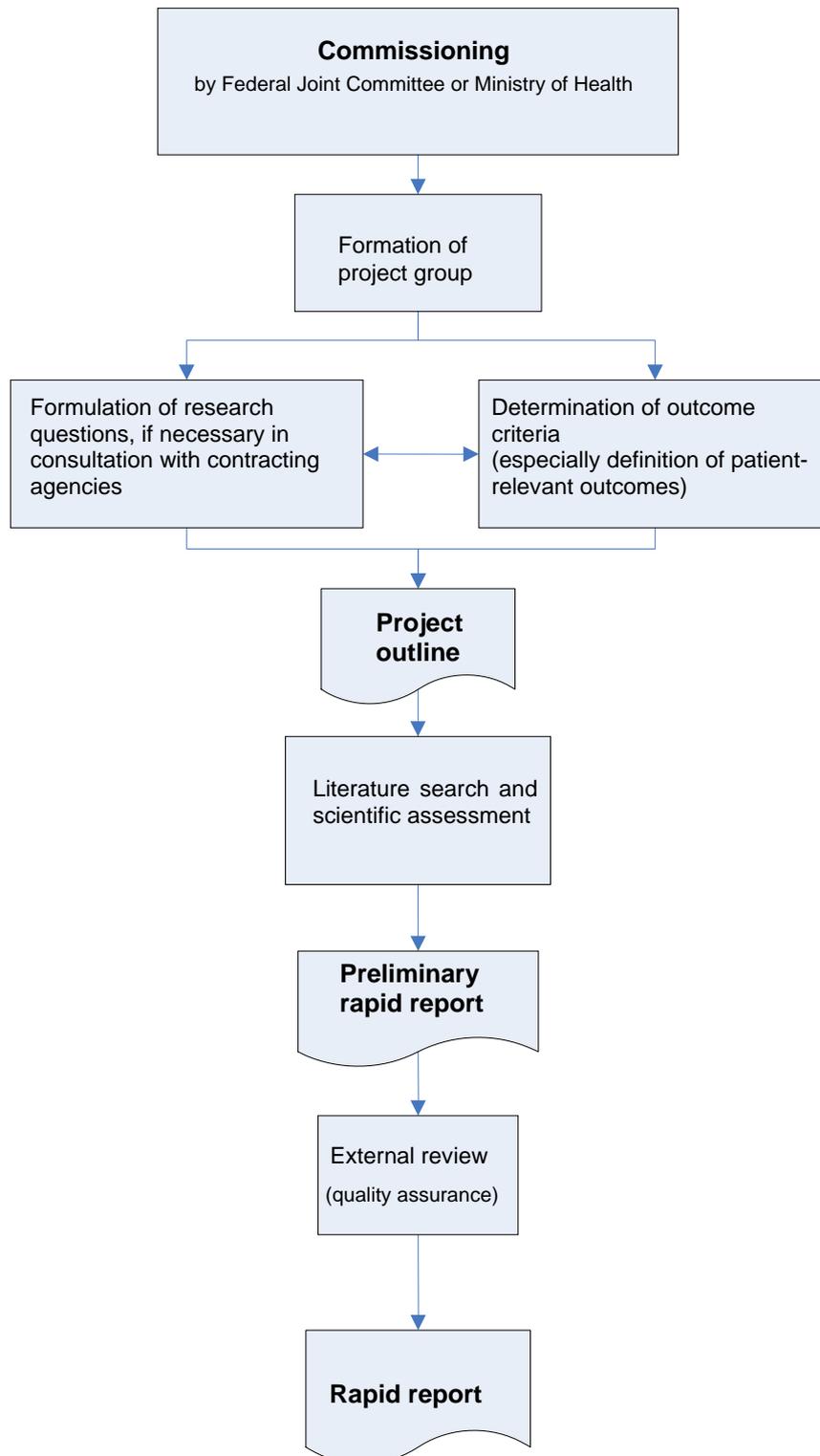


Figure 3: Production procedure for an IQWiG rapid report

The **project outline** summarises the main steps of the literature search and scientific assessment and is not published.

The preliminary **rapid report** is initially prepared under the responsibility of the project group, usually involving external experts. After completion of the internal quality assurance process, as an additional quality assurance step, the preliminary rapid report is reviewed by one or more external experts with recognised methodological and/or topic-related competence. Subsequently the final rapid report is prepared. After completion of the internal quality assurance process on this final rapid report and after approval by the Institute's Steering Committee, the document is then forwarded to the contracting agency. It is also forwarded to the Foundation's Board of Trustees and Board of Directors, and published on the IQWiG website (usually 4 weeks later). If comments on rapid reports are received that contain substantial evidence not considered, or if the Institute receives such evidence from other sources, the contracting agency will be provided with well-founded information as to whether, in the Institute's opinion, a new commission on the topic is necessary (rapid report update) or not. The contracting agency then decides on the renewed commissioning of the Institute. The general methodological and procedural requirements for the Institute's products apply in the updating process.

2.1.3 Health information

The Institute produces **health information** for patients and the general public in various formats, which are presented in more detail in Section 5. The Institute's main health information products include:

- **Feature articles:** comprehensive article which forms the basis of a set of related articles on an important health issue;
- **Fact sheets:** short, easily-understandable information (as commissioned, for example, by the Federal Joint Committee to explain directives);
- **Research summaries:** brief summaries of systematic reviews, health technology assessments or large studies, including summaries of the other products of the Institute, insofar as they are relevant for patients.

In addition to these core products, supplementary items are also produced. These aim to make the key messages of the health information more understandable and interesting.

The production process for health information is presented in Figure 4. The Institute maintains responsibility for all stages of the process, although it may also involve external expertise. The

internal review process is *not* outlined in the flow chart below. The process for the selection of health information topics on the Institute's own initiative is described in Section 5.3.1

The Institute's health information for patients and the general public is produced:

- in response to commissions received from the G-BA (Federal Joint Committee) or the German Ministry of Health;
- to summarise other products published by the Institute and as accompanying information for these products; and
- to fulfil its legislative responsibility to provide consumers with health information, as well as on its own initiative within the framework of the G-BA's general commission.

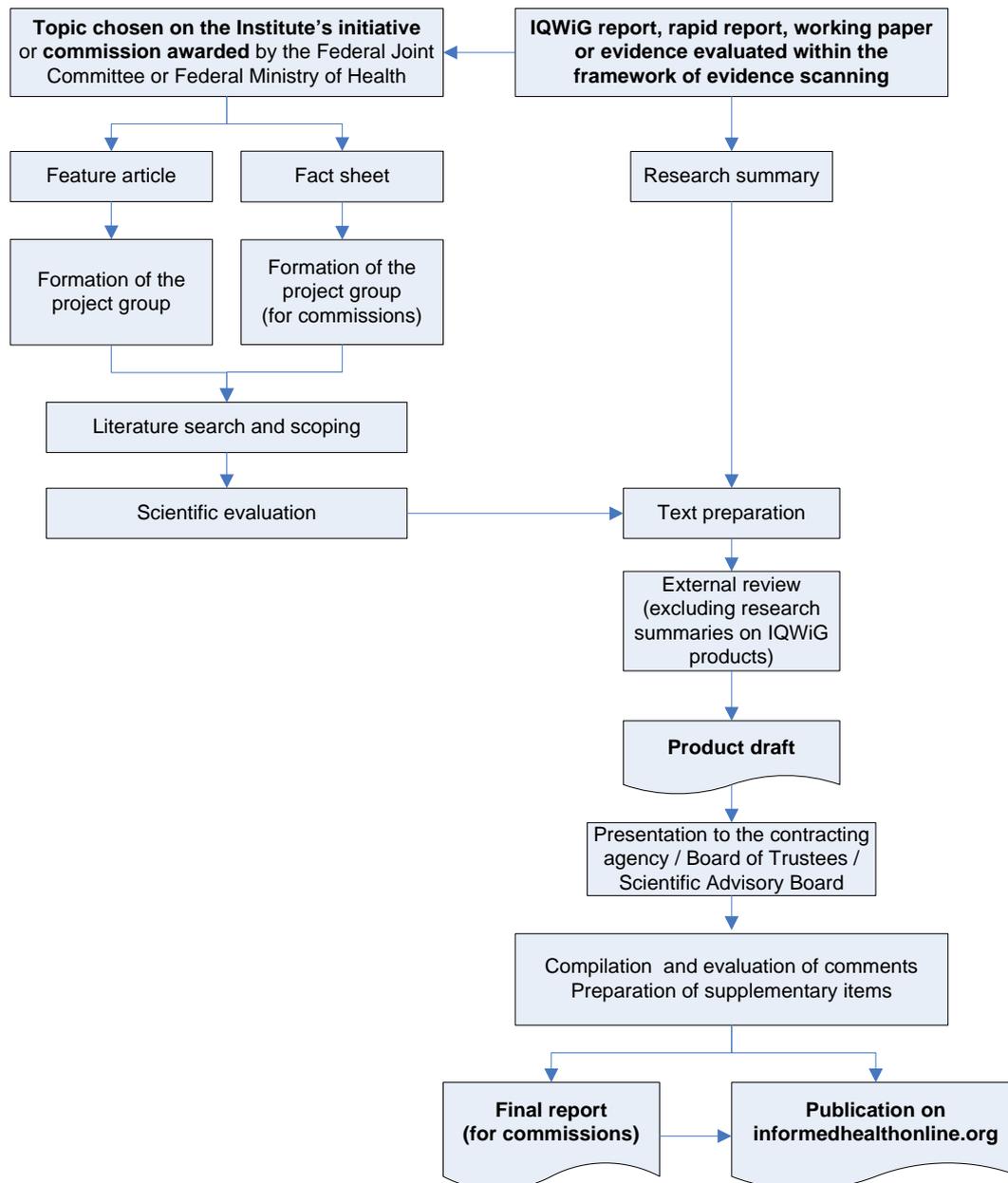


Figure 4: Procedure for the production of health information

The Institute's general commission (see Section 2.1.4) was amended in July 2006 to specifically include informing the general public. The process of evidence scanning, which the Institute undertakes to develop potential topics for this information, is described in Section 5.3.1. Section 5 also describes the methodology of literature searches, as well as patient involvement.

After **commissioning** by the Federal Joint Committee or Federal Ministry of Health, an internal project group is formed. A project manager is appointed. A project group includes at least one

IQWiG staff member who does not belong to the Health Information Department. A project group is also formed for each feature article.

The accompanying information for the Institute's reports and other products are developed in close cooperation with the project manager and the department responsible for the Institute's product. The departments are also consulted in relation to health information within their area of responsibility.

After the **text** has been prepared and a departmental quality assurance process undertaken, the drafts are sent out for **external review**. The drafts of "feature articles" and "fact sheets" are sent to at least one external reviewer in Germany and usually to at least one external reviewer abroad. "Research summaries" are sent to the author of the research that was summarised, as well as to at least one internal or external reviewer. The "research summaries" of the Institute's products are usually reviewed internally, but can also be sent to external reviewers (e.g. external experts) involved in the report production. All of the core products undergo patient or "user" testing by at least one person.

The "supplementary items" are subjected to the same internal review processes as the corresponding core products. If necessary, they may also be reviewed externally. Patient stories (see Section 5.4.3) are only published if written consent of the patient involved has been given.

After approval by the Steering Committee, the final draft of a health information product is sent to the commissioning agency and the Board of Trustees for **commenting** within a one-month consultation period. The Board of Trustees represents relevant stakeholders, including patient representatives. Unlike preliminary reports, patient information drafts are not published on the Institute's website.

The comments submitted during the consultation period are summarised and reviewed. The summaries and the comprehensive versions are also provided to the project group (if applicable) and to the Steering Committee.

With commissioned patient information, the commissioning agency is also provided with a report on the comments received during consultation. Final reports on commissioned patient information are subjected to the same publication procedures as other final reports. They are sent initially to the commissioning agency and then forwarded to the Foundation's Board of Directors and Board of Trustees, usually 4 weeks later. They are then published on the IQWiG website *www.iqwig.de* (usually a further 4 weeks later). The corresponding health information is subsequently published on the Institute's website for consumers and patients at *www.informedhealthonline.org*. The patient

information explaining directives by the Federal Joint Committee are published only after publication of the directives themselves on www.informedhealthonline.org.

Research summaries for the Institute's products are usually published at the same time as the final report on www.informedhealthonline.org.

Corrections, improvements, and updates of the Institute's health information are made internally. If more extensive updates of content are made, external experts may be involved. A more detailed description of the Institute's updating mechanisms is provided in Section 5.

2.1.4 Working papers

The production procedure for **working papers** is presented in Figure 5. All steps are conducted under the Institute's responsibility, in exceptional cases involving external expertise. If necessary, the Institute's Scientific Advisory Board may also be involved. The internal quality assurance process is not presented in this flow chart.

The preparation of working papers is conducted (among other things) within the framework of the **general commission** awarded by the Federal Joint Committee on 21.12.2004. According to this document, the Institute was commissioned "by means of documenting and analysing the relevant literature, continuously to study and assess medical developments of fundamental importance and their effects on the quality and efficiency of health care in Germany, and to relay its findings to the Federal Joint Committee on a regular basis. In this context, the Committee assumes that the Institute, according to the tasks assigned following §139a (3) SGB V, will not only work on individual commissions awarded by the Federal Joint Committee, but will also take on scientific projects on its own responsibility, and relay essential information on relevant health care developments to the Committee so that the Committee can fulfil its legal obligations. Against the background of this information, the Institute will also develop concrete proposals for individual commissions that it considers relevant." The need to conduct independent scientific projects therefore results from the Institute's legal remit and the general commission. This also includes projects on the further development of methods, which can also be published as working papers.

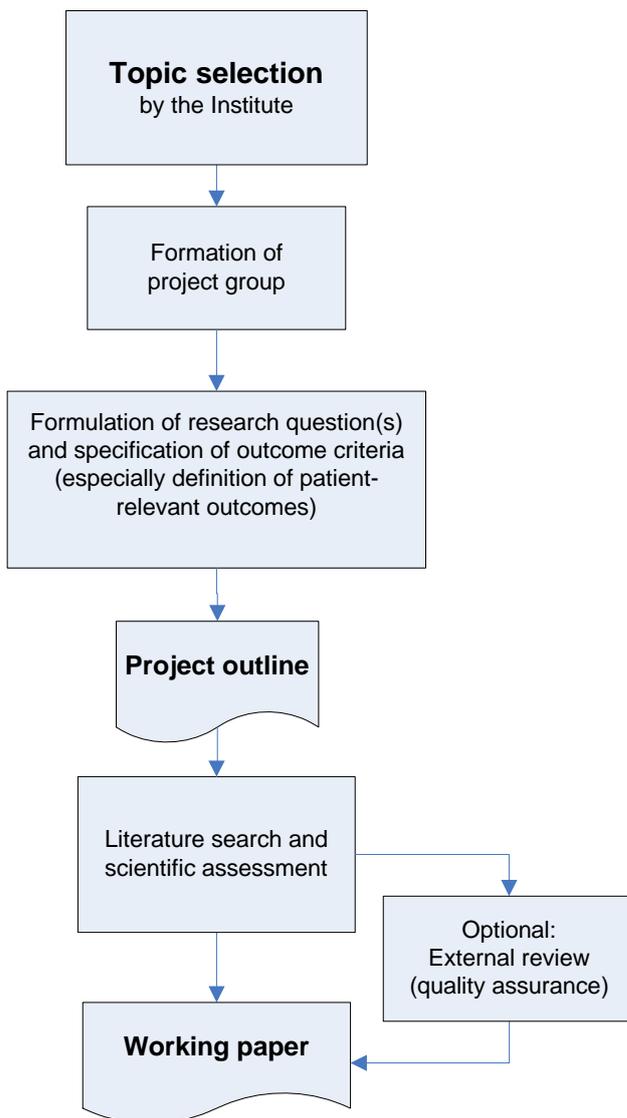


Figure 5: Production procedure for an IQWiG working paper

The **topic selection** is usually conducted by continuous screening of the scientific literature and other information sources by the Institute's departments. On the basis of the information retrieved, a potential topic for investigation is proposed to the Steering Committee, which then supports or rejects the proposal.

The **project outline** summarises the main steps of the literature search and the scientific assessment. In this context, relevant patient organisations may be involved, especially in the definition of patient-relevant outcomes. The project outline is not published.

The **working paper** is prepared under the responsibility of the Institute; external experts are only involved in exceptional cases. The quality assurance process can (optionally) include an external

review. After completion of the internal and (optional) external quality assurance process and approval by the Steering Committee, the document is sent to the Federal Joint Committee. The final version of the working paper is usually sent to the Foundation's Board of Directors and Board of Trustees 4 weeks later, and published on the IQWiG website. If comments on working papers are received that contain substantial unconsidered evidence, or if the Institute receives such evidence from other sources, the Institute assesses whether it considers it necessary to update the document or not. The general methodological and procedural requirements for the Institute's products apply in the updating process.

2.2 General aspects of the preparation of products

The following procedures and aspects that are valid for all products are presented in this chapter:

- Choice of external experts for cooperation in the preparation of products;
- Guarantee of scientific independence in the preparation of products;
- Review of products;
- Publication of products.

2.2.1 Choice of external experts

In accordance with its legal remit, the Institute involves external experts in its work. External experts are persons who are awarded research commissions within the framework of the preparation of the Institute's products or the review of these products.

The award of research commissions follows the Institute's own procurement regulations, which are based on the terms of §22 of the Regulation of Budgetary Matters in Social Insurance^p as well as the legal regulations of procurement law. The Institute announces in scientific journals that commissions to be awarded to external experts are published on its website on a regular basis (www.iqwig.de). Current commissions to be awarded are listed under <http://www.iqwig.de/index.174.en.html>. Commissions with a volume above the current threshold value of the procurement regulations of the European Union (EU) are advertised throughout the EU.

^p Verordnung über das Haushaltswesen in der Sozialversicherung (SVHV)

The specific requirements regarding the suitability of applicants are published in the corresponding announcements or tendering documents. The Institute has established a tendering office to ensure proper conduct of the tendering procedure.

2.2.2 *Guarantee of scientific independence*

The scientific independence of the Institute and of the products it is responsible for and publishes have their legal basis in §139a SGB V, as well as in the Charter of the Foundation.

A) *Guarantee of the independence of external experts*

Before a contract is signed between the Institute and an external expert or institution with regard to the preparation of a product, it must be decided whether any reservations exist as to potential conflicts of interest. For this purpose, all external experts and institutions must disclose all activities that may *potentially* influence their scientific independence. In accordance with the relevant guidelines of scientific journals, the following criteria are particularly viewed as conflicts of interest: all financial agreements, employment, consultations, fees, reimbursed expert opinions, reimbursed travel expenses, patent applications, and share ownership within the previous three years that could influence the commission in question, as well as all current existing personal contacts with other persons or organisations that could influence the commission in question [214]. A list of these criteria is also included in a form published on the Institute's website, which is updated whenever necessary. The downloadable version of the form on the website always applies. The names of external experts involved in the preparation of reports or rapid reports are usually published in these reports. If specifically requested by external experts or the contracting agency, or if other relevant reasons exist, it is possible not to publish the external experts' names in order to ensure their independence and prevent interest-driven attempts to influence them. Corresponding stipulations have been made for the work of the Federal Joint Committee (§ 35 [1] S. 8 SGB V). If the names of the external experts are published, any potential conflicts of interest they disclose are also published. This is presented in such a way that it is stated whether or not any potential conflicts of interest were reported for each criterion listed in the "Form for disclosure of potential conflicts of interest". Other details, for example, concerning the amount of any financial remuneration received, will not be published. The selection process for external experts is described in Section 2.2.1.

B) *Guarantee of internal scientific independence*

Internal scientific independence is guaranteed as far as possible by the selection of staff. When appointed, staff must outline their previous activities in a credible manner and, within the framework of their employment in the Institute, are obliged to cease all (external) assignments

likely to call their scientific independence into question. The Institute's scientific staff is prohibited from performing paid external assignments that could in the broadest sense be associated with their professional duties. As a matter of principle, all external assignments must be declared by all members of staff to the Institute's Management or the Department of Administration. External assignments in the broadest sense also include unpaid honorary positions such as positions on boards or in organisations and societies. In individual cases, violations may lead to a reprimand or, in recurrent or serious cases, to dismissal. The Institute's Management, after consultation with the Steering Committee, will decide on a case-by-case basis whether a member of staff should be excluded from a certain activity or project on grounds of a suspected conflict of interest.

2.2.3 Review of the Institute's products

The review of the Institute's products aims in particular at ensuring their high scientific quality. Other aims may be relevant for individual products, such as ensuring the product's comprehensibility for the general public.

All products (including interim ones) are subjected to a comprehensive multi-stage internal quality assurance process. In addition, in the preparation of reports, rapid reports, and some health information products, an external review is performed as a further quality assurance step. An external review is optional for working papers and certain health information products (see Section 2.1.3). The choice of internal and external reviewers is primarily made on the basis of their methodological and/or professional expertise.

The issue of the effectiveness of various external review procedures has only recently been investigated in specific studies. However, very few relevant studies on this topic are available [135,300]; in particular, there is a lack of sufficiently valid intervention studies. No evidence is available that one particular approach is better than another. For example, it is unclear how many people should be involved in a review. According to the studies available, lack of clarity exists as to the relevance of conventional procedures applied in medical journals [135,215], including evaluation by consumers and patients [32]. Various methods exist to assess the quality of individual reviews. Here too, however, no evidence is available as to whether one method is superior, and if so, which one.

The identification and choice of potential external reviewers depend on the extent of the review commissioned. Comprehensive external reviews can also be commissioned as research projects, whereby the conditions outlined in Section 2.2.1 apply. External reviewers can also be identified by a literature search, expertise of the project group, by contacting scientific societies, or by an

application during the tendering procedure for a commission, etc. In each case, potential conflicts of interest must be disclosed.

External reviewers are selected by the project group and other persons involved in the product preparation. The number of reviewers is not limited. The external reviews are evaluated with regard to their relevance for the corresponding product. The external reviewers' reports are not published. As with the procedure for external experts, the names of the external reviewers are usually published in the final reports and rapid reports, including a disclosure of potential conflicts of interest.

In addition to the external quality assurance process described above, which involves reviewers selected and commissioned by the Institute, an open and independent review process is ensured by the publication of the Institute's products together with the opportunity to submit comments.

2.2.4 Publication of the Institute's products

One of the Institute's main tasks is to determine the available evidence on a topic by performing a careful assessment of the information available, and to publish the results of this assessment. It is legally specified that the Institute "must at regular intervals publicly report on its working processes and results, including the bases for decision-making" (§139a [4] SGB V).

To maintain the Institute's independence, it must be ruled out that the contracting agencies or any other interested third parties can exert any influence on the content of the reports. This could lead to conflation of scientific findings with political and/or economic aspects and/or interests. At the same time, it must be avoided that the Institute itself withholds certain findings. All the results obtained by the Institute within the framework of its legal responsibilities are therefore published as soon as possible. In the case of reports, this also includes the report plan. Product-specific features are noted in the sections in which procedures are described. In justified exceptional cases, timelines may deviate from the stipulated norm (period between completion and publication of a document).

Unless otherwise agreed, all copyright is held by the Institute.

3 Assessment of the benefit and harm of medical interventions

3.1 Patient-relevant medical benefit

3.1.1 Definition of a patient-relevant medical benefit

To define a patient-relevant medical benefit, first of all, it is meaningful to distinguish between the terms “necessity” and “benefit”. According to §27 SGB V, SHI-insured persons are entitled to a medical intervention if this intervention is necessary to diagnose a disease, cure it, prevent its worsening, or alleviate its symptoms.

This refers to key aspects of the benefit of an intervention, which fall under the superordinate terms “mortality”, “morbidity” and “quality of life (QoL)”. However, the term “necessity” goes beyond the term “benefit”. The demonstration of the benefit of an intervention is thus a necessary (but not a sufficient) requirement for the demonstration of the necessity of an intervention. As a consequence, an intervention for which no benefit has been proven cannot be described as being necessary.

The term “benefit assessment” refers to the whole process of the assessment of medical interventions with regard to their positive and negative causal effects compared with a clearly defined alternative active treatment, a placebo (or a different sham intervention), or no treatment. In this context, “causal” means that it is sufficiently certain that the observed effects can be ascribed solely to the intervention to be assessed [381].

The term “benefit” refers to the positive causal effects, and the term “harm” refers to the negative causal effects of a medical intervention on patient-relevant outcomes (see below). The description of benefit or harm is always performed based on the intervention to be assessed. An intervention’s benefit or harm is determined by way of comparison with a placebo (or another sham intervention) or no treatment. In the case of a comparison between a medical intervention to be assessed and a clearly defined alternative (active) medical intervention, the following terms are used in the comparative assessment of beneficial or harmful aspects:

- Beneficial aspects:
 - In the case of a greater benefit, the term “additional benefit” is used;
 - In the case of a lower or comparable benefit, the terms “lower” or “comparable” benefit are used.

- Harmful aspects:
 - The terms “greater”, “comparable” and “less” harm are used; the term “additional harm” should be avoided.

If possible, a weighing of the benefit and harm of an intervention is conducted (see Section 3.1.4).

The assessment of the evidence should come to a clear conclusion that either:

1. There is proof of a(n) (additional) benefit or harm of an intervention, or
2. There is proof of a lack of a(n) (additional) benefit or harm, or
3. There is no proof of a(n) (additional) benefit or harm or the lack thereof, and it is therefore unclear whether the intervention results in a(n) (additional) benefit (or harm) for the patient.

The assessment of the medical necessity of an intervention is based on (a) the type and extent of any (additional) benefit proven and also, if appropriate, on (b) whether this (additional) benefit can be achieved by the intervention to be assessed alone.

As the benefit of an intervention should be related to the patient, this assessment is based on the results of studies that have investigated the effects of an intervention on patient-relevant outcomes. In this context, “patient-relevant” refers to how a patient feels, functions or survives [44]. In this context, consideration is given to both the intentional and unintentional effects of the intervention that allow an assessment of the impact on the following patient-relevant outcomes, in order to determine the changes related to disease and treatment:

1. Mortality,
2. Morbidity (complaints and complications),
3. Health-related QoL.

In addition, invested time and effort related to the disease and the intervention can be considered, as well as treatment satisfaction of patients. However, these aspects are normally regarded only as secondary outcomes.

According to §35b SGB V, the following outcomes related to patient benefit are to be given particular consideration: increase in life expectancy, improvement in health status and QoL, as well as reduction in disease duration and adverse effects. These dimensions of benefit are represented by the outcomes listed above: for example, the improvement in health status and the reduction in disease duration are aspects of direct disease-related morbidity; the reduction in adverse effects is an aspect of therapy-related morbidity.

Those outcomes that reliably and directly represent specific changes in health status are primarily considered. In this context, in particular individual affected persons as well as organisations of patient representatives and/or consumers are involved in the topic-related definition of patient-relevant outcomes. In the assessment of QoL and patient satisfaction, only instruments should be used that are suited for application in clinical trials and have been evaluated accordingly [122,365]. In addition, valid surrogate outcomes can be considered in the benefit assessment.

Both beneficial and harmful aspects can be of different relevance for the persons affected; these aspects may become apparent through qualitative surveys or through IQWiG's consultations with affected persons and organisations of patient representatives and/or consumers in connection with the definition of patient-relevant outcomes. In the summarising assessment of beneficial and harmful aspects, a different weighting of outcomes can therefore be performed. Consequently, the precondition for a comparative appraisal of specific outcomes (with lower weighting) is the comparative appraisal of other outcomes (with higher weighting). For example, a lower adverse event rate of the test intervention can only mean an additional patient benefit if, at the same time, outcomes with a higher weighting are beneficially affected by the test intervention to a similar (or acceptably smaller) extent compared with the comparator intervention. However, this requires an a priori definition of what can still be regarded as "acceptably smaller" (irrelevance range; see Section 6.4.5).

Diagnostic measures can be of indirect benefit by being a precondition for therapeutic interventions through which the achievement of an effect on the patient-relevant outcomes outlined above is made possible. The precondition for the benefit of diagnostic tests is therefore the existence and the proven benefit of the treatment for patients, depending on the test result. In addition, diagnostic tests can enable patient-relevant personal decisions and may therefore also be of benefit.

Interventions can also have consequences for those indirectly affected, for example, relatives and nursing staff. If appropriate, these consequences can also be considered within the framework of IQWiG reports.

3.1.2 Surrogates of a patient-relevant medical benefit

Surrogate outcomes are frequently used in medical research as a substitute for patient-relevant outcomes, mostly to obtain conclusions on patient-relevant (additional) benefits earlier and more simply [13,134,299]. Most surrogate outcomes are, however, unreliable in this regard and can be misleading when used in a benefit assessment [153,161]. As a rule, in the Institute's benefit assessments, surrogate outcomes are therefore considered only as proof of a(n) (additional) benefit of an intervention if appropriate statistical methods applied beforehand showed that the effect of an

intervention (with a comparable mechanism of action) on the patient-relevant outcome to be substituted was explained to a sufficient degree by the effect on the surrogate outcome [26,378]. For this purpose, clear proof is normally required from intervention studies of a plausible, strong, consistent, and unidirectional association between the change in the surrogate outcome and the change in the patient-relevant outcome. A unidirectional association means that a positive or negative change in the surrogate outcome is associated with a positive or negative change in the patient-relevant outcome.

The validity of a surrogate outcome is regarded as not proven if no relevant studies are available describing an association between the change in the surrogate outcome and the change in the corresponding patient-relevant outcome. In addition, a surrogate outcome is not seen as valid if it has been shown in studies that an intervention:

- Had an effect on the surrogate outcome, but not on the patient-relevant outcome, or
- Had an effect on the patient-relevant outcome, but not on the surrogate outcome, or
- Produced inconsistent effects on the surrogate outcome and on the patient-relevant outcome.

Associations noted between a surrogate outcome and the corresponding patient-relevant outcome regarding an intervention with a specific mechanism of action are not necessarily applicable to interventions that are used for the treatment of the same disease, but have a different mechanism of action [153,161].

Surrogate outcomes of unclear or controversial validity may be presented in the Institute's reports. However, such outcomes are not suitable to provide proof of the (additional) benefit of an intervention. In the case of extremely serious diseases in terms of morbidity and mortality without treatment alternatives, surrogate outcomes of unclear validity may have to be accepted as outcomes that potentially indicate a benefit of an intervention.

Depending on the proximity to the corresponding patient-relevant outcome, various other terms for surrogate outcomes are used in the literature (e.g. intermediate outcome). Such a distinction is dispensed with here, as it does not affect the issue of the validity required for surrogate outcomes.

3.1.3 Assessment of the harm potential of medical interventions

The use of any type of medical intervention (drug, non-drug, surgical, diagnostic, preventive, etc.) carries the risk of adverse effects. The term "adverse effects" refers to all events and effects that represent individually perceived or objectively detectable physical or mental harm that may cause a short- or long-term reduction in life expectancy, an increase in morbidity, or a more or less severe

impairment in QoL. If the term “adverse effects” is used, a causal relationship to the intervention is assumed, whereas the issue of causality still remains open with the term “adverse events” [79].

The term “harm potential” describes the risk of the occurrence of adverse effects when using a medical intervention. The description of the harm potential is an essential and equal component in the benefit assessment of an intervention. It ensures an informed, population-related, but also individual weighing of benefit and harm [390]. A prerequisite for this is that the effect sizes of a medical intervention can be described from the data available, both for its desired as well as for its adverse effects, and compared with therapy alternatives, for example.

However, in a systematic review, the analysis, assessment, and reporting of the potential harm of a medical intervention are often far more difficult than those of the (additional) benefit. In particular, this refers to unexpected adverse events [79]. Studies are typically designed to measure the effect of a medical intervention on a few predefined effectiveness outcomes. In such studies, results with regard to adverse events depend greatly on the underlying methodology on how these events were recorded [41,206]. In addition, studies designed to specifically detect rare, serious adverse effects (including the description of a causal relationship to the medical intervention) are considerably underrepresented in medical research [48,109,205]. Furthermore, reporting of adverse events in individual studies is of poor quality, which recently led to an amendment to the CONSORT⁹ statement on RCTs [204]. Finally, the systematic assessment of adverse effects of an intervention is also made difficult by the fact that the corresponding coding in literature databases is insufficient; the specific search for relevant scientific literature therefore often produces an incomplete picture [88].

The consequence of the obstacles noted above is that in many cases, despite enormous efforts, the uncertainty of conclusions on the harm potential of an intervention will be greater than that of conclusions on positive effects [252]. It is necessary to find a meaningful balance here between the completeness of the evaluation and the resources invested. Consequently, it is necessary to limit the evaluation and reporting to relevant adverse effects. In particular, adverse effects can be defined as relevant that may:

- Completely or almost completely counterbalance the benefit of an intervention;
- Substantially differ from adverse effects occurring with (an) otherwise equivalent treatment option(s);
- Occur predominantly with treatment options that may be particularly effective;

⁹ Consolidated Standards of Reporting Trials

- Have a dose-effect relationship;
- Be regarded by patients as especially important;
- Be accompanied by serious morbidity or even increased mortality, or be associated with substantial impairment in QoL.

In the interests of patient safety and the medical axiom “primum nil nocere”, the Institute observes the following principles when assessing and reporting adverse effects:

- In the benefit assessment, the initial aim is to compile a selection of potentially relevant adverse effects that are essential in making the decision for or against the use of the intervention to be assessed. In this context, the selection of adverse effects and events is made in accordance with the criteria outlined above. This compilation is made within the framework of the preliminary literature search regarding the research question posed, especially on the basis of data from controlled intervention studies in which the benefit of the intervention was specifically investigated. In addition, and if appropriate, the compilation is made on the basis of available epidemiological data (e.g. from cohort or case-control studies), as well as pharmacovigilance and regulatory data, etc. In individual cases, data obtained from animal trials and experiments to test pathophysiological constructs may be useful.
- The compilation of potentially relevant adverse effects described above forms the foundation for the assessment of the potential harm of an intervention on the basis of the studies included in the benefit assessment.

3.1.4 Summarising assessment

The benefit assessment and the assessment of the extent of the (un)certainty of results (see Section 3.2.1) generally follow international EBM standards as developed, for example, by the GRADE^r group [22,166,328].

Medical interventions are compared with other clearly defined active or sham interventions (e.g. placebo), or with no intervention in respect of their (additional) beneficial and harmful effects on defined patient-relevant outcomes, and are subsequently summarised. For this purpose, exactly one of the five following evaluating conclusions is first drawn for each predefined patient-relevant outcome on the basis of the analysed scientific data available:

^r Grading of Recommendations, Assessment, Development and Evaluation

1. Proof of a(n) (additional) benefit or harm exists.
2. Indications of a(n) (additional) benefit or harm exist.
3. Proof of the lack of a(n) (additional) benefit or harm exists.
4. Indications of the lack of a(n) (additional) benefit or harm exist.
5. No proof and no indication of a(n) (additional) benefit or harm exist.

Conclusion 1 is drawn if scientific evidence of a(n) (additional) benefit or harm exists. Conclusion 2 is drawn if indications of a(n) (additional) benefit or harm exist, but these indications do not fulfil the requirements for scientific evidence. Conclusion 3 is drawn if scientific evidence exists that this intervention is not associated with a(n) (additional) benefit or harm. Conclusion 4 is drawn if indications of the lack of a(n) (additional) benefit or harm exist, but these indications do not fulfil the requirements for scientific evidence. Well-founded definitions of irrelevance ranges are the precondition for indications or proof of the lack of a(n) (additional) benefit or harm (see Section 6.4.5). Conclusion 5 is drawn if no proof or indications exist of a(n) (additional) benefit or harm, for example, due to insufficient or inconsistent data.

As a rule, if the conclusion is drawn that “proof” is available, it is required that a meta-analysis of studies shows a corresponding statistically significant effect (with outcome-related minor uncertainty of results). If a meta-analysis is not feasible, at least 2 studies conducted independently of one another should be available that show outcome-related minor uncertainty of results and a statistically significant effect, and whose results are not questioned by further comparable studies with outcome-related sufficient certainty of results (“consistency of results”). The two studies conducted independently of one another need not necessarily be of exactly identical design. Which deviations in design between studies are still acceptable depends on the research question posed. Despite showing statistically significant effects, as a rule a meta-analysis of studies with outcome-related high uncertainty of results or results from individual studies can consequently at most provide indications of the effects of an intervention. If, in exceptional cases, proof of the benefit of an intervention is inferred from only one study, then specific requirements apply to this study and its results [76].

These conclusions, drawn separately for each patient-relevant outcome, are then summarised (as far as possible) in one evaluating conclusion in the form of a weighing of benefits and harms. If proof of a(n) (additional) benefit and/or harm exists with regard to outcomes 1-3 from Section 3.1.1, the Institute describes (insofar as possible on the basis of the data available):

1. A benefit potential,

2. A harm potential, and
3. A weighing of benefits and harms.

In this context, the Institute will follow the principle of risk prevention, i.e. if in doubt and depending on the relevant context, assume that a harm potential exists. Furthermore, factors related to age, gender, and personal circumstances will be considered.

The precise conduct of the weighing of benefits and harms is topic-specific and should, if this is possible prospectively, be described in the report plan (protocol), or otherwise in the preliminary report. In addition to the weighing of benefits and harms of an intervention by means of a comparison of the benefit and harm potential, it is possible to perform a comparative weighting of benefit and harm. In this case, IQWiG's conclusions would be reported for each patient-relevant outcome through the weighting of benefit and harm (e.g. using a summarising score). The weighting of benefit and harm is also specific to a therapeutic area, and how they are weighted should be defined prospectively at the time the outcomes to be investigated are selected.

3.2 Specific aspects of the benefit assessment

3.2.1 (Un)certainty of results

In principle, every result of an empirical study or a systematic review of empirical studies is uncertain. In this context, one distinguishes between qualitative and quantitative uncertainty of results. Qualitative uncertainty is determined by the study design, from which evidence levels can be inferred (see Section 6.2.2), as well as by (outcome-related) measures to further avoid or minimise potential bias, which needs to be assessed depending on the study design (see Section 6.2.3). These measures include, for example, blinded outcome assessment, analysis on the basis of all included patients (if necessary, using appropriate replacement methods for missing values), and (if applicable) the use of appropriate and valid measurement instruments. In addition to the qualitative uncertainty of results, measurable quantitative uncertainties exist due to statistical principles, which in turn are directly related to the sample size, i.e. the number of patients investigated in a study or the number of (primary) studies included in a systematic review, as well as to intra- and inter-study variability. If the underlying data allow, statistical uncertainty can be quantified and evaluated as the standard error or the confidence interval of parameter estimates (precision of the estimate).

The appraisal of both the qualitative and quantitative certainty of results, as well as the size of the effects observed (and their consistency) form the basis of the recommendations to be inferred and of their grading. For example, evidence from studies with a lower qualitative certainty of results

can be upgraded by large estimated effects. Conversely, a low quantitative certainty of results, i.e. few available data (studies), can lead to downgrading despite a possible high qualitative certainty of results [328].

3.2.2 *Dramatic effect*

If the course of a disease is predictable with sufficient certainty, and no treatment options are available to influence this course, then proof of a benefit of a medical intervention can also be provided by the observation of a reversal of the (more or less) deterministic course of the disease in well-documented case series. If, for example, it is known that in many cases a disease leads to death within a short time after diagnosis, and it is described in a case series that, after application of a specific intervention, most of those affected survive for a longer period of time, then this “dramatic effect” may be sufficient to provide proof of a benefit. An example for such an effect is the substitution of vital hormones in diseases with a failure of hormone production (e.g. insulin therapy in patients with diabetes mellitus type 1). An essential prerequisite for classification as a “dramatic effect” is sufficiently reliable documentation of the fateful course of the disease in the literature and of its diagnosis in the patients included in the case series. Possible harms of the intervention should also be taken into account. Glasziou et al. recently presented an approach to operationalise the classification of an intervention as a dramatic effect, and proposed regarding an observed effect as not explicable solely by the impact of confounding factors if it was significant at a level of 1% and, expressed as the relative risk, if it exceeded the value of 10 [147].

If, in the run-up to the work on a specific research question, sufficient information is available indicating that a dramatic effect caused by the intervention to be assessed can be expected (e.g. because of a preliminary literature search), then studies will also be sought in the literature search that show a higher uncertainty of results because of their design.

3.2.3 *Study duration*

Study duration is an essential criterion in the selection of studies relevant to the benefit assessment. In the assessment of a therapeutic intervention for acute diseases where the primary objective is, for example, to shorten disease duration or alleviate acute symptoms, it is not usually meaningful to call for long-term studies, unless late complications are to be expected. On the other hand, in the assessment of therapeutic interventions for chronic diseases, short-term studies are not usually suitable to achieve a complete benefit assessment of the intervention. This especially applies if treatment is required for several years, or even lifelong. In such cases, studies that cover a treatment period of several years are particularly meaningful and desirable. As both beneficial and

harmful aspects can be distributed differently over time, in long-term interventions a meaningful comparison of benefits and harms is only feasible with sufficient certainty if the study duration is long enough. Individual aspects of the benefits and harms of an intervention may quite well be investigated in short-term studies.

With regard to the selection criterion “minimum study duration”, the Institute primarily follows standards for demonstrating the effectiveness of an intervention. In the assessment of drugs, the Institute will in particular resort to information provided in the guidelines specific to therapeutic areas, which are published by regulatory authorities (e.g. [119]). As the benefit assessment of an intervention also includes aspects of its harm potential, the generally accepted standards in this respect are also relevant when determining the minimum study duration. Furthermore, for long-term interventions as described above, the Institute will resort to the relevant guidelines for the criterion “long-term treatment” [201]. In individual cases, the Institute may deviate from this approach (and will justify this deviation), for example, if a topic requires longer follow-up, or if specific (sub-) questions apply to a shorter period. Such deviations may also be indicated if short-term effects are a subject of the assessment (e.g. in the assessment of newly available/approved interventions and/or technologies where no appropriate treatment alternative exists).

3.2.4 Patient-reported outcomes

The patient-relevant dimensions of benefit outlined in Section 3.1.1 can also include patient-reported outcomes (PROs). In addition to health-related QoL and treatment satisfaction, PROs can also cover other dimensions of benefit, for example, disease symptoms. As in the assessment of QoL and treatment satisfaction, instruments are required that are suitable for use in clinical trials [122,365]. In the selection of evidence (especially study types) to be considered for the demonstration of an effect, the same principles as with other outcomes usually apply [365]. This means that also for PROs (including health-related QoL and treatment satisfaction), RCTs are best suited to demonstrate an effect.

As PROs are 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 can be inferred for PROs from open studies. Empirical evidence shows a high bias potential for subjective outcomes in open studies [386]. This should be considered in the interpretation of these studies (see also Section 6.2.3). However, situations are conceivable where blinding of physicians and patients is not possible. In such situations, if possible, other efforts are required to minimise and assess bias (e.g. blinded documentation and

assessment of outcomes). Further aspects on the quality assessment of studies investigating PROs are outlined in [365].

3.2.5 *Benefit in small populations*

In small populations (e.g. patients with rare diseases or special subgroups of patients with common diseases), there is no convincing argument to deviate in principle from the hierarchy of evidence levels. Patients with rare illnesses also have the right to the most reliable information possible on treatment options [117]. Non-randomised studies require larger sample sizes than randomised ones because of the need of adjustment for confounding factors. Due to the rarity of a disease it may sometimes be impossible to include enough patients to provide the study with sufficient statistical power. A meta-analytical summary of smaller studies may be particularly meaningful in such cases. Smaller samples generally result in lower precision of an effect estimate, accompanied by wider confidence intervals. Because of the relevance of the assumed effect of an intervention, its size, the availability of treatment alternatives, and the frequency and severity of potential therapy-related harms, for small sample sizes it may be meaningful to accept a higher p -value than 5% (e.g. 10%) for the demonstration of statistical significance, and therefore to increase quantitative uncertainty. Similar recommendations are made for other problematical constellations [121]. Such an approach must, however, be specified a priori and well justified. Likewise, for small sample sizes it may also be necessary to substitute a patient-relevant outcome that occurs too rarely with surrogate outcomes. However, these surrogate outcomes must also be valid for small sample sizes [120].

In the case of extremely rare diseases or very specific disease constellations, the demand for (parallel) comparative studies may be inappropriate [383]. Nevertheless, in such cases it is also possible to at least document and assess the course of disease in such patients appropriately, including the expected course without the application of the intervention to be assessed (e.g. using historical patient data) [62]. It is also specified and highlighted in the report plan that a situation is being assessed involving an extremely rare disease or a very specific disease constellation.

3.2.6 *Benefit in individual cases*

The aim of a benefit assessment is to make robust predictions for future patients using results of studies (with more or less large sample sizes) suited to demonstrate causal effects. Experiences based on individual cases (except for specific situations, e.g. dramatic effects), cannot be applied to future (different) patients, as no demonstration of causality is possible.

However, under certain conditions, single patient trials are possible that allow predictions based on causal effects precisely for this individual case (but not for other future patients). Single patient trials are common in practical medicine but are mostly conducted non-systematically; their interpretation is therefore often unreliable, due to the usually poor control for confounding factors [324].

Single patient trials (or “n-of-1” trials) include only one patient, and the outcomes must be completely and relatively quickly reversible [167,170,219,324]. In order to conduct an n-of-1 trial appropriately, reversibility is required, which itself requires the existence of a chronic disease or symptoms. If such a study is possible and well performed, it allows conclusions to be drawn as to whether a particular patient profits from a particular treatment. Such n-of-1 trials consist of several “test and control” study periods, which are applied in random succession in a patient [170,324]. In this type of study design, the treatment periods, rather than the patients, are randomised. The intervention should be blinded and include either an active control or placebo control. Unfortunately, the methodological quality of these trials is often insufficient, so that their findings must be interpreted with caution. The use of a non-blinded design requires a comprehensible justification.

3.3 Benefit assessment of drugs

One main objective of the benefit assessment reports is to support the Federal Joint Committee’s decisions on directives concerning the reimbursement of drugs by the SHI. For this purpose, it is necessary to establish whether a drug’s benefit has been demonstrated (or whether, when compared with a drug or non-drug alternative, a higher benefit [additional benefit] has been demonstrated).

The Federal Joint Committee’s decisions on directives do not usually consider particular cases, but the general case. Consequently, the Institute’s reports do not usually refer to decisions on particular cases.

Because of the objective of the Institute’s benefit assessments, these assessments only include studies with an evidence level principally suited to demonstrate a benefit of an intervention. Thus, studies that can only generate hypotheses are not generally relevant for the benefit assessment. The question as to whether a study can demonstrate a benefit of an intervention mainly depends on the certainty of results of the data analysed.

3.3.1 Relevance of the drug approval status

The commissioning of the Institute by the Federal Joint Committee to assess the benefit of drugs usually takes place within the framework of approval of the drug to be investigated (therapeutic area, dosage, contra-indications, concomitant treatment, etc.). For this reason, the Institute's recommendations to the Federal Joint Committee, which are formulated in the conclusions of the benefit assessment report, usually refer to the use of the assessed drug within the framework of the current approval.

It is clarified on a project-by-project basis how to deal with studies (and the evidence inferred from them) that were not conducted according to the use of a drug as outlined in the approval. In principle, it is conceivable that studies in which a drug was used outside the scope of the approval described in the Summary of Product Characteristics ("off-label use"), over- or underestimated a drug's benefit and/or harm. This may lead to a false assessment of the benefit and/or harm in patients treated within the framework of the drug's approval. However, if it is sufficiently plausible or has even been demonstrated that the results obtained in these studies are applicable to patients treated according to the drug's approval, these results can be considered in the benefit assessment.

Therefore, for studies that were excluded from the assessment only because they were off-label studies (or because it was unclear whether they fulfilled the requirements of the approval status), each case is assessed to establish to what extent the study results are applicable to patients treated according to the approval requirements.

Results from off-label studies are regarded as "applicable" if it is sufficiently plausible or has been demonstrated that the effect estimates for patient-relevant outcomes are not considerably affected by the relevant characteristic of the drug approval (e.g. pre-treatment required). As a rule, the equivalence of effects should be proven with appropriate studies. These studies should be targeted towards the demonstration of equivalence of the effect between the group with and the group without the characteristic. Results applicable to patients treated according to a drug's approval can be considered in the conclusion of the assessment.

Results from studies are to be regarded as "not applicable" if their applicability has not been demonstrated and if plausible reasons against the transferability of results exist. 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 range or severity, if off-label combinations including other substances were used, or if studies were conducted in patients with contra-indications for the intervention investigated. The results of these studies are not presented in the reports, as they cannot be considered in the assessment of the drug.

If results from off-label studies are regarded as applicable, this is specified in the report plan. Independent of the applicability of study results to the use specified in the approval of the drug, as a rule the results of studies showing the following characteristics are discussed:

- They refer to patients with the disease specified in the commission;
- They refer to patients treated with the drug to be assessed;
- They are of particular relevance due to factors such as sample size, study duration, or outcomes investigated.

3.3.2 Studies on the benefit assessment of drugs

The results of the Institute's benefit assessment of drugs may have an impact on patient health care in Germany. For this reason, high standards are required regarding the certainty of results of studies included in the benefit assessment.

The certainty of results is defined as the certainty with which an effect (or the lack of an effect) can be inferred from a study. This refers to both "positive" aspects (benefit) as well as "negative" aspects (harm). The certainty of results of an individual study is essentially influenced by three components:

- The study design;
- The internal validity (which is design-specific and determined by the concrete way the study was conducted);
- The size of an expected or observed effect.

In the benefit assessment of drugs, not only individual studies are assessed, but the results of these studies are incorporated into a systematic review. The certainty of results of a systematic review is in turn based on the certainty of results of the studies included. In addition, it is determined in particular by the following factor:

- The consistency of the results of several studies.

The study design has considerable influence on the certainty of results insofar as a causal association between intervention and effect cannot usually be shown with prospective or retrospective observational studies, whereas controlled intervention studies are in principle suited for this purpose [160]. This particularly applies if other factors influencing results are completely or almost completely eliminated. For this reason, an RCT represents the gold standard in the assessment of drug and non-interventions [241].

In the assessment of drugs, RCTs are usually possible and practically feasible. As a rule, the Institute therefore considers RCTs in the benefit assessment of drugs and only uses non-randomised intervention studies or observational studies in justified exceptional cases. Reasons for exception are, on the one hand, the non-feasibility of an RCT (e.g. if the therapist and/or patient have a strong preference for a specific therapy alternative) or, on the other, the fact that other study types may also provide sufficient certainty of results for the research question posed. For diseases that would be fatal within a short period of time without intervention, several consistent case reports may provide sufficient certainty of results that a particular intervention prevents this otherwise inevitable course [249] (dramatic effect, see also Section 3.2.2). The particular obligation to justify a non-randomised design when testing drugs can also be found within the framework of drug approval legislation in the directives on the testing of medicinal products (Directive 2001/83/EC, Section 5.2.5 [235]).

In the preparation of the report plan (see also Section 2.1.1), the Institute therefore determines beforehand which study types can be regarded as feasible on the basis of the research question posed, and provide sufficient certainty of results (with high internal validity). Studies not complying with these minimum quality standards (see also Section 6.2.3) are not given primary consideration in the assessment process.

Sections 3.2.1, 3.2.2, and 6.2 present information on the assessment of the internal validity of studies, as well as on further factors influencing certainty of results, such as the consistency of the results of several studies and the relevance of the size of the effect to be expected.

In addition to the description of the certainty of results of the studies considered, it is necessary to describe whether – and if so, to what extent – the study results are transferable to local settings (e.g. population, health care sector etc.), or which local study characteristics had or could have had an effect on the results or on their interpretation. From this perspective, studies are especially relevant in which the actual German health care setting is represented as far as possible. However, the criteria for certainty of results outlined above must not be ignored. Finally, the transferability of study results (generalisability or external validity) must be assessed in a separate process that is initially independent of the study design and quality.

3.4 Non-drug therapeutic interventions

Even if the regulatory preconditions for the market access of drugs and non-drug therapeutic interventions differ, there is initially no reason to apply a principally different standard concerning the certainty of results in the assessment of the benefit and harm of an intervention. For example, the Code of Procedure of the Federal Joint Committee envisages, as far as possible, the preferential

consideration of RCTs, independent of the type (drug/non-drug) of medical intervention to be assessed (§20) [145]. For medicinal products, this is weakened by the conformity evaluation in the EN ISO Norm 14155-2 (Section 4.7 [92]), where RCTs are not presented as the design of choice. However, the choice of design must be justified.

Compared with studies on drug interventions, studies on non-drug interventions are often associated with specific challenges and difficulties [261]. For example, the blinding of the staff performing the intervention will often be impossible, and the blinding of patients will either be difficult or also impossible. In addition, it can be assumed that therapists' and patients' preferences for certain treatment options will make the feasibility of studies in these areas particularly problematic. In addition, in the assessment of complex interventions in particular, it may be necessary to consider the possibility of contamination effects. It may also be necessary to consider the distinction between the effects caused by the procedure or (medicinal) product to be assessed on the one hand, and those caused by the skills of those applying the intervention on the other. Moreover, learning effects related to the time of assessment need to be taken into account.

In order to give consideration to the aspects outlined above, studies of particularly good quality are required in order to achieve sufficient certainty of results. Paradoxically, the opposite has rather been the case in the past; i.e. sound randomised studies are often lacking, particularly in the area of non-drug interventions (e.g. in surgery [261]). In order to enable any conclusions at all to be made on the relevance of a specific non-drug therapeutic intervention, it may therefore also be necessary to consider non-randomised studies in the assessment. Nonetheless, quality standards also apply in these studies, in particular regarding measures taken to ensure structural equality. However, such studies will usually and at best be able to indicate a(n) (additional) benefit or harm of an intervention because of their inherently higher uncertainty of results.

The inclusion of studies with lower evidence levels is consistent with the corresponding regulation in the Federal Joint Committee's Code of Procedure (§20) [145]. However, the specific obligation to provide a justification is emphasised. In this regulation it is noted: "However, in order to protect patients, recognition of a method's medical benefit on the basis of documents with lower evidence levels requires all the more justification the greater the deviation from evidence level 1 (in each case, the medical necessity of the method must also be considered). For this purpose, the method's potential benefit for patients is in particular to be weighed against the risks associated with the demonstration of effectiveness based on studies of lower evidential value" [145]. This means that the non-availability of studies of the highest evidence level cannot generally be viewed as sufficient justification for a benefit assessment conducted on the basis of studies with lower evidence levels. In the assessment of non-drug therapeutic interventions, it may also be necessary to consider the

approval status of interventions to be assessed (e.g. [medicinal] products, components of complex interventions, or comparator interventions) (see Section 3.3.1). The corresponding consequences must subsequently be specified in the report plan (see Section 2.1.1).

3.5 Diagnostic tests

In general, the evaluation process for diagnostic tests can be categorised into different hierarchy phases or levels, analogously to the evaluation of drugs [141,232]. Phase 4 prospective, controlled diagnostic studies according to Köbberling et al. [232], or Level 5 studies according to Fryback and Thornbury [141] have an (ideally random) allocation of patients to a strategy with or without application of the diagnostic test to be assessed or to a group with or without disclosure of the (diagnostic) test results. These studies can be seen as corresponding to Phase 3 (drug) approval trials (efficacy trials). Accordingly, they are allocated to the highest evidence level (see, for example, the Federal Joint Committee's Code of Procedure [145]). The FDA also recommends such studies for specific indications in the approval of drugs and biological products developed in connection with diagnostic imaging techniques [136]. Examples show that – depending on the effect expected – they can be conducted with comparatively moderate effort [368].

The Institute follows this logic and primarily conducts benefit assessments of diagnostic tests on the basis of studies designed as described above that investigate patient-relevant outcomes. The main features of the assessment comply with the explanations outlined in Sections 3.1 to 3.4. In this context, patient-relevant outcomes refer to the same benefit categories as in the assessment of therapeutic interventions, namely mortality, morbidity, and health-related QoL. The influence of diagnostic tests on these outcomes can be achieved by the avoidance of high(er) risk interventions or by the (more) specific use of interventions. Generally speaking, the patient-relevant benefit of a diagnostic test unfolds through the support of clinical and/or personal decision-making.

Studies in which the interaction between the diagnostic information and the therapeutic benefit is investigated also have a high evidence level and are to be given preference in the benefit assessment of diagnostic tests [319].

If such studies are not available or are of insufficient quantity or quality, an assessment of the diagnostic chain can be performed [262]. In this context, the accuracy of the diagnostic test is assessed by means of generally applied test quality criteria (e.g. diagnostic sensitivity and specificity or likelihood ratios) determined in studies showing sufficient certainty of results (usually Phase 3 according to Köbberling et al. [232]). It is also reviewed to what extent it is proven that the consequences resulting from the test results are associated with a benefit. In the case of therapeutic consequences (which is mostly assumed), such proof can be inferred from

randomised intervention studies (with patient-relevant outcomes) in which a specific (test) result of the diagnostic test to be assessed was defined as an inclusion criterion.

In assessing the certainty of results of studies on diagnostic accuracy, the Institute primarily follows the QUADAS criteria [380], which are adapted, if necessary, for the specific project. The STARD criteria [50,51] are applied in order to decide on the inclusion or exclusion of studies not published in full text on a case-by-case basis.

Level 3 and 4 studies according to Fryback and Thornbury [141] investigate the effect of the (diagnostic) test to be assessed on considerations regarding (differential) diagnosis and/or subsequent therapeutic (or other management) decisions, i.e. it is investigated whether the result of a diagnostic test actually leads to any changes in decisions. However, such studies or study concepts have the major disadvantage that they are not sharply defined, and are therefore of rather theoretical nature. A principal (quality) characteristic of these studies is that it was clearly planned to question the physicians involved regarding the probability of the existence of the disease (and their further diagnostic and/or therapeutic approach) *before* the conduct of the diagnostic test to be assessed or the disclosure of results. This is done in order to determine the change in attitude caused by the test result. In contrast, retrospective appraisals and theoretical estimates are susceptible to bias [141,171]. The relevance of such ultimately uncontrolled studies within the framework of the benefit assessment of diagnostic tests is largely unclear.

It will not always be necessary to reinvestigate the whole diagnostic chain regarding modifications of diagnostic tests already available and for which a patient-relevant benefit has been demonstrated or can be postulated with sufficient plausibility. In such cases it can, for example, be sufficient to verify only equivalent or improved intratest variability. In a comparison between two or more diagnostic tests regarding specific test characteristics, studies with the highest certainty of results are those with a random allocation of the sequence of the test performance (conducted independently of one another and preferably blinded) in the same patient, or with random allocation of the test to different patients. These studies are therefore given primary consideration in the Institute's reports.

It is also conceivable that a new diagnostic test is incorporated in an already existing diagnostic strategy; for example, if a new test precedes (triage test) or follows (add-on test) an established test in order to reduce the frequency of application of the established test or new test, respectively [49]. However, against the background of the subsequent therapeutic (or other types of) consequences, it should be considered that through such a combination of tests, the patient populations ensuing from the respective combined test results differ from those ensuing from the individual test results. This difference in turn could influence subsequent therapeutic (or other types of) consequences and their

effectiveness. If such an influence cannot be excluded with sufficient certainty, comparative studies on diagnostic strategies including and excluding the new test may be required [136,253].

In the assessment of diagnostic products, it may also be necessary to consider their approval status (see Section 3.3.1). The corresponding consequences must subsequently be specified in the report plan (see Section 2.1.1).

3.6 Population-wide interventions

3.6.1 Screening

Screening programmes are composed of different modules, which can be examined either in part or as a whole. The assessment of a screening test generally follows internationally accepted standards and criteria, for example, those of the UK National Screening Committee (UK NSC [362]), the US Preventive Services Task Force (US PSTF [177]), or the New Zealand National Health Committee (NHC) [274].

According to the criteria outlined above, the Institute primarily assesses the benefit of screening tests by means of prospective comparative intervention studies on the whole screening chain, which include the (ideally random) allocation of participants to a strategy with or without application of the screening test (or to different screening strategies) and which investigate patient-relevant outcomes. In this context, the main features of the assessment comply with the explanations outlined in Sections 3.1 to 3.4.

If such studies are not available or are of insufficient quantity or quality, an assessment of the single components of the screening chain can be performed. In this context, the accuracy of the diagnostic test is assessed by means of generally applied test quality criteria (see Section 3.5), determined in studies showing sufficient certainty of results (usually Phase 3 according to Köbberling et al. [232]), and it is reviewed to what extent it is proven that the consequences resulting from the test outcomes are associated with a benefit. In the case of therapeutic consequences (which are mostly assumed), proof can be inferred from randomised intervention studies in which an early (earlier) intervention was compared with a late(r) one. The benefit of an early (earlier) vs. a late(r) intervention may also be assessed by means of intervention studies in which the interaction between the earliness of the start of the intervention and the intervention's effect can be investigated. This can be performed either directly within a study or indirectly by comparing studies with different starting points for the intervention, but with otherwise comparable study designs. Here too, the main features of the assessment comply with the explanations outlined in Sections 3.1 to 3.4.

3.6.2 *Prevention*

Prevention is directed at avoiding, reducing the probability of, or delaying health impairment [376]. Whereas primary prevention comprises all measures employed before the occurrence of detectable biological impairment in order to avoid the triggering of contributory causes, secondary prevention comprises measures to detect clinically asymptomatic early stages of diseases, as well as their successful early therapy (see also Section 3.6.1). Primary and secondary prevention measures are characterised by the fact that, in contrast to curative measures, whole population groups are often the focus of the intervention. Tertiary prevention in the narrowest sense describes specific interventions to avoid permanent (especially social) functional deficits occurring after the onset of disease [376]. This is not the focus of this section, but is addressed in the sections on the benefit assessment of drug and non-drug interventions (see Sections 3.3. and 3.4).

The Institute also primarily performs benefit assessments of prevention programmes (other than screening programmes) by means of prospective, comparative intervention studies that have an (ideally random) allocation of participants to a strategy with or without application of the prevention measure, and that investigate patient-relevant outcomes. Alternatively, due to potential “contamination” between the intervention and control group, studies in which clusters were allocated to the study arms may also be eligible [357].

In individual cases, it needs to be assessed to what extent the consideration of other study designs is meaningful [209]. For example, mass-media campaigns are often evaluated within the framework of “interrupted time-series analyses” (e.g. in [370]), and the use of this study design is also advocated for community intervention research [43]. In the quality assessment of these studies, the Institute uses for orientation the criteria developed by the Cochrane Effective Practice and Organisation of Care Review Group [72].

For the benefit on the population level, not only the efficiency of the programme is decisive, but also the participant rate. In addition, the question is relevant as to which persons are reached by prevention programmes; research indicates that population groups with an increased risk of disease participate less often in such programmes [242]. Special focus is therefore placed on both of these aspects in the Institute’s assessments.

4 Clinical practice guidelines and disease management programmes

4.1 Background of guideline appraisal

Clinical practice guidelines (CPGs) are seen as key instruments in the improvement and assurance of the medical quality of patient care [118]. Their objective is to reduce inappropriate differences in patient care and improve care by formulating concrete recommendations for clinical decision making. In Germany, they are also used as a basis for decisions on steering procedures in the health care system, such as in connection with the formulation of requirements for disease management programmes (DMPs) (in acc. with §137f SGB V). Consequently, CPGs are increasingly influencing decisions affecting the structural level of the German health care system.

Taking this into consideration, it should be ensured that CPGs are based on the best and most up-to-date scientific evidence available, and are formulated under consideration of clinical experience.

However, in many cases the reference to current scientific evidence is lacking [182,321], and some CPGs on the same topics reveal considerable differences in their recommendations [63,207].

The main reason for this is that the internationally stipulated quality standards for CPG development are not followed consistently [29,70,334].

4.2 Aims of guideline appraisal

The appraisal of CPGs aims to improve care through greater transparency in the health care system. For this purpose, it is particularly important to:

- Discriminate between CPGs of good or bad quality with regard to methodology and content;
- Elaborate and review the evidence base on which key CPG recommendations are founded;
- Make clear scientific statements on the reasonableness and effectiveness of the implementation of different medical recommendations;
- Provide the Federal Joint Committee or its panels with a basis for decisions in discussions on DMPs;
- Ensure that only verified (quality-assured) CPGs, for which an improvement in outcome has been indicated, are introduced into health care;

- Identify research needs and initiate meaningful projects for the development and implementation of evidence-based recommendations;
- Promote the inclusion of CPGs in total quality management (TQM) processes.

Furthermore, the results of this work provide the users of CPGs (physicians, health facilities, health policy committees, decision-makers in the health care system, and patients) with orientation towards meaningful and appropriate recommendations on high-priority health care problems.

For the appraisal of CPG content, the available methodological competence and topic-related expertise of external institutes, facilities, or organisations should be used and incorporated as much as possible [132].

4.3 Methods of guideline appraisal

Essential approaches in the appraisal and review of the quality of CPGs are:

- The examination of formal criteria that primarily reflect the transparency of the developmental process and assume that CPGs applying these criteria will make correct recommendations with a higher probability (comparable to the internal validity of studies) [162,166];
- The precise review of content with regard to the underlying evidence;
- The appraisal of the appropriateness of recommendations;
- The appraisal of the effects caused by the implementation of CPGs (outcome evaluation).

The approaches can vary substantially (depending on how they are performed, their evidential value, and the effort invested), and they are applied according to the research question posed and the commission awarded. The methodology applied by the Institute in this context is reviewed regularly and, if necessary, updated under consideration of current scientific publications, as well as national and international experience. Different aspects are usually combined with each other in a step-by-step procedure [132]. The restriction to largely formal and methodological aspects of the appraisal, which can be well operationalised, was critically discussed before the revision of the methods. On the one hand, such restrictions are imposed by the current legal framework (§139a SGB V), which so far envisages a review of CPGs (only) with regard to the underlying evidence. On the other, the methods for an evidence-based (and consented) CPG development and update are still insufficiently applied [20]. For this reason, a formal appraisal cannot be dispensed with at present.

With regard to their relevance, appropriateness, and practicability, individual recommendations and key points in CPGs may be judged very differently by those affected.

So far, there is no internationally harmonised procedure for the appraisal of the appropriateness of content of CPGs. Such a procedure needs to consider different aspects and must also involve other stakeholders, such as the Federal Joint Committee, professional representatives, and patients.

Statements on the appropriateness of content of CPG recommendations comprise two key questions:

1. Are the presentation and interpretation of current evidence presented in the CPG, from which concrete individual recommendations are made, appropriate and comprehensible?
2. Is the intervention recommended appropriate?

To answer these questions, the following aspects need to be considered:

- A benefit of the intervention must have been demonstrated, and the weighing of benefit and harm must have resulted in a positive decision (see Section 3.1);
- The intervention must be relevant for the German health care system;
- The intervention must be available and approved;
- The intervention must be necessary;
- Safe application (by physician/patients) must be possible;
- The intervention can be implemented and financed with the resources available.

In a shared risk community, the last point in particular requires a generally agreed decision on the available overall financial frame, as well as on the prioritisation of health care services. So far, the Federal Joint Committee, at least indirectly and partly, has taken on this function with regard to the inclusion and exclusion of services from the SHI benefits catalogue, and forms a consensus between the different interests and stakeholders (including patients). It is conceivable that in future, the Federal Joint Committee will also take on a similar function within the framework of the appraisal of the content of CPGs (regarding their appropriateness), and will publish corresponding national CPGs after the review of the evidence base by IQWiG. Against the background of the appraisal of the appropriateness of CPGs, IQWiG has entered a developmental process in which aims and methods of this procedure are developed.

4.3.1 Formal review

The issue of CPG quality can be approached through a formal review using methodological criteria [15,63,70,133]. The formal CPG appraisal is conducted in a structured manner following the methods of CPG clearing procedures, and referring to CPG appraisal criteria of the German Medical Association and the Association of Statutory Health Insurance Physicians [15,17] by means of DELBI,^s the German Instrument for Methodological Guideline Appraisal [16]. DELBI is the German adaptation of the AGREE^t instrument [3], extended by a domain that refers to the applicability of the particular CPG to the German health care system. The formal appraisal is performed by two researchers independently of one another. Where conflicting appraisals are made, the issues are discussed and the affected CPGs are appraised again. If the dissent continues and cannot be solved by a query to the CPG authors, it will be documented separately.

On an international level, several CPGs are available on specific medical topics that in part vary greatly in methodology and content [207]. In this context, a formal review has an important filter function that ultimately enables an appraisal of the content of key statements and specific recommendations of relevant CPGs.

First, a comprehensive search for CPGs is conducted in relevant databases (CPG and additional literature databases) in order to identify up-to-date CPGs available on the particular research question (under consideration of the procedures outlined in Section 6.1). The filtering procedure for CPG appraisal has a multi-step approach. Inclusion and exclusion criteria and the search strategy (search terms, choice of data bases, etc.) are prespecified and documented depending on the research question posed. A first screening step follows in which the hits are selected according to predefined thematic criteria. A short methodological appraisal is conducted in a second step in which CPGs are selected that comply with minimum international standards. An appraisal of the content of the resulting documents follows.

4.3.2 Comparison of guidelines and appraisal of content of key recommendations

The appraisal of CPG content is of particular relevance. The criteria for identifying and interpreting evidence and formulating CPG recommendations, which have so far been checked with conventional instruments (“ÄZQ^u-Checkliste” [15], AGREE [3], DELBI [16]), are essentially

^s Deutsches Leitlinien-Bewertungs-Instrument

^t Appraisal of Guidelines Research and Evaluation

^u Ärztliches Zentrum für Qualität in der Medizin (Agency for Quality in Medicine)

transparency criteria.^v On the basis of these criteria, only the reporting of a process (e.g. complete presentation of the steps of a literature search) is appraised, without, for example, appraisal of the completeness and topicality of the search.

For the essential key recommendations of a CPG, their derivation from the underlying evidence must therefore be reviewed in each case. Besides the appraisal of completeness and topicality of the literature consulted, the review of content also includes the appraisal and interpretation of study results. As this procedure involves a great deal of time and effort, the appraisal of content must be limited to the research questions commissioned by the Federal Joint Committee, or to the CPG's key recommendations. The key recommendations are identified within the context of each specific commission in consultation with the Department Heads and external experts concerned. In this context, focussing on selected key issues has several advantages. As CPGs are very heterogeneous with regard to their content and recommendations, even if they refer to the same medical topic, one can reach a good comparison between different CPGs by reducing the scope of the appraisal to predefined key issues. Within this framework, a review of the underlying evidence and the resulting recommendations is also possible, provided this is presented in a transparent manner in the underlying CPG. A comparison of content in respect of the key question defined is not only possible between CPGs by different publishers, but also between CPGs and other evidence-based sources (systematic reviews, HTA reports, IQWiG's evidence reports, etc.). In consultation with IQWiG's other departments, our department also works on individual research questions that are not investigated in available CPGs.

A comparative summary of the content of CPGs can be helpful in identifying key recommendations. In particular, questions that are the subject of scientific dissent can be identified. Methodologically, the comparative summary only facilitates the appraisal process; a review of the evidence base is also meaningful for interventions that are consistently recommended. In the appraisal of the content of key recommendations it can also be reviewed whether the outcome parameters used in the CPG are relevant to patients, and whether a weighing of the benefit and harm of an intervention [166] was considered in the formulation of the recommendation.

4.3.3 *Improvement in outcomes*

The key question in the preparation and implementation of CPGs is whether their implementation leads to a measurable improvement in health care [118,163,387]. Strictly speaking, this can only be

^v Efforts are currently being made to initiate methodological improvements to these aspects (e.g. by the GRADE working group [154]).

reviewed by a rigorous evaluation of effects [182]. However, this cannot be realised for every existing CPG. For example, only unsystematic studies with heterogeneous research questions and results investigate the topic “Outcome evaluation after guideline implementation”. Moreover, pilot studies have so far only been conducted for a few guidelines before publication [18,19].

If results of pilot studies or projects testing CPGs are available, these are to be included in the overall appraisal (e.g. by describing methods, quality indicators, results and consequences).

It would also be meaningful to compare CPG recommendations with the conventional procedures applied in routine health care. If complex changes are recommended in a CPG, its implementation is more difficult and has to be accompanied by supportive measures and tools [63].

In particular, CPGs from other countries must be reviewed in respect of the transferability of their conclusions to the German health care system and/or which structural prerequisites need to be implemented for the successful introduction of these CPGs. The Institute can also be commissioned by the Federal Joint Committee to appraise CPGs.

4.4 Presentation of quality assessment

Structured reports (guideline appraisal reports) are prepared from the results of the appraisals to provide the Federal Joint Committee with a basis for further consultations. The reports can also serve as a basis for producing topic-related information or be used by scientific societies in the revision of CPGs.

4.5 Issue of recommendations on disease management programmes

In accordance with §91 SGB V, the Federal Joint Committee names diagnoses to be included in DMPs (in acc. with §137f SGB V), and develops requirements with regard to the content of these programmes. In accordance with §139a (3) SGB V, it is IQWiG’s task to issue recommendations on DMPs. This includes:

- Supporting the Federal Joint Committee in naming new diagnoses for DMPs;
- Revising existing requirements for DMPs;
- Developing new requirements for the content of DMPs.

The concrete possibilities of supporting the Federal Joint Committee and its panels beyond the commissioning of specific research questions within the framework of IQWiG’s benefit assessments are currently being reviewed in consultation with the responsible panels.

5 Evidence-based health information for consumers and patients

5.1 Background and goals

The Institute has a legislative responsibility for providing health information to consumers and patients, but not direct advice to individuals. The Institute's goal is to improve health and patient autonomy through the provision of health information that aims to advance health and scientific literacy [33,95,228,239]. The goals of the Institute's health information are therefore to:

- Support active and informed decision-making about health issues;
- Promote the critical use of health care services;
- Improve understanding of physical, mental and emotional health;
- Improve understanding of medical and scientific information, including the concept of evidence-based medicine; and
- Enable support of patients by family and friends.

To achieve these goals, the Institute needs to be a reliable, trusted and patient-centred information provider. The integration of patient values in decision-making is core to the concept of evidence-based medicine [314], and thus to evidence-based health information. There are several definitions of evidence-based patient information [77,115,317,345]. Each definition requires the information to include evidence on benefits, harms and uncertainties of interventions, but the requirements of evidence-based health information go beyond that [345]. The Institute defines evidence-based health information as information where:

- The content is based on clear scientific evidence, particularly systematic reviews;
- The information is developed following systematic methods which aim to minimise bias and maintain neutrality;
- Evidence-based communication techniques are used to meet the goals of informing, supporting and empowering consumers and patients;
- Uncertainties as well as the potential for benefit and harm are discussed;
- Language and framing are neutral and non-directive, so that decisions are made in accordance with the patients' own values; and
- The information is updated so that it remains evidence-based.

The Institute's primary medium for communication is the internet, through which it aims to reach consumers, patients and those who advise or provide information to patients ("information

multipliers”). Internet-based and offline computer-based health information can positively affect consumers’ and patients’ knowledge, choices, health and wellbeing [42,221,246,273,276,341,377,385]. However, information interventions can also be ineffective or harmful [106,108,159,276]. Evidence-based information is also unfamiliar to many users, which brings a set of communication challenges [116,150,346]. The Institute therefore needs to maintain its expertise in evidence-based and electronic communication and monitor the use of its health information.

5.2 Patient-centred communication

5.2.1 Communication standards

A key challenge for evidence-based health information is to communicate in a way that is widely understandable while remaining scientifically accurate and objective. The objective and easy-to-use methods of measuring the level of readability in English, French and Spanish have only limited applicability in Germany, and no similarly validated local tool exists [146]. There is also a wide variation in literacy levels among consumers and patients [228].

To support the differing levels of understanding in the community, the Institute produces information at a variety of comprehension levels – from fairly simple fact sheets to more complex articles - and including multi-media elements (see below). As the Institute produces its information in English as well as German, it can also make use of English readability tools in assessing the level of readability of its information. The primary means of quality assessment for understandability, however, will remain reviews of drafts by test readers and reader ratings of understandability (see below).

Explaining evidence and remaining objective in communicating about health-related information raise additional challenges [116,327,358]. To be objective and non-directive, the Institute’s health information should, on the one hand, not exaggerate what is scientifically known and, on the other, not tell people what they “should” do. This is achieved by not making recommendations and by using neutral language. Drawing on the evidence that has accumulated on communicating research findings, the Institute aims to:

- Present information in consistent formats to aid understanding, supplemented with additional elements that could enhance the understandability of medical words and numerical information;
- Explain the degree of uncertainty associated with evidence;
- Indicate to which groups of people the evidence is applicable;

- Distinguish carefully between “absence of evidence” and “evidence of no effect”;
- Supplement relative risk information with absolute risk data and other information (such as baseline risk) whenever possible; and
- Avoid biasing information for or against the products of any particular company, by using generic names of products whenever possible and only using brand names when it is essential for understanding and / or all products on the market in Germany can be named.

There is evidence that the presentation of personalised or individualised risk estimates can be effective [108]. The Institute uses tools which can help people estimate their personal risk. Although uncertainties remain about formal decision aids, especially for individual use on the internet [68,113], the Institute may develop decision aids for some topics, particularly drawing on the experience of specific decision aids which have been shown to be effective in randomised trials. The Institute develops its decision aids in accordance with the International Patient Decision Aid Standards (IPDAS) [113,189].

The provision of information is not the only purpose of health communication. An important role of health information is to provide patients and consumers with emotional support [129], and it can also play a role in enhancing patient autonomy. Health communication needs to be patient-centred if it is to be empowering and emotionally supportive. According to the definition of the World Health Organisation, empowerment in health includes the ability to make choices and take actions in accordance with your own goals [283]. These abilities enable patients to think and act autonomously. Empowering health communication addresses what patients want to know, shows interest and respect for what patients think and respects their competence [97,227,372].

Historically, patient information has tended to be paternalistic [97], assuming ignorance on the part of the patient, together with a need for them to be protected from uncertainty and distressing information: they should be directed about what they should do. This remains, to a significant extent, a feature of patient information and discussions about its role today [97]. In Germany it is sometimes argued that patients are not ready for evidence-based and non-directive information, or that they do not want this kind of information.

However, studies consistently show that patients in Germany receive less information than they would like [95], and that more patients in Germany report that they do not have enough information in comparison with people in other similar countries [322,366]. German users of the internet are possibly even more interested in hearing about the latest research results than they are in hearing the opinions of experts [297]. Nevertheless, the effects of making people more aware of scientific uncertainty are to a large extent unknown. It has been argued that coping with uncertainty is an integral part of coping and self-realisation for adults [137]. The possibility of harm cannot be

ruled out though. The Institute will therefore keep up-to-date with the evidence on communicating about risks, uncertainty and difficult information and monitor the impact of its information locally.

As well as seeking to be understandable, objective and accurate in its information, the Institute aims to:

- Demonstrate sensitivity and respect for consumer and patient knowledge, values and concerns, autonomy, cultural differences as well as gender, age and disability-related interests;
- Maintain a patient-centred, non-judgmental, non-directive and neutral style of language; and
- Respect readers' time.

On the basis of available evidence and the experience of other groups, the Institute has developed a style guide for its products. It will continue to develop its communication standards in the light of the monitoring and evaluation of its products, as well as emerging evidence in the area of evidence-based communication.

5.2.2 Visual communication and multi-media

Text alone may not be as understandable and memorable as information where explanations are supported by pictures [116,193,236,250,358]. Spoken text may also enhance understanding [193,318]. Explanations where text, pictures and sound are combined may be the most understandable form of communication, especially for people of lower literacy [193]. Where possible, the Institute supports its texts with visuals and sound to enhance the effectiveness of its information and widen the audience that could understand it. This includes anatomical diagrams and short animated films on key topics that combine visuals, text and sound. Graphics and pictorial symbols may also be helpful to many people in explaining numerical data and other aspects of evidence [107,250,327]. Visual and multi-media elements should not replace text, but enhance the material covered in the texts. This ensures that the information is still accessible to people who are visually or hearing impaired.

The internet enables health information to be presented in multi-media formats. As the technology of the internet improves and access to the internet is no longer limited only to computers, communicating effectively with vision and sound on websites is becoming increasingly feasible for more users. The internet also enables interactivity with users, so that communication need not flow only towards the user. Showing an interest in what is important to patients is a critical element in patient-centred and empowering communication [97,227,372]. While the Institute cannot provide

individual health advice, there are nevertheless multiple ways in which the Institute offers its users the opportunity to share their views and concerns, including:

- Online quality rating of articles;
- Topic suggestion and general online contact form;
- An ongoing survey of the website's usability; and
- Occasional online polls on specific health topics.

5.2.3 Accessibility

The internet has both particular advantages and disadvantages in terms of accessibility. For example, its availability 24 hours a day to those with access to the internet makes it a highly accessible medium. Access to the internet continues to increase: more than half the people in Germany use the internet for health information, and the number continues to grow [95,325].

The rate of use of the internet by people with chronic diseases may be particularly high. Studies in patients attending orthopaedic clinics in Germany, for example, found that up to 70% were using the internet for information about their condition [290,301]. Over one-third (38%) of patients had visited the internet on the subject of their consultation before they arrived at the clinic [290]. However, it is not necessary for people to have direct personal access to the internet for them to benefit from health information on the internet: relatives or friends will often search the internet on their behalf, and information "multipliers" such as doctors, self-help groups and journalists routinely use the internet. Health information might be shared widely among family members [329]. In the early years of the World Wide Web there were clear gender differences in terms of access to the internet, but these appear to be getting smaller [276].

For internet use, there are several accessibility issues, including:

- Disabilities, particularly (but not only) visual and hearing impairment;
- Poor reading skills;
- Insufficient computer skills;
- Technological capacity (affecting speed and access to multi-media); and
- Language (the user's mother tongue).

The Institute ensures that its website meets internationally accepted disability accessibility criteria [374], and the German BITV ("Barrierefreie Informationstechnik-Verordnung") [61]. It will continue to evaluate and optimise the usability of the website, and develop tools that assist with understanding the Institute's information.

Close to 10% of people living in Germany have another nationality and close to a further 10% of Germans have a migrant family background [344]. The largest cultural group among these is people from Turkey [344]. People from non-German backgrounds as a group may have greater needs for health information [84]. For this reason it is important that the fact sheets are written in easily-understandable language. Ideally, culturally and gender-appropriate health information would be available in people's original languages. The Institute will cooperate with external partners to enable at least some of its health information to be translated into the most widely spoken languages in Germany.

The Institute publishes health information in both German and English, and has incorporated the technical capacity for publishing in further languages. The best possible quality assurance requires broad international involvement. Publishing in English enables the Institute's information to profit from feedback from international scientists and reviewers, particularly the authors of systematic reviews. The availability of an English version broadens the opportunities for translation into further languages.

High standards in translating health information are required, but the quality of translations of health information is often inadequate [144]. It is difficult to assess the quality of translations according to objective criteria. Translations can be literal ("word for word") or can aim to capture the intent of the original text in the target language ("sense for sense") [271]. Non-literal translation offers the highest level of understandability in the translated language, and is therefore the style of translation used by the Institute. All translations of the Institute's health information are double-checked by a second person with proficiency in both languages.

5.3 Topic selection, research and evaluation of evidence

5.3.1 Topic selection

The Institute's health information is produced:

- in response to commissions received from the G-BA (Federal Joint Committee) or the German Ministry of Health;
- to summarise other products published by the Institute and as accompanying information for these products; and
- to fulfil its legislative responsibility to provide consumers with health information, as well as on its own initiative within the framework of the G-BA's general commission.

This section addresses methods for topic selection for information accompanying other products and information produced by the Institute on its own initiative in accordance with the general

commission. Health information is potentially limitless in scope, and informing everybody about everything is not possible. As with other health care priority-setting decisions, deciding priorities for health information involves simultaneous analysis of multiple streams of dissimilar information [27,28].

§139a of the German Social Code Book V (SGB V) sets the following task for the Institute: “Provision of easily-understandable general information to citizens on the quality and efficiency of health care services, as well as on the diagnosis and therapy of diseases of substantial epidemiological relevance”. The Institute’s general commission was amended in July 2006 to specifically include informing the general public – after identifying and evaluating relevant literature. The general commission states that all the Institute’s departments, including the Department of Health Information, are required to “continuously monitor and evaluate fundamentally important developments in medicine” and report on these.

It was not possible to come up with a broadly acceptable definition or a clear list of “diseases of substantial epidemiological relevance”. The epidemiological relevance of a disease can only be determined based on various types of burden of disease data. Epidemiologically relevant factors could include:

- Mortality;
- Frequency of occurrence (prevalence / incidence);
- Frequency of utilisation of health care services;
- Treatment costs;
- Absence from work due to illness; and
- Impairment of quality of life and other consequences that have a significant impact on the lives of those affected by the condition.

The Institute uses a variety of sources when setting priorities for topics, including causes of death, diagnoses for hospital admission, absence from work due to illness, common diagnoses and prescriptions in Germany, as well as the list of diseases chosen within the framework of the morbidity-oriented risk structure compensation scheme in Germany [64].

The legislative responsibility of the Institute to provide citizens with information concerns “healthy” consumers as well as patients. For this reason the spectrum of information should also include topics that cover the perspective of people who are not directly affected by an illness. To meet its goals, the Institute needs to offer information that helps people make choices and take actions to realise their own health goals [283], support self-management, and address what people want to know [97,227,372]. There is some evidence that involving patients in the development of

patient information can increase its relevance [282]: such “consumer-informed” patient information might also be more likely to increase patients’ knowledge [282]. The Institute’s choice of topics therefore also needs to be based on what might interest patients and consumers and reflect the realities of the health care system.

The Institute uses a number of sources to find out what consumers and patients would like to know:

- Surveys, primary qualitative research and reviews of qualitative research on people’s information needs;
- The experiences of other information providers, patient advice services and self-help groups;
- Enquiries made to the Federal Government Commissioner for Patients’ Affairs [326];
- Topics that are entered into the search engine of the IQWiG website for consumers and patients (www.gesundheitsinformation.de / www.InformedHealthOnline.org) as well as other data concerning internet use;
- Topics suggested by the website users; and
- Results of the Institute’s own online polls about priorities and interests.

The Institute also considers reviews of effective information interventions in specific health and illness-related topics, to help it determine which health information might be valuable. The Institute’s choice of topics is not solely based on predefined (interest-driven) issues. It gives priority to questions for which evidence-based answers are possible. Therefore, the Institute searches for evidence (evidence scanning) to find potential topics. The scanning system is based predominantly on identifying systematic reviews and health technology assessments, including reviews of adverse effects. A scanning system was piloted by the Institute in 2006 to assess its feasibility. It is being further developed in cooperation with the Centre for Reviews and Dissemination to take advantage of the Centre’s extensive experience in identifying and appraising systematic reviews [280].

The Institute is developing a list of criteria to help guide its topic choices. The evidence-scanning system will also play a critical role in the process for updating the Institute’s health information (see below).

5.3.2 Research

The Institute relies predominantly on systematic reviews and qualitative research to develop its information. When researching a topic in depth, the Institute generally looks for the following information to help identify issues of interest and concern for patients and consumers:

- Rapid appraisals of primary qualitative studies as well as reviews of qualitative studies (see Section 6.5);
- Reviews of the effects of communication;
- Reviews of adherence studies; and
- Freely accessible patient information on the internet as well as self-help group websites.

The internet and other sources are also searched to identify the interventions being used by, or offered to, consumers and patients.

The results of this primary assessment of patients' information needs inform the Institute's picture of the patient journey in relation to a health issue, the psychological and emotional issues in relation to that topic, and the decisions that individuals may need to make. Representatives of patients or consumers are also generally interviewed to identify further issues and discuss the relevance in Germany of the Institute's findings from research.

The Institute also searches for systematic reviews of causes, prognosis, diagnosis, treatments and adverse effects. This usually covers the entire disease, with a scoping exercise conducted later to focus on areas that the health information will cover. Searches include, but are not limited to, the Database of Reviews of Effects (DARE), [54,280] and the 'systematic review' filter of PubMed [268]. Only reviews with search strategies undertaken in the last five years are considered. Reviews are generally considered to be up-to-date if the search was carried out within the last three years [335,336].

The Institute sometimes considers doing a rapid search for trials. For example, if there is no more recent review on an important subject, an update search is considered if there is a high quality review with a search strategy conducted more than three years ago [335]. In some instances, updates are considered relevant for a time span that is more or less than three years. This depends on the strength of the evidence in the review and the extent of research activity in that field.

An update search for trials, to test how up-to-date a review is, is generally conducted using the Cochrane Controlled Trials Register and the 'clinical trial' filter of PubMed [179].

Rapid searches of qualitative literature are conducted by one person in the Department, with a second person reviewing the final results. Searches for other evidence include screening of results by at least two people independently. Exclusions are agreed between the two people, with a third person when necessary.

5.3.3 *Evaluation of evidence*

The health information produced for patients and consumers is mainly based on systematic reviews (see Section 6.3). The Institute only uses systematic reviews concerning treatment effects if they fulfil certain minimum requirements. For example, the review is only allowed to have few methodological flaws according to the Oxman and Guyatt Index [210,286,288]. To be the subject of a research summary suggesting treatment benefit, a review should include at least one trial judged to be of adequate quality by the review's authors, and include data on at least one patient-relevant outcome. The Institute also takes into consideration the relevance and applicability of the evidence to the reader, particularly in terms of gender and age (see Section 6.5).

Independence is a critical attribute of the Institute's legislative mandate, the way the Institute is portrayed by others [95] and the way the Institute portrays itself (www.gesundheitsinformation.de / www.InformedHealthOnline.org: "Independent, objective and evidence-based"). Independence might be a contributor to ensuring trust in the Institute's patient information [226,297,337], although this is not necessarily so for all people [34]. Independence in the conduct of research might also be of importance to many members of the general public [230]. Systematic reviews should provide an impartial evaluation of the state of knowledge. Some studies have suggested that systematic reviews which have been sponsored by a product's manufacturer may be biased towards an over-positive assessment of effectiveness [218,388]. A further study also found a similar association, although it was not statistically significant [293]. To maintain the reputation of the Institute as a provider of independent information, reviews sponsored by manufacturers are not used for the Institute's patient information.

When more than one systematic review of adequate quality addresses a particular subject or outcome, a further quality assessment is carried out to determine if any of the reviews are of inferior quality. The results of the highest quality review for a particular outcome are the source of numerical data used in the Institute's health information. Where reviews come to contradictory conclusions, the possible reasons for this are explored [211].

The methods of the GRADE working group may be used to formally assess the strength of primary evidence in a certain systematic review [23,24,154,172]. The GRADE system explicitly assesses the quality of the evidence and describes how trustworthy the estimates of specific treatment effects are, such as the estimated mortality rate associated with the treatment in question.

For issues concerning the aetiology or prognosis of a condition, or qualitative descriptions of patients' experiences, other types of primary studies are suitable for inclusion in systematic reviews [149]. When assessing such systematic reviews, the Institute uses the criteria of the Oxford Centre

for Evidence-Based Medicine and the McMaster University evidence-rating system [67,179]. The Institute's methods for assessing qualitative research are described in Section 6.5.

5.3.4 Updating

A critical part of evidence-based health information is making sure that its conclusions are not out-of-date. Regular updating is one of the quality criteria determined by the European Union for health-related websites [75]. Evidence is growing exponentially. This is the case for both trials [359] and for systematic reviews [266]. New evidence can render existing reviews obsolete or out-of-date [336], although new evidence often leads to no change or a strengthening of the original conclusions [203,298].

A study of guideline recommendations concluded that after three years, over 90% of recommendations may still be current, while after less than six years, about 50% of the recommendations in guidelines may be obsolete [335]. For some topics, for example where the evidence is very strong, the half-life of evidence will be much longer, and in other areas it will be less than three years [336]. However, as evidence continues its exponential growth, the half-life of information is likely to shorten: that is, information will become out-of-date more quickly. The Institute sees three years as the usual minimum time after which its information requires review. The half-life of the Institute's health information is monitored to inform future updating methods.

Updating can be very resource-intensive: it has been estimated that a full update of a guideline, for example, can take almost as long as developing a new guideline [105]. Traditional mechanisms of updating are to schedule a review for a set date. However, this is only sustainable for providers of multiple pieces of evidence-based information in the long term if there is also a continuous increase in resources: an updating workload will continue to grow exponentially over time. The Cochrane Collaboration, for example, has the goal of having the searches for at least 80% of its reviews updated every two years [233]. However, this has not been possible and the reviews are instead probably becoming ever more out-of-date [233]. Using this standard approach of a two-yearly update, the updating workload for the Institute's health information would exceed the Institute's capacity to both stay up-to-date and keep producing new information within about five years.

The Institute uses the following model to keep its information relatively updated:

- The new reviews identified through the evidence-scanning system with the Centre for Reviews and Dissemination (see Section 5.3.1) are used to trigger updating (“refreshing” of information);

- The updates of a key group of other providers of evidence-based health information and safety alerts are electronically networked by the Institute and used to trigger “refreshing” of information; and
- The remainder of information which has not automatically been “refreshed” is reviewed at the end of a predetermined time period (usually three years).

In addition, the Institute monitors developments in methodology for determining when evidence has become so strong that updates are no longer necessary.

Information on the Institute’s health information website is therefore usually reviewed or “refreshed” at least once every three years. Updates for individual items are not usually done more often than once every six months.

The dates of all updates and scheduled reviews are made known online. Information about updates is sent to all www.gesundheitsinformation.de / www.InformedHealthOnline.org newsletter subscribers, and logged online. The Institute will continue to work with cooperation partners on updating methods, including the Centre for Reviews and Dissemination (NHS England), the Centre for Evidence-Based Medicine (Oxford University), the Haute Autorité de Santé (Paris) and the Cochrane Collaboration.

5.4 Information products

5.4.1 Core products

The Institute produces health information for patients and consumers in different formats. This is intended to help meet the needs of audiences who have differing information needs, differing reading levels, and varying time for reading.

The core products of the Institute include:

- Feature articles: comprehensive article which forms the basis of a set of related articles on an important health issue;
- Fact sheets: short, easily-understandable information; and
- Research summaries: brief summaries of systematic reviews, health technology assessments or large studies, including those produced by the Institute itself.

Together, these products and supplementary items constitute an evidence-based health encyclopaedia. Studies show that the greatest interest most patients and consumers have is in information about treatments and what they can actively do themselves [129,226,267,311,337]. In addition, German patients may feel particularly under-informed about diagnostic tests and their

results [366]. This reflects the need for a major focus of evidence-based information on the effects of treatments, diagnostic tests and self-management strategies. The majority of the Institute's information items are fact sheets and research summaries.

The feature articles and fact sheets are similar in format to conventional patient information, while the research summaries are more similar to newspaper reports.

Feature articles are articles that are usually more than 20 pages in length, directed to people who are interested in more detailed information. The level of readability of these extended articles reflects the more interested and motivated readers likely to read them. These readers may feel patronised by overly simple language [97].

The contents of the individual feature may vary from topic to topic. For each topic, information taken from the following areas is taken into account:

- Descriptions of the health condition or symptoms, including
 - Anatomy
 - Physiology
 - Different forms of the condition
 - Causes of the condition
 - Recognising symptoms
 - Natural course of the condition
 - Prognosis
 - Possible complications
 - Recovery / rehabilitation
 - Possible recurrence of the condition (relapse)
 - Recognising recurrences
 - Risk groups (including caregivers)
- Preventive and health promotion measures, including
 - Diet
 - Physical activity
 - Screening methods
 - Information
- Diagnostic options, including complementary diagnostic procedures
- Treatment options, including
 - Medication
 - Surgical interventions

- Other non-pharmacological options
- Rehabilitation
- Other health services
- Psychosocial aspects, including the personal experiences of patients who have the condition in question, as well as of other people affected by the condition, such as carers, family members and friends.

Fact sheets are written at a reading level that should be more widely understandable, again reflecting the expected use of this information. They should be suitable for use within patient-doctor consultations and for people who want a quick overview of information.

Research summaries can be thought of as research-based “FAQs” (frequently asked questions). One of the greatest information interests for German internet users is the latest research, which they may be more interested in than the opinion of experts [297]. The research summaries offer the opportunity to make the results of high quality scientific evidence more widely accessible in Germany.

The production process for health information is described in Section 2.1.4. Information based on health care research should involve a similar level of quality assurance as the report of the research itself [306]. The quality and relevance of patient information might also be improved if reviewed by patients [282].

Feature articles and fact sheets are therefore externally reviewed by patients too. The quality assurance for all the core products includes the opportunity for the authors of the systematic reviews to comment on the patient information. The draft versions of the core products are sent to the Institute’s Board of Trustees, which also provides the patient representatives there with the opportunity to comment. In addition, the products are tested by potential readers (user testing). The patients who share their stories of illness and health are also invited to comment on the drafts of the patient information that their stories accompany (see Section 5.4.3).

5.4.2 *Supplementary items*

In addition to the core products, additional products (“supplementary items”) are produced. These aim to make the key messages of the health information more understandable and interesting. The inclusion of pictures, sound and animated films may increase the understandability of the website, especially for people with lower literacy levels (see Section 5.2.2). The animated films aim to be the most easy to understand of all of the Institute’s health information products.

People may also prefer and trust websites which are more attractive and which include multi-media elements [77,226,337]. Indeed, high quality content can be rejected solely because of poor design [337].

The supplementary items include:

- Graphics, photos and other images;
- Short animated films with text and sound;
- Interactive quizzes;
- An online dictionary, which can be switched on or off;
- Short explanatory texts on subjects such as recognising the signs of a disease;
- Articles and interactive tools explaining evidence-based medicine, to improve understanding of research and numbers;
- Calculators;
- Online polls and questionnaires;
- Evidence tables; and
- Decision aids.

The goals of the supplementary items are to:

- Promote general understanding of health and medical issues;
- Help patients and consumers to understand and weigh up the potential benefits and harms of medical interventions; and
- Support self-management strategies.

The content of supplementary items usually arises from core products. If the item includes content that is original, or content such as an evidence table or decision aid, then it usually goes through the same external testing and external review process as core products.

Interactive supplementary items are tested for usability in-house. They are also a critical focus of any user testing and evaluation of the website. Accessibility is a further particular focus.

The Institute may sometimes develop decision aids, particularly in areas where decision aids have been shown to be effective (see Section 5.2.1). In developing a decision aid, the Institute applies the International Patient Decision Aids Standards (IPDAS) [113,189].

5.4.3 Patient stories

Many patients would like to hear or read about the experiences of people affected by the same health condition as them [183,351]. Patient stories are commonly used to impart information in

both the fields of journalism and patient information. They represent one means of conveying scientific evidence and making it accessible to the general public [150]. The importance of patient stories in medical practice and in health care is increasingly being recognised [156,347,389]. Patients may trust health information websites more if they include the experiences of other patients [337].

By presenting the individual stories of patients as well as those close to them, the Institute would like to enable patients and other interested people to find out about the different aspects of living with a condition and health care. This is intended as a complementary source of health information, in addition to the other products. The content of the patient stories should not contradict the evidence-based health information.

One example of patient stories associated with evidence-based health information is the Database of Personal Experiences of Health and Illness (DIPEX), an evaluated multimedia website (www.dipex.org) which is available free of charge on the internet [183,184,389]. The Institute's methods for gathering, editing and publishing patient stories are based on DIPEX's established approach.

The Institute prepares patient stories using the following process:

1. Interview partners are found, most often via self-help organisations.
2. Informed consent is sought from interview partners regarding the interview procedure and how the story will be used.
3. The interviews are carried out.
4. The interviews are documented and edited, and the interview partners give their informed consent regarding the publication of the final version.
5. The patient story is published on the website with the permission of the interview partner.

Particular importance is placed on extensively briefing the interview partners before the interview, on the fact that they can withdraw their informed consent to publish the story at any time, on preparing the interviews well, on carrying out the interviews based on predefined criteria, as well as on the anonymity of the interviews. If possible, every feature article should be accompanied by at least 2 patient stories.

5.4.4 Website

The primary dissemination vehicle for the Institute's health information is the bi-lingual website, www.gesundheitsinformation.de / www.informedhealthonline.org. The Institute aims to maintain high website standards in:

- Usability and accessibility (see Section 5.2.3) [196,240,281];
- Privacy and data protection [199];
- Transparency;
- Search engine visibility [355];
- Attractiveness to users; and
- User interactivity.

The Institute aims to achieve usability and user interaction through a variety of means, including:

- Navigation through a graphic of the human body;
- Linkage of related topics to each other;
- Online rating of the quality of individual information items; and
- Help and website tour functions.

The website also includes a free electronic newsletter, with the choice of weekly or monthly subscription. The newsletter contains information on what is new on the website, including when information is updated. In addition, the Institute maintains a version of the website for handheld computers (personal digital assistants) and provides RSS feeds to enable individuals to subscribe by RSS. This also allows the contents of the website to be automatically integrated into other websites.

People's statements about what they trust in a website often focus on factors such as credibility and being clearly non-commercial. User behaviour suggests that, in practice, good design and attractiveness also play a large role in user trust of websites [34,78,226].

The website is based on the open source content management system Papaya [96]. In order to ensure the highest technical and privacy standards, the hardware platform and technical services are provided by DIMDI, the German Institute of Medical Documentation and Information. DIMDI (www.dimdi.de) is an institute set up by the German Federal Ministry of Health (BMG). Its responsibilities include making information from all areas of medicine accessible to interested members of the public.

The Institute's website is certified by the HON ("Health On the Net") Foundation and fulfils the 8 requirements of the HON Code of Conduct (HONcode) for medical and health-related websites [181]. Based in Switzerland, this is an accreditation program for health website standards in content, transparency and privacy. IQWiG has chosen HON because it is internationally recognised, covers multiple quality dimensions, and because HON regularly reviews its accredited websites to ensure that they continue to meet the HONcode.

5.5 Monitoring and evaluation

5.5.1 Routine monitoring

The Institute routinely monitors and analyses the use of its health information website, in particular:

- Website usage, including comparison with similar websites;
- User ratings and feedback, including responses to its ongoing online user survey;
- The information's position in Google searches and the website's Google PageRank [57];
- The technical performance of the website;
- Newsletter subscriptions and retention of subscribers; and
- Adoption of the Institute's information by those who advise or provide information to patients ("information multipliers").

Commonly used metrics for website use, such as number of hits, provide an inflated impression of the use of websites: numbers in the hundreds of thousands can in fact represent very small numbers of people actually reading the information. Terms such as "hits" are not good indicators of website readership as they measure technical aspects of delivering internet information rather than actual readers or readership [196,281]. The Institute differentiates between several main categories of website metrics:

- Measurement of website "traffic" (the number of people who have looked for the website or come across it by chance);
- Determination of the "source" of visitors (search engines and links from other websites);
- Measurement of the number of pages of information viewed; and
- Readership and interaction with the website, including searches.

In order to be able to compare user traffic with that of other websites, the Institute routinely gathers and analyses data on [196,197]:

- The number of individual pages opened by users (*page impressions* or *page views*); and
- The total number of individual website viewing sessions (*visits*)

Page impressions and visits by so-called internet robots (*crawlers*) are excluded, as are the use of the website by the Institute itself and its website development team. Care is taken not to gather data in forms that can identify the user. The Institute's privacy and data protection policy is described in detail on the website [199].

The *traffic* of the website is the total number of people who enter the website, not the number of people who actually read items on it [196]. The Institute therefore monitors and analyses more

critically indicators of the number of people who are apparently actually reading information, for example:

- Searching for information;
- Navigating through articles;
- Clicking on glossary terms and visiting related information;
- Viewing animated films or using quizzes;
- Downloading PDFs; and
- Visiting the website to view new information after receiving the newsletter.

The Institute also monitors the standing of the website in Google, including the positioning of its information in Google searches and the Google “PageRank” of the website itself. The PageRank is a citation-based methodology for rating the importance of a website [57].

In addition, the Institute estimates the extent to which “information multipliers” have adopted the Institute’s health information. This involves some analysis of how many other websites link to www.gesundheitsinformation.de or www.InformedHealthOnline.org, in particular the websites of the German statutory health insurance (SHI) funds. In addition, the number of printed versions of the Institute’s information, particularly those produced by SHI funds, is monitored.

5.5.2 Feedback, corrections and improvements

As well as pre-publication quality assurance, the Institute’s website encourages post-publication comments for improvements by readers: between March and August 2006, over 100 readers offered suggestions for improvements or asked for additional information to be addressed in items of information. In addition, over 3,000 emails containing feedback and suggestions for topics to cover arrived over the website in those six months. Although the Institute does not give medical advice, answer individual health-related questions or distribute any information other than that on its website or other information published by the Institute, reader feedback is an important element of the Institute’s efforts to offer patient-friendly and useful health information.

Amendments to health information are classified as minor, moderate or major. Minor changes only include linguistic improvements, while more major changes are content-related. Standard procedures cover the allocation of these categories and the level of quality assurance required for corrections and improvements associated with them. Each piece of health information is accompanied by a document history online, showing the date and level of change that was made, and each version is archived. Other than minor changes or urgent content changes, information is not usually changed more often than once or twice a year.

5.5.3 *Evaluation*

As well as routine monitoring, the extent to which the website is meeting its goals is assessed by:

- User testing;
- Online user surveys;
- Consultation with users and patient representatives; and
- Independent evaluation.

User testing of individual pieces of information is part of the pre-publication process for some interactive products. These tests are undertaken by members of the Department of Health Information. More formal user testing of the website is undertaken by people outside of the Department. Two user tests were undertaken in 2006, including one specifically on accessibility. Further user tests will be done annually.

From the launch of the website, the online survey used by the University of Bielefeld to evaluate online pilot projects in accordance with §65b of the Social Code Book (SGB V), was implemented with permission of the developers on the Institute's website [323]. This was done to enable benchmarking between www.gesundheitsinformation.de and the more formally evaluated pilot projects (15 websites). Analysis of the first 2,561 completed surveys on www.gesundheitsinformation.de showed that on many criteria, such as understandability, www.gesundheitsinformation.de's rating fell around the median of the 15 other websites.

Methods for evaluating websites are an ongoing area of methodological work for the Department of Health Information. There are many instruments and guidelines for the evaluation of health information and health websites. Many evaluate process but not content, or content but not process [78,81,237]. There is no instrument that has been shown to be a reliable indicator of the quality of health information or health websites [124,143,212,269]. There are major omissions in commonly used instruments: DISCERN, for example, does not address the quality of content [78,237].

Studies with patients, including patients in Germany [125,366], indicate that some of the issues suggested as important in evaluating health information may not in fact be important to most patients. Some recommendations common in such instruments may actually reduce the quality of health information. One example is expecting links to other information and to self-help groups: only a minority of users may rate this as of significance [297], and poor quality in linked information may result in misinformation. Regularly checking links to see if they remain of high quality is a very resource-intensive task that few websites could realistically maintain.

External evaluation, particularly qualitative evaluation by potential users, is important for the ongoing development of the Institute's information and website. External experts are responsible

for the content evaluation of the individual information products and information packets by potential users. Furthermore, the patients who share their stories of illness and health also evaluate the drafts of the patient information that their stories accompany. The outcomes of evaluations are continuously used as feedback to improve the processes for the production and editing of information products.

6 General methodological aspects

In research, the term “bias” refers to a systematic deviation of research results from the “truth” [312]. This could be a misleadingly high (or low) estimate of a treatment’s effect.

A key goal of the evaluation of healthcare is to achieve as reliable and unbiased an estimate of the real effects of treatments and interventions as possible. Internationally, a variety of approaches are used to minimise bias in healthcare evaluation, including scientific rigour, ensuring broad participation and avoiding conflicts of interest [71]. Each of these approaches also forms the legislative basis for the Institute’s work.

6.1 Literature search

Various types of information form the basis of the Institute’s reports (e.g. results of studies on the medical effectiveness and safety of an intervention, registry data and other data collections, or documents from regulatory authorities). This section describes the process of a topic-related search for scientific literature.

6.1.1 *General principles of topic-related literature searches*

A topic-related literature search aims to identify all publications relevant to the research question (i.e. publications that contribute to a gain in knowledge on the topic). The methodology of the systematic literature search therefore follows the general principle that the topic-related literature search concerned must consider the following binding aspects, which, according to the results of research on this issue, have an important influence on the answer to the research question posed. These aspects include:

- Selection of data sources (e.g. public bibliographic databases, private databases, handsearching in selected scientific journals, contacts with experts/industry/patient organisations, etc.);
- Search techniques related to the selection of study types (e.g. RCTs, case reports, etc.);
- Search techniques related to the medical criteria specified by the research question (e.g. target population, type of intervention, outcomes, etc.);
- Search techniques related to formal characteristics of the publication (e.g. abstract publications, language, etc.).

Studies and examples on this topic are provided by numerous publications [112,126,131,152,158,190,191,220,255,260,264,294,308,315,316,340].

The relevance of these criteria varies, depending on the different research questions. The type of product to be prepared (e.g. report, rapid report, working paper) and the resulting timeframe also have an impact on the approach to the literature search.

The selection of databases for each product depends on the focus (i.e. regarding content, methods, and region) of the bibliographic databases. At least two large biomedical databases (e.g. MEDLINE) are always selected. The extent of the handsearch in scientific journals is determined by the existence of relevant journals (e.g. key journals), and the extent of their indexing in the databases considered.

The study type with the minimum evidence level to be considered is defined in the inclusion criteria: this study type and all study types with higher evidence levels are selected for the literature search.

Depending on the topic to be assessed, the medical criteria of the research question aim for high sensitivity and the resulting maxim “as many search components as necessary, as few as possible”. Search strategies for assessments on drug interventions tend to contain fewer components than those for non-drug interventions.

6.1.2 Search procedure for primary publications

The search for primary publications in bibliographic databases consists of the following nine components:

1. If necessary, specification of the research question posed.
2. Modification of the research question to a searchable research question.
3. Formulation of a search concept (e.g. language, period).
4. Selection of data sources.
5. Identification of search terms per component of the search concept.
6. Formulation of the search strategies.
7. Performance of the search.
8. Storage of the search results in text files and import into a reference management software.
9. Documentation of the search strategies.

In each case, it is reviewed whether consultation with external experts would be useful. This may be the case if no specific expertise regarding the research question is available in the Institute.

Relevant publications identified in the preliminary search are usually drawn upon to identify search terms and formulate the search strategy. As a quality assurance step, it is tested whether the search strategy developed in this way identifies known relevant primary publications (test set) with sufficient certainty. The test set is generated by using previous publications of other working groups (HTA reports, systematic reviews).

Under certain conditions, the HTA reports and systematic reviews identified can also be used to limit the Institute's search to areas not covered by these documents (e.g. regarding the search period). The necessary prerequisite for this is that the particular search is in line with the Institute's methodology, and that transferability of the results to the research question is assured, particularly under consideration of the inclusion and exclusion criteria specified in the report plan.

6.1.3 Other data sources for literature searches

Besides bibliographical database searches, it can be useful (depending on the research question) to conduct a handsearch in selected scientific journals. This is decided on a case-by-case basis.

In addition, depending on the research question, further data sources may be of considerable relevance, e.g. study registries or volumes of abstracts from scientific congresses. In the case of drug assessments, but also of assessments of specific (non-drug) medicinal products, publicly accessible drug approval databases or correspondence with regulatory authorities are further potential sources of information. Moreover, the manufacturers of the technologies to be assessed are usually asked to provide previously unpublished information. The aim of this request is to identify all studies and other information relevant to the benefit assessment, independent of their publication status.

For drug assessments, this request is usually made in two steps. In the first step, the Institute asks the manufacturer to supply a complete overview of all studies conducted on the drug to be assessed. If appropriate, the Institute defines project-specific inclusion criteria for this overview. In the second step, the Institute identifies studies relevant to the benefit assessment from the overview provided and requests detailed information on these studies. This may refer to a request for unpublished studies, or for supplementary, previously unpublished information on published studies. Previously unpublished information considered in the benefit assessment will also be published in the Institute's reports in order to ensure transparency. The basis for the incorporation of previously unpublished information in the benefit assessment is the conclusion of an agreement on the transfer and publication of study information, concluded between the Institute and the manufacturer concerned before the submission of data (see sample contract [198]). This agreement specifies the procedure, the requirements for the documents to be submitted, as well as the

confidential and non-confidential components of the documents submitted. If the manufacturer concerned does not agree to the conclusion of this contract and therefore in particular does not agree to the complete transfer of all information requested by the Institute, or does not completely transfer the information requested despite conclusion of an agreement, no further requests to the manufacturer will be made. This is to prevent bias of results through selective provision of information.

If documents are provided by the contracting agency (Federal Joint Committee or Federal Ministry of Health), they are regarded as a component of the literature search. In the subsequent procedure these documents are handled following the other principles of the literature search and assessment.

6.1.4 Selection of relevant publications

Due to the primarily sensitive approach, the literature search in bibliographic databases results in a large number of citations that are not relevant to the assessment. The selection of relevant publications is usually made in several steps:

- Exclusion of definitely irrelevant publications (i.e. publications not fulfilling the inclusion or exclusion criteria of the report plan or project outline) through perusal of the titles, and, if available, the abstracts.
 - This step can be divided into two in order to distinguish completely irrelevant publications from topic-related ones which, however, do not fulfil the inclusion or exclusion criteria. “Topic-related” refers, for example, to studies that investigate the topic of interest but have a different study design or duration from that specified in the report plan or project outline.
- The full texts of the remaining potentially relevant publications are obtained. The decision on the inclusion of the study in the assessment concerned is then made on the basis of these documents.

All selection steps are, as a matter of principle, performed by two persons independently of each other. Discrepancies are resolved by discussion. In the first selection step, if doubts exist as to the relevance of a study, the corresponding full text is obtained and assessed. In this step, completely irrelevant publications may also be distinguished from topic-related ones.

The languages of publication are usually restricted to those of Western Europe. However, other foreign-language publications may also be included if the available information on these publications indicates that additional and relevant information answering the research question posed is to be expected.

6.1.5 Documentation

All steps in the literature search are documented. This especially includes:

- The search strategy for the databases selected;
- The search date;
- The user interface;
- The number of hits;
- After perusal of the hits: documentation of the publications judged relevant to the research question posed (citations);
- After perusal of the full texts: documentation of the citations not judged relevant; alternatively, documentation of the topic-related publications that were, however, irrelevant for the report (in each case providing a reason for exclusion).

All other steps of the literature search are also documented (e.g. correspondence with authors, queries to manufacturers, etc.).

6.1.6 Literature monitoring

Besides topic-related retrospective searching, early detection and assessment of up-to-date and relevant publications are necessary, based on the systematic monitoring of important scientific data sources. The term “data sources” includes not only scientific journals, but also the lay press, daily, weekly and monthly press, electronic media, etc.

A publication is especially classified as “relevant” in this regard if:

- The publication is likely to have a considerable influence on the current health care situation;
- The publication is considered to be a milestone study;
- A topic of prime public interest is discussed in the publication;
- The publication refers to a topic on the Institute’s internal priority and project list.

The individual departments evaluate data sources and forward relevant publications to the other departments. A publication may result in an ad hoc assessment, including an official comment from the Institute, or may initiate the preparation of a working paper. This is decided by the Institute’s Steering Committee.

6.1.7 Consideration of legal aspects of data protection/confidentiality

The processing of personal data within the Institute is conducted according to the relevant federal data protection laws.^w The data protection officer appointed by the Institute is responsible for ensuring compliance with these laws.

The Institute may in future also process personal data (attributed to an *identifiable* individual) obtained from research projects. In exceptional cases, personal data attributed to an *identified* individual may be used. If personal data were originally collected or are being collected by a third party, the corresponding declarations on compliance with legal regulations need to be provided; otherwise, for each individual case the fulfilment of the necessary legal requirements (informed consent, patient information, etc.) needs to be thoroughly reviewed in advance.

A further aim is to obtain personal data primarily attributed to an identified individual in an anonymous or pseudonymous form from third parties and process them. In most cases, it will be sufficient to use data thus coded for research purposes and individual research questions. In this way, any possible reservations about transferring data to the Institute could be dispelled.

If data not allowed to be published are transferred to the Institute, these data cannot be considered in the Institute's benefit assessments, as this would contradict the obligation for transparency.

With regard to the confidential handling of data from commercial enterprises, appropriate declarations ensuring the Institute's confidentiality will be made to third parties if necessary (see also Section 6.1.3). Not only does the Institute have the necessary technical infrastructure to ensure data safety, but it also includes clauses in all its employment contracts obliging personnel to observe confidentiality. In individual cases, externally appointed persons or institutions must also undertake corresponding obligations towards the Institute.

6.2 Quality assessment of individual studies

6.2.1 Relationship between study type and research question

Only the most relevant study designs that play a role in benefit assessments in medical research (depending on the research question posed) are summarised here.

It is primarily the inclusion of a control group that is called for in the benefit assessment of interventions. In a design with dependent samples without a control group, proof of the effect of an

^w German data protection laws distinguish between personal data attributed to an identified individual (e.g. name, address) and personal data attributed to an identifiable individual (e.g. medical diagnosis). The respective German terms are "personenbezogene Daten" and "personenbeziehbare Daten".

intervention cannot usually be inferred from a pure “before-after” comparison. Exceptions include diseases with a deterministic (or practically deterministic) course (e.g. ketoacidotic diabetic coma; see Section 3.2.2). Randomisation and blinding are quality criteria that increase the evidential value of controlled studies. Parallel group studies [295], cross-over studies [217], and cluster randomised studies [102] are common designs for clinical trials. If interim analyses are planned, the use of appropriate sequential designs must be considered [379].

Case reports or case series often provide initial information on a topic. These are susceptible to all kinds of bias, so that, depending on the research question, only limited reliable evidence can be inferred from this type of study. The prevalence of diseases can be estimated from population-based cross-sectional studies. Other fundamental and classical study types in epidemiology are case-control studies [55] to investigate the association between exposures and the occurrence of rare diseases, as well as cohort studies [56] to investigate the effect of an exposure over time. Cohort studies designed for this purpose are prospective, although retrospective cohort studies are also conducted in which past exposure is recorded (this type of study is frequently found in occupational or pharmacological epidemiology). In principle, prospective designs are preferable to retrospective designs. However, case-control studies, for example, are frequently the only feasible way of obtaining information on associations between exposures and rare diseases. Newer study designs in modern epidemiology contain elements of both case-control and cohort studies and can no longer be clearly classified as retrospective or prospective [223].

Diagnostic and screening studies may have very different aims, so that the assessment depends on the choice of an appropriate design (see Sections 3.5 and 3.6).

6.2.2 Ranking of different study types/evidence levels

Different approaches exist within the framework of systematic reviews or guideline development for allocating specific evidence levels to particular study types [169,173]. These levels can be used to create a ranking with regard to the validity of evidence from different study types. However, no evidence assessment system currently exists that is generally accepted and universally applicable to all systematic reviews [224]. Due to the complexity of the appraisal of studies, no conclusive judgement on quality can be inferred from the hierarchy of evidence. In general, the Institute follows the rough hierarchy of study types described below, which is widely accepted and is also largely consistent with the evidence classification of the Federal Joint Committee [145]. The highest evidence level is allocated to RCTs and systematic reviews of RCTs, at least within the framework of therapeutic studies. In some classifications, individual RCTs are further graded into those of higher or lower quality. In this context, the conflation of the quality of concept and the

quality of results has been criticised by some authors [384]. The next levels include non-randomised intervention studies, prospective observational studies, retrospective observational studies, non-experimental studies (case reports and case series) and, at the lowest evidence level, expert opinions not based on scientific rationale. The Institute will adapt this rough grading system to the particular situation and research question and, if necessary, present it in more detail [173].

6.2.3 Aspects of the assessment of the bias potential of results

One main aspect of the interpretation of study results is the assessment of the bias potential (see qualitative uncertainty of results, Section 3.2.1). In this context, the research question, the study type and design, and the conduct of the study play a role, as well as the availability of information. The bias potential is substantially affected by the study quality; however, its assessment is not equivalent to the quality assessment of a study. For example, individual outcomes may also be considerably biased in a high-quality study. Other studies, however, may provide high certainty of results for specific outcomes in individual cases, despite being of low quality. As a rule, the Institute will therefore estimate the extent of the bias potential in a problem-orientated manner for all relevant results (both for the study and the specific outcomes).

In principle, a recognised standardised concept should be followed in a study; from planning to conduct, data analysis, and reporting. This includes a study protocol describing all the important methods and procedures. For (randomised) clinical trials, the usual standards are defined by the basic principles of good clinical practice (GCP) [202,234]; for epidemiological studies, they are defined by guidelines and recommendations to ensure good epidemiological practice (GEP) [91]. In this context, a key criterion to avoid bias is whether the study was actually analysed in the way planned. This cannot usually be reliably concluded from the relevant publications. However, a section on sample size planning may at least provide indications in this regard. In addition, a comparison with the study protocol (possibly previously published) or with the corresponding publication on the study design is useful.

The following important documents were developed to improve the quality of publications:

- The CONSORT^x statement on randomised clinical trials [265] and the corresponding explanatory document [12];
- A proposal for an extension of the CONSORT statement for randomised studies on non-drug interventions [53] and the corresponding explanatory document [52];

^x Consolidated Standards of Reporting Trials

- The CONSORT statement on cluster randomised trials [65];
- The QUORUM^y statement on meta-analyses of randomised trials [263];
- The TREND^z statement on non-randomised intervention trials [89];
- The STROBE^{aa} statement for observational studies in epidemiology [373] and the corresponding explanatory document [369];
- The MOOSE^{bb} checklist for meta-analysis of observational studies in epidemiology [349];
- The STARD^{cc} statement on diagnostic studies [50] and the corresponding explanatory document [51].

If a publication fails to conform to these standards, this may be an indicator of an increased bias potential of the results of the relevant study. Additional key publications on this issue describe fundamental aspects concerning the assessment of bias potential [110,168,185].

Key aspects of the Institute's assessment of the bias potential of the results of RCTs comprise:

- Adequate concealment, i.e. the unforeseeability and concealment of allocation to groups (e.g. by external randomisation in trials that cannot be blinded);
- Blinded outcome assessment in trials where blinding of physicians and patients is not possible;
- Appropriate application of the “intention-to-treat” principle;
- Specification of a clear primary outcome or appropriate consideration of potential multiple testing problems.

The overall effect of an intervention consists of various components: the actual effect to be ascribed to the intervention alone, the natural course of disease, and the “placebo effect” [354] (which describes the context in which the treatment takes place). This summarises all the influences accompanying an intervention, for example, the expectations of the patient and treating staff, the suggestive behaviour of those treating, or any effect deriving from the simple fact that patients are being cared for. Placebo-controlled studies serve to balance the influence of these concomitant effects in both groups by including a group that receives placebo therapy. The decisive factor in this case is the possibility of blinding patients and treating staff with regard to the

^y Quality of Reporting of Meta-analyses

^z Transparent Reporting of Evaluations with Non-randomized Designs

^{aa} Strengthening the Reporting of Observational Studies in Epidemiology

^{bb} Meta-analysis of Observational Studies in Epidemiology

^{cc} Standards for Reporting of Diagnostic Accuracy

intervention. The blinding of treating staff is designed to ensure that the care, attention, and suggestive behaviour going beyond the intervention itself are equally distributed between groups.

Placebo treatment is not limited to drug interventions alone, but can also be applied in non-drug interventions, at least to achieve the blinding of patients. In this respect one also speaks of “sham interventions”. The extent of a placebo effect can depend on the type of intervention [222].

During the course of a trial it can happen that in certain interventions, despite original blinding, an unblinding of patients and treating staff occurs, for example, if there are specific adverse effects of the therapy under investigation. A possible unblinding or the omission of a blinding procedure in the first place can lead to a bias of results. The exact extent of this bias will not usually be determinable. To achieve this, an unbiased estimate of the degree of the placebo effect would be necessary, for example, by including a third group that does not receive any intervention at all. However, particularly in situations where unblinding occurs, the placebo effect cannot be estimated without bias, as the difference between the placebo group and the group without an intervention is also biased. If, for example, it can be assumed that a possible placebo effect is substantially reduced after unblinding of patients in the placebo group, the difference between the placebo group and the group without an intervention would present too small an estimate of the placebo effect.

Despite the problems described, the Institute discusses the validity of the results of unblinded studies (or studies in which unblinding occurred), in case indications of a possible major placebo effect exist.

It has been proposed to assess the success of blinding in end-of-trial tests by comparing how many patients and treating staff correctly guess treatment assignment [130]. Such assessments, however, show methodological problems that have not yet been satisfactorily solved and are a matter of controversy (e.g. Which approach should be applied for which null hypotheses? [11,313,333]). In the case of effective therapies with effects that can be directly experienced by patients, unblinding is possible or even probable to a certain extent. This means that in such a situation it is difficult or even impossible to judge whether unblinding (for whatever reason) led to a biased therapy effect, or whether conversely the therapy effect led to unblinding. Despite these problems, verification of the blinding of a trial is to be welcomed, as it indicates that this issue has been given appropriate consideration in the planning and conduct of a study.

There must be a more cautious interpretation of the results of unblinded trials, or of trials where unblinding (possibly) occurred, compared with the interpretation of blinded studies. Randomisation and the choice of appropriate outcome variables are important instruments to prevent bias in studies where a blinding of the intervention was not possible. In studies that cannot be blinded, it is crucial to ensure adequate concealment of the allocation of patients to the groups to be compared. It is also

necessary that the outcome variable is independent of the (non-blinded) treating staff or assessed in a blinded manner independent of the treating staff (blinded assessment of outcomes). If a blinded assessment of outcome measures is not possible, a preferably objective outcome should be chosen which can be influenced as little as possible (with regard to its dimension and the stringency of its recording) by the (non-blinded) person assessing it.

If an IQWiG project involves the assessment of older studies that do not satisfy current quality standards because they were planned and conducted at a time when these standards did not exist, then the Institute will present the disadvantages and deficiencies of these studies and discuss possible consequences. A different handling of these older studies compared with the handling of newer studies that have similar quality deficits is however only necessary if this is clearly justifiable from the research question posed or other circumstances of the assessment.

The assessment of formal criteria provides essential information on the bias potential of the results of studies. However, the Institute always conducts an assessment of the bias potential that goes beyond purely formal aspects in order, for example, to present errors and inconsistencies in publications, and to assess their relevance in the interpretation of results.

6.2.4 Interpretation of subgroup analyses

In the methodological literature, subgroup analyses are a matter of controversy [21,287]. The interpretation of their results is complicated mainly by three factors:

- No characteristic of proof: subgroup analyses are rarely planned a priori, and are rarely a component of the study protocol (or its amendments). If post-hoc subgroup analyses are conducted, the results cannot be regarded as a methodologically correct testing of a hypothesis.
- Multiple testing: if several subgroups are analysed, results in a subgroup may well reach statistical significance, despite actually being random.
- Lack of power: the sample size of a subgroup is often too small to enable the detection of moderate differences (by means of inferential statistics), so that even if effects actually exist, significant results cannot be expected. The situation is different if an adequate power for the subgroup analysis was already considered in the sample size calculation and a correspondingly larger sample size was planned [60].

If one or more of the three factors mentioned above are present, the results of subgroup analyses should only be considered in the assessment with strong reservations and should not dominate the result of the primary analysis, especially if the primary study objective was not achieved.

Moreover, subgroup analyses are not interpretable if the subgroup-forming characteristic was defined after initiation of treatment (after randomisation), e.g. in responder analyses.

The statistical demonstration of different effects between various subgroups should be conducted by means of an appropriate homogeneity or interaction test. The finding that a statistically significant effect was observed in one subgroup, but not in another, cannot be interpreted (by means of inferential statistics) as the existence of a subgroup effect.

Despite the limitations specified above, for some research questions subgroup analyses may represent the best scientific evidence available in the foreseeable future in order to assess effects in subgroups [139], since factors such as ethical considerations may argue against the reproduction of findings of subgroup analyses in a validation study. Rothwell presents an overview of indications for applying subgroup analyses [305].

Possible heterogeneity of an effect in different, clearly distinguishable patient populations is an important reason for conducting subgroup analyses [238,305]. If a priori information is available on a possible effect modifier (e.g. age, pathology), it is in fact essential to investigate possible heterogeneity in advance with regard to the effect in the various patient groups. If such heterogeneity exists, then the estimated total effect across all patients cannot be interpreted meaningfully [238]. It is therefore important that information on a possible heterogeneity of patient groups is considered appropriately in the study design. It may even be necessary to conduct several studies [164]. Within the framework of systematic reviews, the analysis of heterogeneity between individual studies is a scientific necessity.

The gold standard for subgroup analyses are analyses in which the subgroup was specified a priori. This approach includes the use of stratified randomisation on the basis of these subgroups, and the application of an appropriate statistical method (homogeneity test, interaction test) for the data analysis [80].

Taking into account the above-named factors, the Institute interprets results of subgroup analyses very cautiously. However, it does not exclude them from the assessment as a matter of principle.

6.2.5 Assessment of prognosis studies

An essential basis for the assessment of prognosis studies is the precise formulation of a research question, as studies conducted to evaluate prognostic characteristics have different objectives (e.g. evaluation of risk factors, score development or validation). The discrimination between prognosis studies and diagnostic and/or screening studies can be difficult. Depending on the study objective, in the quality assessment of prognosis studies, different assessment principles are applied.

A prognostic characteristic provides information that should not be an end in itself, but should have a consequence that constitutes a verifiable benefit for the patient. In this context, the (general) requirements applying to a prognostic procedure are similar to those applying to a diagnostic test. If a prognostic characteristic is to be applied in the sense of a screening or prevention programme, then the principles formulated in Section 3.6 need to be considered in the assessment.

No generally accepted quality criteria exist for the assessment of prognosis studies [8,178,339]. Simon and Altman [339] describe guidelines for the planning and conduct of prognosis studies in oncology. Laupacis et al. [244] suggest a general framework for the quality assessment of prognosis studies. Hayden et al. [178] developed guidelines for the quality assessment of prognosis studies with regard to potential sources of bias. The points listed below, which result from the underlying data source as well as the data analysis applied, should always be considered. As multifactorial regression models often play a central role in prognosis studies, Section 6.4.6 should also be taken into account. The following points are especially relevant:

- Clear formulation of a research question and the corresponding planning of the study. This includes sample size planning, which can for example be orientated towards the desired precision of the estimate (width of the confidence interval), and requires an estimate of both the prevalence and incidence of the exposure regarding the outcome variable concerned.
- Clear description of the target and sample population (e.g. population-, register- or general practitioner-based) and justification of their selection.
- Clear description of the selection of and the recruitment procedure for study participants.
- Homogeneity of the population investigated. If the population is heterogeneous, it needs to be considered that a prognostic statement can be made as constantly as possible across the subgroups causing heterogeneity (e.g. existence of different baseline risks for the outcome variable in question).
- Clear definition of the outcome variable(s) towards which the prognostic significance should be orientated.
- Clear definition of the prognostic characteristics, including their statistical handling (e.g. dichotomisation or assessment of terziles or quartiles, etc., for a quantitative characteristic), and justification of the procedure selected.
- Clear specification and definition of potential confounders and interactions, including their statistical handling.

- For cohort studies: completeness of follow-up or measures to achieve as complete a follow-up as possible. Estimation of possible selection effects if follow-up is incomplete.
- When assessing prognostic scores, it should be noted that a distinction is made between score development and score validation, e.g. score development within a “learning sample” and validation in a test sample. Ideally, score development and score validation are carried out in different studies.

Typical study designs for the evaluation of prognostic characteristics in terms of risk factors include cohort studies and case-control studies. In exceptional cases (e.g. when investigating constant characteristics), cross-sectional studies may also play a role. The basic principles for the assessment of such studies beyond the aspects mentioned above are described in Section 6.2.3.

The literature search for the evaluation of prognostic characteristics (within the framework of a systematic review) is more difficult than for example for therapeutic studies, and no generally accepted optimum search strategy exists (yet). Furthermore, it is assumed that this research field is especially susceptible to publication bias [8,339]. The methodological quality of studies (or their publications) on prognostic characteristics is frequently insufficient [296], so that the extraction of required data is difficult or even impossible. Meta-analyses of prognosis studies (not, however, systematic reviews per se) are therefore often inappropriate and their findings should be utilised with reservation [8]. Some important problems with meta-analyses of prognosis studies can be avoided if individual patient data (IPD) are available [8].

Besides using the results of studies investigating single or (mainly) multiple prognostic characteristics, risk charts (also called risk engines) are being increasingly used to assess the individual risk of patients (or clinically healthy persons) of experiencing an adverse event. Multi-factorial estimates for the concurrence of several risk factors are made in these charts (e.g. the Sheffield Table [375] or the Joint British Chart [58]).

The basis for these risk charts are mainly formed by multi-factorial regression models, whose results, for easier handling, are presented in tables or points systems [350]. It should be noted that risks derived for such charts are not “personal” estimates for specific individuals, but statistical estimates of the average risks of a population with a specific risk profile for a defined period (e.g. ten years). The following factors should be considered when assessing such instruments:

- Which population the estimated risks apply to;
- What type of study the underlying data originate from;
- Whether the variables included were analysed together in these studies;

- Whether, and if so, how a multi-factorial statistical analysis was conducted in these underlying studies;
- Whether these instruments were ever validated in subsequent studies (test samples).

6.2.6 Assessment of the consistency of published data

To assess the evidential value of published results, the Institute will review the consistency of data with regard to their plausibility and completeness. Implausible data are not only produced by incorrect reporting of results (typing, formatting, or calculation errors), but also by the insufficient or incorrect description of the methodology, or even by forged or invented data [7]. Inconsistencies may exist within a publication, and also between publications on the same study.

One problem with many publications is the reporting of incomplete information in the methods and results sections. In particular, the reporting of lost-to-follow-up patients, withdrawals, etc., as well as the way these patients were considered in the analyses, are often not transparent.

It is therefore necessary to expose potential inconsistencies in the data. For this purpose, the Institute reviews, for example, calculation steps taken, and compares data presented in text, tables, and graphs. In practice, a common problem in survival-time analyses arises from inconsistencies between the data on lost-to-follow-up patients and those on patients at risk in the survival curve graphs. For certain outcomes (e.g. total mortality), the number of lost-to-follow-up patients can be calculated if the Kaplan-Meier estimates are compared with the patients at risk at a point in time before the minimum follow-up time. Statistical techniques may be useful in exposing forged and invented data [7].

If relevant inconsistencies are found in the reporting of results, the Institute's aim is to clarify these inconsistencies and/or obtain any missing information by contacting authors, for example, or requesting the complete clinical study report and further study documentation. However, it should be considered that firstly, enquiries to authors often remain unanswered, especially concerning older publications, and that secondly, authors' responses may produce further inconsistencies. In the individual case, a weighing-up of the effort involved and the benefit of such enquiries is therefore meaningful and necessary. If inconsistencies cannot be resolved, the potential impact of these inconsistencies on effect sizes (magnitude of bias), uncertainty of results (increase in error probability), and precision (width of the confidence intervals) will be assessed by the Institute. For this purpose, sensitivity analyses may be conducted. If it is possible that inconsistencies may have a relevant impact on the results, this will be stated and the results will be interpreted very cautiously.

6.3 Consideration of secondary literature

In this section, the term “secondary literature” refers to publications that summarise and assess the results of primary studies in a systematic, reproducible, and transparent way. These include both systematic reviews and HTA reports that usually attempt to answer a clinical or patient-relevant question. HTA reports also often seek to answer additional questions of interest to policymakers [103,245,291]. There is no need to differentiate between systematic reviews and HTA reports for the purposes of this section. Therefore, for simplicity, HTA reports are included in the term “systematic review”.

6.3.1 Classification of systematic reviews

Relying on individual research studies can be misleading. Looking at one or only some studies in isolation from other similar studies on the same question can make treatments appear to be more or less useful than they actually are [213]. High quality systematic reviews aim to overcome this form of bias by identifying, assessing and summarising evidence systematically rather than selectively [103,110,149,291].

Systematic reviews identify, assess and summarise evidence from the type of primary study or studies that can best answer a specific and clearly formulated question. Systematic and explicit methods are then used to identify, select and critically assess the studies relevant to that question. When studies are identified, data are collected and analysed systematically. Systematic reviews are observational studies, and their methodology must aim to minimise bias at each stage of the reviewing process [110,185,213].

For systematic reviews of the effects of healthcare treatments, randomised controlled trials provide the most reliable answers. However, for other questions such as aetiology, prognosis or the qualitative description of patients’ experiences, other primary study types will provide the appropriate evidence base for a systematic review [149]. Systematic reviews of tests used for diagnosis and screening also show some methodological differences from reviews of treatments [85].

In the production of the Institute’s reports, systematic reviews are primarily used to identify potentially relevant (primary) studies. However, an IQWiG report can be based partially or even solely on systematic reviews (see Section 6.3.2). Health information produced by the Institute for patients and consumers is to a large part based on systematic reviews. This includes systematic reviews of treatments, and reviews addressing other questions such as aetiology, adverse effects and syntheses of qualitative research (see Section 5.3.3).

The minimal prerequisite for a systematic review on the effects of treatments to be used by the Institute is that it has only minimal methodological flaws according to the Oxman and Guyatt index [210,286,288]. In addition to considering the strength of evidence within systematic reviews, the Institute will also consider the relevance and applicability of the evidence. This includes considering whether the results have been consistent among different populations and subgroups as well as in different healthcare contexts. The following issues are usually considered: the population of the participants in the included studies (including gender and baseline disease risk); the healthcare context (including the health settings and the professionals providing care); and the applicability and likely acceptability of the intervention in the form in which it was assessed [47,83].

6.3.2 Benefit assessment on the basis of systematic reviews and HTA reports

A benefit assessment on the basis of systematic reviews and HTA reports (both sometimes referred to as “secondary literature” in the following text) can provide a resource-saving and reliable evidence base for recommendations to the Federal Joint Committee or the Ministry of Health, provided that specific preconditions have been fulfilled. The applicability of a benefit assessment prepared on the basis of systematic reviews and HTA reports depends on the availability of secondary literature that, in order to draw a clear conclusion, [22,286]:

- Is of sufficiently high quality and shows only minimum potential for bias;
- Is directly relevant;
- Includes high-quality primary studies.

The methodology applied must provide sufficient certainty that a new benefit assessment based on primary literature would not reach different conclusions from one based on systematic reviews and HTA reports. For example, this is usually not the case if a relevant amount of previously unpublished data is to be expected.

A) Research questions

In principle, this method is suited for all research questions insofar as the criteria named above have been fulfilled. The following points should be given particular consideration in the development of the research question:

- Definition of the population of interest;
- Definition of the test intervention and comparator intervention of interest;
- Definition of all relevant outcomes;

- If appropriate, specification of the health care setting or region affected (e.g. Germany, Europe).

The research question defined in this way also forms the basis for the specification of the inclusion and exclusion criteria to be applied in the benefit assessment, and subsequently for the specification of the relevance of the content and methods of the publications identified. On the basis of the research question, it is also decided which type of primary study the systematic reviews and HTA reports must be based on. Depending on the research question, it is possible that questions concerning certain parts of a commission are answered by means of systematic reviews and HTA reports, whereas primary studies are considered for other parts.

B) Literature search strategy including a search update/supplementation

The search does not aim to achieve completeness in terms of the complete consideration of all primary studies. However, a high robustness of results is required, i.e. sufficient certainty that the results will not be greatly affected by the inclusion of an additional study.

Development of a literature search:

The basic procedure for literature searches is presented in Section 6.1.

Database selection:

Some of the sources to be considered in literature searches for systematic reviews/HTA reports are different from those considered in the search for primary literature. As a rule, databases are primarily searched that exclusively or largely contain secondary literature (including HTA reports). In addition, a selection of biomedical databases which also (but not primarily) contain secondary literature are searched (e.g. MEDLINE and EMBASE).

Depending on the topic investigated, it is decided whether and, if so, which databases or other sources (e.g. websites of HTA agencies) are also relevant and should be searched. CPGs are not categorically excluded as an information source. However, a systematic search for CPGs is not usually conducted. HTAs that are not free of charge are considered in exceptional cases, if, for example, it is assumed that additional information can be retrieved from them, or if no information is otherwise available.

Selection of relevant publications:

The selection of relevant publications is conducted following the selection process for primary publications outlined in Section 6.1.4.

Minimum number of relevant systematic reviews:

All systematic reviews and HTA reports that are of sufficient quality and relevant to the topic are considered. In order to be able to assess the consistency of results, at least two high-quality publications (produced independently of each other) should as a rule be available as the foundation of a report based on secondary literature. If only one high-quality publication is available and can be considered, then it is necessary to justify the conduct of an assessment based only on this one systematic review/HTA report.

Search update/supplementation:

In most cases, a supplementary literature search for current primary studies will be necessary (search update), which covers the period between the time of the searches performed in the secondary literature and the production of the IQWiG report. For the benefit assessment of interventions, a search update for primary literature can only be forgone in exceptional cases with a specified justification. In addition, it may be necessary to conduct supplementary searches for primary literature regarding specific research questions not investigated in the secondary literature.

C) Quality assessment of publications, including minimum requirements

The assessment of the general quality of secondary literature is performed with Oxman and Guyatt's validated quality index for systematic reviews [285,286,288]. According to this index, systematic reviews are regarded to be of sufficient methodological quality if they have been awarded at least five of seven possible points in the overall assessment, which is performed by two reviewers independently of one another. In addition, as a rule, the sponsors of systematic reviews and HTA reports, as well as conflicts of interests of authors, are documented and discussed. Depending on the requirements of the project, the Oxman and Guyatt checklist can be supplemented by additional items (completeness of the search, additional aspects regarding secondary literature on diagnostic studies, etc.).

D) Results

For each research question, the results of a benefit assessment based on systematic reviews and HTA reports are summarised in tables, where possible. If inconsistent results on the same outcome are evident in several publications, possible explanations for this heterogeneity are described [211].

If the compilation of systematic reviews and HTA reports on a topic indicates that a new benefit assessment on the basis of primary studies could produce different results, then such an assessment will be performed.

E) Conclusion/recommendations

Reports based on secondary literature summarise the results of the underlying systematic reviews and HTA reports and, if necessary, they are supplemented by a summary of up-to-date primary studies (or primary studies on questions not covered by secondary literature). Independent conclusions are then drawn from these materials.

The recommendations made on the basis of secondary literature are not founded on a summary of the recommendations or conclusions of the underlying systematic reports or HTA reports. In HTA reports, they are often formulated against the background of the specific socio-political and economic setting of a particular health care system, and are therefore rarely transferable to the health care setting in Germany.

6.3.3 Consideration of published meta-analyses

Following international EBM standards, IQWiG's assessments are normally based on a systematic search for relevant primary studies, which is specific to the research question posed. If it is indicated and possible, results from individual studies identified are summarised and evaluated by means of meta-analyses. However, the Institute usually has access only to aggregated data from primary studies, which are extracted from the corresponding publication or the clinical study report provided. Situations exist where meta-analyses conducted on the basis of IPD from relevant studies have a higher value (see Section 6.4.7). This is especially the case if, in addition to the effect caused solely by the intervention, the evaluation of other factors possibly influencing the intervention effect is also of interest (interaction between intervention effect and covariables). In this context, meta-analyses including IPD generally provide greater certainty of results, i.e. more precise results not affected by ecological bias, when compared with those based on aggregated data [338]. In individual cases, these analyses may lead to more precise conclusions, particularly if heterogeneous results exist that can possibly be ascribed to different patient characteristics. However, one can only assume a higher validity of meta-analyses based on IPD if such analyses are actually targeted towards the research question of the IQWiG assessment and also show a high certainty of results. The prerequisite for the assessment of the certainty of results of such analyses is maximum transparency; this refers both to the planning and to the conduct of analyses. Generally valid aspects that are relevant for the conduct of meta-analyses are outlined, for example, in a document published by the European Medicines Agency (EMA) [76]. In its benefit assessments, IQWiG considers published meta-analyses based on IPD if they address (sub)questions in IQWiG reports that cannot be answered with sufficient certainty by meta-analyses based on aggregated data. In addition, high certainty of results for the particular analysis is required.

6.4 Specific statistical aspects

6.4.1 *Description of effects and risks*

The description of intervention or exposure effects needs to be clearly linked to an explicit outcome variable. Consideration of an alternative outcome variable also alters the description and size of a possible effect. The choice of an appropriate effect measure depends in principle on the measurement scale of the outcome variable in question. For continuous variables, effects can usually be described using mean values and differences in mean values (if appropriate, after appropriate weighting). For categorical outcome variables, the usual effect and risk measures of 2x2 tables apply [36]. Chapter 9 of the *Cochrane Handbook for Systematic Reviews of Interventions* provides a well-structured summary of the advantages and disadvantages of typical effect measures [86]. Agresti [4,5] describes the specific aspects to be considered for ordinal data.

It is essential to describe the degree of statistical uncertainty for every effect estimate. For this purpose, the calculation of the standard error and the presentation of a confidence interval are methods frequently applied. Whenever possible, the Institute will state appropriate confidence intervals for effect estimates, including information on whether one- or two-sided confidence limits apply, and on the confidence level chosen. In medical research, the two-sided 95% confidence level is typically applied; in some situations, 90% or 99% levels are used. Altman et al. [10] give an overview of the most common calculation methods for confidence intervals.

In order to comply with the confidence level, the application of exact methods for the interval estimation of effects and risks should be considered, depending on the particular data situation (e.g. very small samples) and the research question posed. Agresti [6] provides an up-to-date discussion on exact methods.

6.4.2 *Evaluation of statistical significance*

With the help of statistical significance tests it is possible to test hypotheses formulated a priori with control for type 1 error probability. The convention of speaking of a “statistically significant result” when the p -value is lower than the significance level of 0.05 ($p < 0.05$) may often be meaningful. Depending on the research question posed and hypothesis formulated, a lower significance level may be required. Conversely, there are situations where a higher significance level is acceptable. The Institute will always explicitly justify such exceptions.

A range of aspects should be considered when interpreting p -values. It must be absolutely clear which research question and data situation the significance level refers to, and how the statistical hypothesis is formulated. In particular, it should be evident whether a one- or two-sided hypothesis applies [45] and whether the hypothesis tested is to be regarded as part of a multiple hypothesis testing problem [360]. Both aspects, whether a one- or two-sided hypothesis is to be formulated,

and whether adjustments for multiple testing need to be made, are a matter of repeated controversy in scientific literature.

Regarding the hypothesis formulation, a two-sided test problem is traditionally assumed. Exceptions include non-inferiority studies. The formulation of a one-sided hypothesis problem is in principle always possible, but requires precise justification. In the case of a one-sided hypothesis formulation, the application of one-sided significance tests and the calculation of one-sided confidence limits are appropriate. For better comparability with two-sided statistical methods, some guidelines for clinical trials require that the typical significance level should be halved from 5% to 2.5% [200]. The Institute follows the central principle that the hypothesis formulation (one- or two-sided) and the significance level must be specified clearly a priori. In addition, the Institute will justify deviations from the usual specifications (one-sided instead of two-sided hypothesis formulation; significance level unequal to 5%, etc.) or consider the relevant explanations in the primary literature.

If the hypothesis investigated clearly forms part of a multiple hypothesis problem, appropriate adjustment for multiple testing is required if the type I error is to be controlled for the whole multiple hypothesis problem. Bender and Lange provide an overview of the situations where this case applies and describe the methods available for this purpose [40]. The specific aspects of dealing with multiplicity issues in systematic reviews are outlined by Bender et al. [37]. If meaningful and possible, the Institute will apply methods to adjust for multiple testing. In its benefit assessments (see Section 3.1), the Institute attempts to control type I errors separately for the conclusions on every single benefit outcome. A summarising evaluation is not usually conducted in a quantitative manner, so that formal methods for adjustment for multiple testing cannot be applied here either.

The Institute does not evaluate a statistically non-significant finding as evidence of the absence of an effect (absence or equivalence) [9]. For the demonstration of equivalence, the Institute will apply appropriate methods for equivalence hypotheses.

In principle, Bayesian methods may be regarded as an alternative to statistical significance tests [342,343]. Depending on the research question posed, the Institute will, where necessary, also apply Bayesian methods.

6.4.3 Evaluation of clinical relevance

In principle, the clinical relevance of an effect or risk cannot be derived from a p -value. Statistical significance is a statement of probability, which is not only influenced by the size of a possible effect but also by data variability and sample size. When interpreting the relevance of p -values,

particularly the sample size of the underlying study needs to be taken into account [307]. In a small study, a very small p -value can only be expected if the effect is marked, whereas in a large study, highly significant results are not uncommon, even if the effect is extremely small [128,195]. Consequently, the clinical relevance of a study result can by no means be derived from a p -value alone.

Widely accepted methodological procedures for evaluating the clinical relevance of study results do not yet exist. Only a few guidelines contain information on the definition of relevant or irrelevant differences between groups. A first approach to assess the clinical relevance of study results is the evaluation of the effect estimate and the corresponding confidence interval using medical expertise. A formal relevance criterion may be the assessment of the lower confidence limit (in the case of favourable effects) for the effect estimate, or the application of a statistical test, shifting the null hypothesis in order to statistically demonstrate clinically relevant effects [382]. A further option is to formulate a relevance criterion individually, e.g. in terms of a responder definition [229]. Moreover, the individual judgement of affected patients plays an important role. In individual cases, the presentation of patient-relevant outcomes may provide information for this purpose. The Institute will perform the evaluation of clinical relevance in a problem-orientated manner, taking these aspects into account.

6.4.4 *Demonstration of a difference*

Various aspects need to be considered in the empirical demonstration that certain groups differ with regard to a certain characteristic. It should be noted that the “demonstration” (of a difference) should not be understood as “proof” in a mathematical sense. With the help of empirical study data, statements can only be made by allowing for certain probabilities of error. By applying statistical methods, these probabilities of error can, however, be specifically controlled and minimised in order to “statistically demonstrate” a hypothesis. A typical method for such a statistical demonstration in medical research is the application of significance tests. This level of argumentation should be distinguished from the evaluation of the clinical relevance of a difference. In practice, the combination of both arguments provides an adequate description of a difference based on empirical data.

When applying a significance test to demonstrate a difference, the research question should be specified a priori, and the outcome variable, the effect measure, and the statistical hypothesis formulation should also be specified on the basis of this question. It is necessary to calculate the sample size required before the start of the study, so that the study is large enough for a difference to be detected. In simple situations, in addition to the above information, a statement on the

clinically relevant difference should be provided, as well as an estimate of the variability of the outcome measure. For more complex designs or research questions, further details are required (e.g. correlation structure, recruitment scheme, estimate of drop-out numbers, etc.) [46,90].

Finally, the reporting of results should include the following details: the significance level for a statement; a confidence interval for the effect measure chosen (calculated with appropriate methods); descriptive information on further effect measures to explain different aspects of the results; as well as a discussion on the clinical relevance of the results, which should be based on the evaluation of patient-relevant outcomes.

6.4.5 Demonstration of equivalence

One of the most common serious errors in the interpretation of medical data is to rate the non-significant result of a traditional significance test as evidence that the null hypothesis is true [9]. To demonstrate “equivalence”, methods to test equivalence hypotheses need to be applied [216]. In this context, it is important to understand that demonstrating exact “equivalence” (e.g. that the difference in mean values between two groups is exactly zero) is not possible by means of statistical methods. In practice, it is not demonstration of exact equivalence that is required, but rather demonstration of a difference between two groups that is “at most irrelevant”. To achieve this objective, it must, of course, first be defined what an irrelevant difference is, i.e. an equivalence range must be specified.

To draw meaningful conclusions on equivalence, the research question and the resulting outcome variable, effect measure, and statistical hypothesis formulation need to be specified a priori (similar to the demonstration of a difference). In addition, in equivalence studies the equivalence range must be clearly defined. This range can be two-sided, resulting in an equivalence interval, or one-sided in terms of an “at most irrelevant difference” or “at most irrelevant inferiority”. The latter is referred to as a “non-inferiority hypothesis” [82,200,303].

As in superiority studies, it is also necessary to calculate the required sample size in equivalence studies before the start of the study. The appropriate method depends on the precise hypothesis, as well as on the analytical method chosen [302].

Specifically developed methods should be applied to analyse data from equivalence studies. The “confidence interval inclusion method” is a frequently used technique. If the confidence interval calculated lies completely within the equivalence range defined a priori, then this will be classified as the demonstration of equivalence. To maintain the level of $\alpha=0.05$, it is sufficient to calculate a 90% confidence interval [216].

Compared with superiority studies, equivalence studies show specific methodological problems. On the one hand, it is often difficult to provide meaningful definitions of equivalence ranges [243]; on the other hand, the usual study design criteria, such as randomisation and blinding, no longer sufficiently protect from bias [332]. Even without knowledge of the treatment group, it is possible, for example, to shift the treatment differences to zero and hence in the direction of the desired alternative hypothesis. Moreover, the ITT principle should be applied carefully, as its inappropriate use may falsely indicate equivalence [216]. For this reason, particular caution is necessary in the evaluation of equivalence studies.

6.4.6 Adjustment principles and multi-factorial methods

Primarily in non-randomised studies, multi-factorial methods that enable confounder effects to be compensated play a key role [225]. Studies with several treatment groups are a further important field of application for these methods [259]. In the medical literature, the reporting of results obtained with multi-factorial methods is unfortunately often insufficient [38,270]. To be able to assess the quality of such an analysis, the description of essential aspects of the statistical model formation is necessary [176,309], as well as information on the quality of the model chosen (goodness of fit) [192]. The most relevant information for this purpose is usually:

- A clear description and a priori specification of the outcome variables and all potential explanatory variables,
- Information on the measurement scale and on the coding of all variables,
- Information on the selection of variables and on any interactions,
- Information on how the assumptions of the model were verified,
- Information on the goodness of fit of the model,
- Inclusion of a table with the most relevant results (parameter estimate, standard error, confidence interval) for all explanatory variables.

Depending on the research question posed, this information is of varying relevance. If it concerns a good prediction of the outcome variable within the framework of a prognosis model, a high-quality model is more important than in a comparison of groups, where an adjustment for important confounders must be made.

Inadequate reporting of the results obtained with multi-factorial methods is especially critical if the (inadequately described) statistical modelling leads to a shift of effects to the “desired” range,

which is not recognisable with mono-factorial methods. Detailed comments on the requirements for the use of multi-factorial methods can be found in various reviews and guidelines [25,39,225].

The Institute uses modern methods in its own regression analysis calculations [175]. In this context, results of multi-factorial models that were obtained from a selection process of variables should be interpreted with great caution. When choosing a model, if such selection processes cannot be avoided, a type of backward elimination will be used, as this procedure is preferable to the procedure of forward selection [175]. A well-informed and careful pre-selection of the candidate predictor variable is essential in this regard [87]. If required, modern methods such as the lasso technique will also be applied [356]. For the modelling of continuous covariates, the Institute will, if necessary, draw upon flexible modelling approaches (e.g. regression using fractional polynomials [310,320]) to enable the appropriate description of non-monotonous associations.

6.4.7 Meta-analyses

Terms used in the literature, such as *literature review*, *systematic review*, *meta-analysis*, *pooled analysis*, or *research synthesis*, are often defined differently and not clearly distinguished [110].

The Institute uses the following terms and definitions:

- A *non-systematic review* is the assessment and reporting of study results on a defined topic, without a sufficiently systematic and reproducible method for identifying relevant research results. A quantitative summary of data from several studies is referred to as a *pooled analysis*. Due to the lack of a systematic approach and the inherent subjective component, reviews and analyses not based on a systematic literature search are extremely prone to bias.
- A *systematic review* is based on a comprehensive, systematic approach and assessment of studies, which is applied to minimise potential sources of bias. A systematic review may, but does not necessarily have to, contain a quantitative summary of study results.
- A *meta-analysis* is a statistical summary of the results of several studies within the framework of a systematic review. In most cases this analysis is based on aggregated study data from publications. An overall effect is calculated from the effect sizes measured in individual studies, taking sample sizes and variances into account.
- More efficient analysis procedures are possible if IPD are available from the studies considered. A *meta-analysis including IPD* is the analysis of data on the patient level within the framework of a general statistical model of fixed or random effects, in which the study is considered as an effect and not as an observation unit.

- The Institute sees a *prospective meta-analysis* as a statistical summary (planned a priori) of the results of several prospective studies that were jointly planned. However, if other studies are available on the particular research question, these must also be considered in the analysis in order to preserve the character of a systematic review.

The usual presentation of the results of a meta-analysis is made by means of forest plots in which the effect estimates of individual studies and the overall effect (including confidence intervals) are presented graphically [247]. On the one hand, fixed effects models are applied, which provide weighted mean values of the effect sizes (for example, weighting by inverting the variance). On the other hand, random effects models are frequently chosen in which an estimate of the variance between individual studies (heterogeneity) is considered.

The question as to which model should be applied in which situation has long been a matter of controversy [114,330,371]. If information is available that the effects of the individual studies are homogeneous, a meta-analysis assuming fixed effects 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 is useful [331]. Moreover, it should be noted that the confidence intervals calculated from a fixed effects 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 [59]. The Institute therefore primarily uses random effects models and only applies fixed effects models in well-founded exceptional cases. In this context, if the data situation is homogeneous, it should be noted that meta-analytical results from models with random and fixed effects at best show marginal differences. As described in the following text, the Institute will only perform a meta-analytical summary of strongly heterogeneous study results if the reasons for this heterogeneity are plausible and still justify such a summary.

Before a meta-analysis is conducted, it must first be considered whether the pooling of the studies investigated is in fact meaningful, as the studies must be comparable with regard to the research question posed. In addition, even in the case of comparability, the studies to be summarised will often show heterogeneous effects [187]. In this situation it is necessary to assess the heterogeneity of study results [148]. The existence of heterogeneity can be statistically tested; however, these tests usually show very low power. Consequently, it is recommended that a significance level between 0.1 and 0.2 is chosen for these tests [208]. However, it is more important to quantify the impact of heterogeneity. For this purpose, specific new statistical methods are available, such as the I^2 measure [186]. Studies exist for this measure that allow a rough classification of heterogeneity (for example, low/medium/high: for I^2 values in the range 25/ 50/ 75%) [187]. If the heterogeneity of the studies is too great, the statistical pooling of the study results may not prove meaningful [86].

In this context, the location of the effect also plays a role. If the individual studies show a clear effect in the same direction, heterogeneous results can also be pooled appropriately by means of a random effects model. In other situations the Institute will not conduct a meta-analysis. However, not only statistical measures, but also issues of content must be considered when making such a decision, which must be presented in a comprehensible way. In this context, the choice of the effect measure also plays a role. The choice of a certain measure may lead to great study heterogeneity, yet another measure may not. For binary data, relative effect measures are frequently more stable than absolute ones, as they do not depend so heavily on the baseline risk [142]. In such cases, the data analysis should be conducted with a relative effect measure, but for the descriptive presentation of data, absolute measures for the specific baseline risks should be inferred from them.

In the case of great heterogeneity of the studies, it is necessary to investigate potential causes. Factors that could explain the heterogeneity of the effect size may possibly be detected by means of meta-regression [353,367]. In a meta-regression, the statistical association between the effect sizes of individual studies and the study characteristics is investigated, so that study characteristics can possibly be identified that explain the different effect sizes; i.e. the heterogeneity. However, when interpreting results, it is important that the limitations of such analyses are taken into account. Even if a meta-regression is based on randomised studies, only evidence of an observed association can be inferred from this analysis, not a causal relationship [353]. Meta-regressions that attempt to show an association between the different effect sizes and the average patient characteristics in individual studies are especially difficult to interpret. These analyses are subject to the same limitations as the results of ecological studies in epidemiology [157]. Due to the high risk of bias, which in analyses based on aggregate data cannot be balanced by adjustment, definite conclusions are only possible on the basis of IPD [338,353] (see also Section 6.3.3).

6.4.8 Handling of unpublished or partially published data

In the quality assessment of publications, the problem frequently arises in practice that essential data or information is partially or entirely missing. This mainly concerns “grey literature” and abstracts, but also full-text publications. Moreover, it is possible that studies have not (yet) been published at the time of the Institute’s technology assessment.

It is the Institute’s aim to conduct an assessment on the basis of a data set that is as complete as possible. If relevant information is missing, IQWiG therefore tries to complete the missing data, among other things by contacting the authors of publications or the study sponsors. However, depending on the type of product prepared, requests for unpublished information may be restricted due to time limits.

A common problem is that important data required for the conduct of a meta-analysis (e.g. variances of effect estimates) are lacking. However, in many cases, missing data can be calculated or at least estimated from the data available [94,194,289]. If possible, IQWiG will apply such procedures.

If data are only partly available or if estimated values are used, the robustness of results will be analysed and discussed, if appropriate with the support of sensitivity analyses (e.g. by presenting best-case and worst-case scenarios). However, a worst-case scenario can only be used here as proof of the robustness of a detected effect. It cannot simply be inferred from a worst-case scenario where an effect previously detected could not be confirmed that such an effect has not been demonstrated. In cases where relevant information is largely or completely lacking, it may occur that a publication cannot be assessed. In such cases, it will merely be noted that further data exist on a particular topic, but are not available for assessment.

6.4.9 Description of types of bias

Bias is the systematic deviation of the effect estimate (inferred from study data) from the true effect. Bias may be produced by a range of possible causes [312]. The following text describes only the most important types; a detailed overview of various types of bias in different situations is presented by Feinstein [127].

Selection bias is caused by a violation of the random principles for sampling procedures. Particularly in the comparison of two groups, selection bias can lead to systematic differences between groups. If this leads to an unequal distribution of important confounders between groups, the results of a comparison are usually no longer interpretable. When comparing groups, randomisation is the best method to avoid selection bias, as the groups formed do not differ systematically with regard to known as well as unknown confounders. However, structural equality can only be ensured if the sample sizes are sufficiently large. In small studies, despite randomisation, relevant differences between groups can occur at random. When comparing groups with structural inequality, the effect of known confounders can be taken into account by applying multi-factorial methods. However, the problem remains of a systematic difference between the study groups due to unknown or insufficiently investigated confounders.

Performance bias is bias caused by different types of care provided (apart from the intervention to be investigated). In addition to the comparability of study groups with regard to potential prognostic factors, equality of care and the equality of observation of participants play an important role.

A violation of the equality of observation can lead to *detection bias*.

Blinding is an effective protection against both performance and detection bias, which are summarised as *information bias* in epidemiology.

Protocol violations and study withdrawals can cause a systematic bias of study results, called *attrition bias*. To avoid attrition bias, the ITT principle can be applied, where all randomised study participants are analysed within the group to which they were randomly assigned, independent of protocol violations.

In diagnostic studies, the assessment of the diagnostic test should be conducted in an appropriate spectrum of patients. If the sample assessed differs systematically from the patient population in which the test is to be applied, this may lead to *spectrum bias*. To avoid this type of bias, the diagnostic test should be assessed in a representative patient population.

When assessing screening programmes, it needs to be considered that earlier diagnosis of a disease often results only in an apparent increase in survival times, due to non-comparable starting points (*lead time bias*). Increased survival times may also appear to be indicated if a screening test preferably detects mild or slowly progressing early stages of a disease (*length bias*). The conduct of a randomised trial to assess the effectiveness of a screening test can protect against these bias mechanisms.

A common problem arising from the estimation of effects is bias caused by measurement errors and misclassifications in the study data collected [66,69]. In practice, measurement errors can hardly be avoided, and it is well known that non-differential measurement errors may also lead to bias in the estimation of an effect. In the case of a simple linear regression model with a classical measurement error in the explanatory variable, *dilution bias* occurs, i.e. a bias of the estimate towards the zero effect. However, in other models and more complex situations, there may be bias in all directions. Depending on the research question posed, the size of potential measurement errors should be discussed and, if necessary, methods to adjust bias of measurement errors should be applied.

Missing values present a similar problem. Missing values not due to a random mechanism can also cause bias in a result [251]. The possible causes and effects of missing values should therefore be discussed on a case-by-case basis and, if necessary, statistical methods should be applied to account or compensate for bias.

Publication bias plays an important role in systematic reviews [35]. Because significant results are more likely to be published than non-significant ones, a systematic bias of the common effect estimate occurs when published results are summarised. Graphical methods such as the funnel plot

[111] and/or statistical methods such as meta-regression are techniques for identifying and considering publication bias [254,292,348].

6.5 Qualitative methods

6.5.1 Qualitative studies

Various information sources can support the preparation of systematic reviews [99,248,352], including research findings from qualitative studies [74,174,248,279,352]. Qualitative studies seem to be establishing themselves in systematic reviews of the benefit assessment of medical services [98,99,279].

Qualitative research methods are applied to explore and understand subjective experiences, individual actions, and the social world [98,174,256,272]. They can enable access to opinions and experiences of patients, relatives, and medical staff with respect to a certain disease or intervention.

Qualitative research can provide information on the acceptability and suitability of interventions in clinical practice [73,98]. The results of qualitative research can be helpful in the interpretation of a systematic review [73,352].

Quantitative research works primarily with numbers of different dimensions and is characterised by strong standardisation, although personal and social experiences may also be taken into account (e.g. QoL studies). Conversely, in qualitative research the emphasis is on subjective data, for example, the conduct of focus groups with participants of RCTs. Qualitative data can also be collected by means of interviews, observations, and written documents, such as diaries. This analytical approach mainly aims to identify and analyse overlapping topics and concepts of the data collected.

Among other things, qualitative methods can be used as an independent research method, in the preparation of or as a supplement to quantitative studies, within the framework of the triangulation or mixed-method approach, or after the conduct of quantitative studies, in order to explain processes or results. Qualitative research is seen as a method to promote the connection between evidence and practice [100].

Qualitative methods can also be used within the framework of primary studies or systematic reviews to identify patient-relevant outcomes [73,98,100,272,279].

Despite the increasing relevance of qualitative methods, no generally accepted approach exists for the synthesis of qualitative studies and the combination of qualitative and quantitative data [73,100,101].

The Institute can use qualitative research findings to identify patient-relevant outcomes and to present background information on patients' experiences and on the patient relevance of the intervention to be assessed. The Institute can also use these findings in the discussion and interpretation of results of a systematic review.

In the development of health information the Institute uses available qualitative research findings to identify information needs, as well as to investigate experiences with a certain disease or an intervention.

A systematic search for qualitative studies is conducted after the specification of the research question. The PubMed, CINAHL, and PsycInfo databases are usually searched. Relevant publications are then selected by means of prespecified inclusion and exclusion criteria, and the study quality is assessed by means of criteria defined beforehand.

The main task of the Institute in the assessment of qualitative studies is to determine whether the study design, study quality, and reliability are appropriate for the research question investigated. There is a weaker general consensus with regard to the validity of criteria for the conduct, assessment, and synthesis of qualitative studies when compared with other research areas [98,101,174,272].

In the development of health information, the results of the studies considered are extracted, organised by topic, and summarised in a descriptive manner. The Institute may also take this approach in the production of reports.

The Institute monitors further methodological developments regarding the combination of qualitative and quantitative methods and the respective research findings, both in primary studies and in systematic reviews of the benefit assessment of medical services. It subsequently adapts its own methods where appropriate.

6.5.2 Consultation techniques

The processing of research questions and tasks commissioned to the Institute often requires the consultation of patients, patient representatives, and national and international experts. In this case the Institute uses various consultation techniques.

In the production of reports, the Institute uses these techniques to identify patient-relevant outcomes and to involve national and international experts, and also uses them in the Institute's formal consultation procedure.

In the development of health information, consultation techniques serve to involve patients and patient representatives in the identification of information needs, the evaluation of health information, and during consultation.

The Institute uses the following consultation techniques:

- Key informant interviews [364], e.g. interviews with patient representatives to identify patient-relevant outcomes.
- Group meetings and consultations [275,277,278], e.g. within the framework of scientific debates on IQWiG products.
- Group interviews and focus groups [98,277,363], e.g. with patients with respect to the evaluation of health information.
- Surveys and polling (including online polling and feedback mechanisms), e.g. to identify information needs of readers of www.informedhealthonline.org.

If a deeper understanding of experiences and opinions is necessary, then the Institute uses the scientific findings obtained from qualitative research.

The use of consultation techniques and the involvement of experts are associated with an additional use of resources. However, the involvement of patients in research processes enables the consideration of patient issues and needs as well as the orientation of research towards these issues and needs [284].

Reference list

1. Gesetz zur Modernisierung der gesetzlichen Krankenversicherung (GKV-Modernisierungsgesetz - GMG) [*Statutory Health Insurance Modernisation Act*]. 19.11.2003. Bundesgesetzblatt 2003; Teil 1(55): 2190-2258.
2. Gesetz zur Stärkung des Wettbewerbs in der gesetzlichen Krankenversicherung (GKV-Wettbewerbsstärkungsgesetz - GKV-WSG) [*Statutory Health Insurance Act to Promote Competition*]. 26.03.2007. Bundesgesetzblatt 2007; Teild 1(11): 378-473.
3. AGREE Collaboration. Appraisal of guidelines for research & evaluation, AGREE Instrument [Online]. 2001 [Accessed on 20.06.2007]. URL: <http://www.agreecollaboration.org/pdf/agreeinstrumentfinal.pdf>.
4. Agresti A. Modelling ordered categorical data: recent advances and future challenges. *Stat Med* 1999; 18(17-18): 2191-2207.
5. Agresti A. *Categorical data analysis*. Hoboken: Wiley; 2002.
6. Agresti A. Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Stat Methods Med Res* 2003; 12(1): 3-21.
7. Al-Marzouki S, Evans S, Marshall T, Roberts I. Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ* 2005; 331(7511): 267-270.
8. Altman DG. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG (Ed). *Systematic reviews in health care: Meta-analysis in context*. London: BMJ Publishing Group; 2001: 228-247.
9. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995; 311(7003): 485.
10. Altman DG, Machin D, Bryant TM, Gardner MJ. *Statistics with confidence: confidence intervals and statistical guidelines*. London: BMJ Publishing Group; 2000.
11. Altman DG, Schulz K, Moher D. Turning a blind eye: testing the success of blinding and the CONSORT statement. *BMJ* 2004; 328(7448): 1135.
12. Altman DG, Schulz K, Moher D, Egger M, Davidoff F, Elbourne DR et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134(8): 663-694.
13. American Society of Clinical Oncology. Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. *J Clin Oncol* 1996; 14(2): 671-679.
14. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *Treatments for myocardial infarction*. *JAMA* 1992; 268(2): 240-248.
15. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ärztliche Zentralstelle Qualitätssicherung (ÄZQ). Das Leitlinien-Manual von AWMF und ÄZQ [*Association of the Scientific Medical Professional Societies (AWMF). Agency for Quality Assurance in Medicine (ÄZQ). Guideline manual of the AWMF and ÄZQ*]. *Z Ärztl Fortbild Qualitätssich* 2001; 95(Suppl 1): 1-84.
16. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ärztliche Zentralstelle Qualitätssicherung (ÄZQ). Deutsches Instrument zur methodischen Leitlinienbewertung (DELBI) [*Association of the Scientific Medical Professional Societies (AWMF). Agency for Quality Assurance in Medicine (ÄZQ). German Instrument for Methodological Guideline appraisal*] [Online]. 2005 [Accessed on 20.06.2007]. URL: <http://www.versorgungsleitlinien.de/methodik/delbi/pdf/delbi05.pdf>.
17. Ärztliche Zentralstelle Qualitätssicherung (ÄZQ). Checkliste "Methodische Qualität von Leitlinien". [*Agency for Quality Assurance in Medicine. Checklist "Methodological quality of guidelines"*]. *Dtsch Arztebl* 1998; 95(41): A2576-A2578.

18. Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „COPD“: Leitlinien-Clearingverfahren von Bundesärztekammer und Kassenärztlicher Bundesvereinigung in Kooperation mit Deutscher Krankenhausgesellschaft, Spitzenverbänden der Krankenkassen und Gesetzlicher Rentenversicherung. [Agency for Quality in Medicine. Guideline clearing report “COPD”: Guideline clearing procedure of the German Medical Association and the Federal Association of Statutory Health Insurance Physicians in cooperation with the German Hospital Federation and the National Confederations of Regional Associations of the Statutory Health Insurance Funds and the Statutory Pension Fund]. Köln: ÄZQ; 2003.
19. Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „Depression“: Leitlinien-Clearingverfahren von Bundesärztekammer und Kassenärztlicher Bundesvereinigung in Kooperation mit Deutscher Krankenhausgesellschaft, Spitzenverbänden der Krankenkassen und Gesetzlicher Rentenversicherung. [Agency for Quality in Medicine. Guideline clearing report “Depression”: Guideline clearing procedure of the German Medical Association and the Federal Association of Statutory Health Insurance Physicians in cooperation with the German Hospital Federation and the National Confederations of Regional Associations of the Statutory Health Insurance Funds and the Statutory Pension Fund]. Köln: ÄZQ; 2003.
20. Ärztliches Zentrum für Qualität in der Medizin. Das Deutsche Leitlinien-Clearingverfahren 1999-2005: Hintergrund, Zielsetzung, Ergebnisse. Abschlussbericht. [Agency for Quality in Medicine. The German guideline clearing procedure 1999-2005. Background, aims, results. Final report]. Berlin: ÄZQ; 2006.
21. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; 355(9209): 1064-1069.
22. Atkins D, Best D, Briss PA, Eccles MP, Falck-Ytter Y, Flottorp S et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328(7454): 1490.
23. Atkins D, Briss PA, Eccles MP, Flottorp S, Guyatt GH, Harbour RT et al. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. *BMC Health Serv Res* 2005; 5(1): 25.
24. Atkins D, Eccles MP, Flottorp S, Guyatt GH, Henry D, Hill S et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 2004; 4(1): 38.
25. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001; 54(10): 979-985.
26. Baker SG. Surrogate endpoints: wishful thinking or reality? *J Natl Cancer Inst* 2006; 98(8): 502-503.
27. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Alloc* 2006; 4: 14.
28. Baltussen R, Stolk E, Chisholm D, Aikins M. Towards a multi-criteria approach for priority setting: an application to Ghana. *Health Econ* 2006; 15(7): 689-696.
29. Bandolier. On care pathways. *Bandolier Forum* 2003: 1-12.
30. Banta D. The development of health technology assessment. *Health Policy* 2003; 63(2): 121-132.
31. Barron BA, Bukantz SC. The evaluation of new drugs. Current Food and Drug Administration regulations and statistical aspects of clinical trials. *Arch Intern Med* 1967; 119(6): 547-556.
32. Bastian H. Non-peer review: consumer involvement in research review. In: Godlee F, Jefferson T (Ed). *Peer review in health sciences*. London: BMJ Publishing Group; 2003: 248-262.
33. Bastian H, Kaiser T, Matschewsky S. Förderung allgemeiner Gesundheits- und Wissenschaftskenntnisse mittels Bürger- und Patienteninformationen: die Rolle des IQWiG [Promotion of general health and scientific literacy via consumer and patient information: The role of IQWiG]. *Z Ärztl Fortbild Qual Gesundh wes* 2005; 99(6): 379-385.

34. Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the internet. *Med Inform Internet Med* 2006; 31(1): 45-52.
35. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *J R Stat Soc [Ser A]* 1988; 151(3): 419-463.
36. Bender R. Interpretation von Effizienzmaßen der Vierfeldertafel für Diagnostik und Behandlung [*Interpretation of efficiency measures of the 2x2 table for diagnosis and treatment*]. *Med Klin* 2001; 96(2): 116-121.
37. Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008; 61: (in press).
38. Bender R, Grouven U. Logistic regression models used in medical research are poorly presented. *BMJ* 1996; 313(7057): 628.
39. Bender R, Grouven U. Ordinal logistic regression in medical research. *J R Coll Physicians Lond* 1997; 31(5): 546-551.
40. Bender R, Lange S. Adjusting for multiple testing - when and how? *J Clin Epidemiol* 2001; 54(4): 343-349.
41. Bent S, Padula A, Avins AL. Brief communication: better ways to question patients about adverse medical events. A randomized, controlled trial. *Ann Intern Med* 2006; 144(4): 257-261.
42. Bessell TL, McDonald S, Silagy CA, Anderson JN, Hiller JE, Sansom LN. Do internet interventions for consumers cause more harm than good? A systematic review. *Health Expect* 2002; 5(1): 28-37.
43. Biglan A, Ary D, Wagenaar AC. The value of interrupted time-series experiments for community intervention research. *Prev Sci* 2000; 1(1): 31-49.
44. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; 69(3): 89-95.
45. Bland JM, Altman DG. One and two sided tests of significance. *BMJ* 1994; 309(6949): 248.
46. Bock J, Toutenburg H. Sample size determination in clinical research. In: Rao CR, Chakraborty R (Ed). *Statistical Methods in Biological and Medical Sciences*. Amsterdam: Elsevier; 1991: 515-538.
47. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ* 2006; 333(7563): 346-349.
48. Bonhoeffer J, Zumbunn B, Heining U. Reporting of vaccine safety data in publications: systematic review. *Pharmacoepidemiol Drug Saf* 2005; 14(2): 101-106.
49. Bossuyt PM, Irwig LM, Craig J, Glasziou PP. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006; 332(7549): 1089-1092.
50. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD Initiative. *Ann Intern Med* 2003; 138(1): 40-44.
51. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138(1): W1-W12.
52. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* 2008; 148(4): 295-309.
53. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacologic treatments. *Ann Intern Med* 2008; 148(4): W60-W66.
54. Boynton J, Glanville J, McDaid D, Lefebvre C. Identifying systematic reviews in MEDLINE: Developing an objective approach to search strategy design. *J Inf Sci* 1998; 24(3): 137-155.

-
55. Breslow NE, Day NE. Statistical methods in cancer research, Vol.I - The analysis of case-control studies. Lyon: International Agency for Research on Cancer; 1980.
 56. Breslow NE, Day NE. Statistical methods in cancer research, Vol. II - The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer; 1987.
 57. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine [Online]. 2002 [Accessed on 20.06.2007]. URL: <http://www.stanford.edu/class/cs240/readings/google.pdf>.
 58. British Cardiac Society, British Hyperlipidaemia Association, British Hypertension Society. Joint British recommendations on prevention of coronary heart disease in clinical practice. Heart 1998; 80(Suppl 2): S1-S29.
 59. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. Stat Med 2001; 20(6): 825-840.
 60. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses. Power and sample size for the interaction test. J Clin Epidemiol 2004; 57(3): 229-236.
 61. Bundesministerium des Innern. Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz [*Federal Ministry of the Interior. Regulation for the Provision of Barrier-free Information Technology in Accordance with the Equal Rights for People with Disabilities Law*]. Berlin: 2002.
 62. Bundesverfassungsgericht. Leitsatz zum Beschluss des Ersten Senats. [*Federal Constitutional Court. Maxim on the Resolution of the First Senate*]. 06.12.2005: 1 BvR 347/98.
 63. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada. Int J Technol Assess Health Care 2003; 19(1): 148-157.
 64. Busse R, Drösler S, Glaeske G, Greiner W, Schäfer T, Schrappe M. Wissenschaftliches Gutachten für die Auswahl von 50 bis 80 Krankheiten zur Berücksichtigung im morbiditätsorientierten Risikostrukturausgleich. Wissenschaftlicher Beirat zur Weiterentwicklung des Risikostrukturausgleichs [*Scientific expertise on the selection of 50 to 80 diseases for consideration in the morbidity-orientated risk adjustment scheme. Scientific Advisory Board for the Further Development of the Risk Adjustment Scheme*]; 2007.
 65. Campbell MK, Elbourne DR, Altman DG, CONSORT Group. CONSORT statement: extension to cluster randomised trials. BMJ 2004; 328(7441): 702-708.
 66. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. London: Chapman & Hall; 2006.
 67. Centre for Evidence-based Medicine. Levels of evidence [Online]. 2001 [Accessed on 07.05.2007]. URL: http://www.cebm.net/levels_of_evidence.asp.
 68. Charles C, Gafni A, Whelan T, O'Brien MA. Treatment decision aids: conceptual issues and future directions. Health Expect 2005; 8(2): 114-125.
 69. Cheng CL, Van Ness JW. Statistical regression with measurement error. London: Arnold; 1999.
 70. Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. Int J Qual Health Care 1999; 11(1): 21-28.
 71. Cochrane Collaboration. The Cochrane Collaboration - Principles [Online]. 2006 [Accessed on 01.05.2006]. URL: <http://www.cochrane.org/docs/tenprinciples.htm>.
 72. Cochrane Effective Practice and Organisation of Care Review Group. The data collection checklist. Ottawa: EPOC; 2002.
 73. Cochrane Qualitative Research Methods Group. Methodological issues arising from the inclusion of qualitative evidence in systematic reviews [Online]. 2007 [Accessed on 30.04.2007]. URL: <http://www.lancs.ac.uk/fass/ihr/research/public/cochrane.htm>.

74. Cochrane Qualitative Research Methods Group. What is qualitative research? What is its role in evidence review? [Online]. 2007 [Accessed on 30.04.2007]. URL: <http://www.joannabriggs.edu.au/cqrmg/role.html>.
75. Commission of the European Communities. eEurope 2002: quality criteria for health related websites. Brussels: 2002.
76. Committee for Proprietary Medicinal Products (CPMP). Points to consider on application with 1. Meta-analyses; 2. One pivotal study. London: EMEA; 2001.
77. Coulter A. Evidence based patient information is important, so there needs to be a national strategy to ensure it. *BMJ* 1998; 317(7153): 225-226.
78. Coulter A, Ellins J, Swain D, Clarke A, Heron P, Rasul F et al. Assessing the quality of information to support people in making decisions about their health and healthcare. Oxford: Picker Institute Europe; 2006.
79. Council of Europe: Committee of Experts on Management of Safety and Quality in Health Care (SP-SQS) EGoSMP. Glossary of terms related to patient and medication safety. Strasbourg: Council of Europe; 2005.
80. Cui L, Hung HM, Wang SJ, Tsong Y. Issues related to subgroup analysis in clinical trials. *J Biopharm Stat* 2002; 12(3): 347-358.
81. Cullen R. Health information on the internet: a study of providers, quality, and users. Westport: Praeger; 2006.
82. D'Agostino RB, Sr., Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues. The encounters of academic consultants in statistics. *Stat Med* 2003; 22(2): 169-186.
83. Dans AL, Dans LF, Guyatt GH, Richardson S, Evidence-Based Medicine Working Group. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998; 279(7): 545-549.
84. David M, Borde T, Kantenich H. Knowledge among German and Turkish women about specifically female bodily functions, contraception, preventative medical examinations and menopause. *Ethn Health* 2000; 5(2): 101-112.
85. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; 323(7305): 157-162.
86. Deeks JJ, Higgins JPT, Altman DG, Eds. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, Eds. *Cochrane handbook for systematic reviews of interventions*, Version 5.0.0 (updated February 2008) [Online]. 2008 [Accessed on 13.05.2008]. URL: <http://www.cochrane-handbook.org>.
87. Derksen S, Keselman HJ. Backward, forward, and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *Br J Math Statist Psych* 1992; 45: 265-282.
88. Derry S, Loke YK, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. *BMC Med Res Methodol* 2001; 1: 7.
89. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004; 94(3): 361-366.
90. Desu MM, Raghavarao D. Sample size methodology. Boston: Academic Press; 1990.
91. Deutsche Gesellschaft für Epidemiologie (DGEpi). Leitlinien und Empfehlungen zur Sicherung von guter epidemiologischer Praxis (GEP). [*German Epidemiological Society. Guidelines and recommendations to ensure good epidemiological practice*] Deutsche Gesellschaft für Epidemiologie (DGEpi); 2008.
92. Deutsches Institut für Normung. Klinische Prüfung von Medizinprodukten an Menschen. Teil 2: klinische Prüfpläne [*German Institute for Standardisation. Clinical testing of medicinal products in humans. Part 2: Clinical protocols*]. (ISO 14155-2:2003). Berlin: Beuth; 2003.

-
93. Deutsches Netzwerk Evidenzbasierte Medizin e.V. Definitionen [German Network for Evidence-based Medicine. Definitions] [Online]. 2007 [Accessed on 25.09.2007]. URL: <http://www.ebm-netzwerk.de/grundlagen/grundlagen/definitionen>.
 94. Di Pietrantonj C. Four-fold table cell frequencies imputation in meta analysis. *Stat Med* 2006; 25(13): 2299-2322.
 95. Dierks ML, Seidel G, Schwartz FW, Horch K. Bürger- und Patientenorientierung im Gesundheitswesen: Gesundheitsberichterstattung des Bundes - Heft 32 [*Consumer and patient orientation in health care. Federal health care reports - Leaflet 32*]. Berlin: Robert Koch Institut; 2006.
 96. Dimensional. Papaya CMS - pflegen Sie Ihre Website ohne Programmierkenntnisse [*Papaya CMS - Maintain your website without programming knowledge*] [Online]. 2007 [Accessed on 27.05.2007]. URL: http://www.dimensional.de/index_ndim85.php.
 97. Dixon-Woods M. Writing wrongs? An analysis of published discourses about the use of patient information leaflets. *Soc Sci Med* 2001; 52(9): 1417-1432.
 98. Dixon-Woods M, Agarwal S, Young B, Jones D, Sutton A. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.
 99. Dixon-Woods M, Fitzpatrick R. Qualitative research in systematic reviews. Has established a place for itself. *BMJ* 2001; 323(7316): 765-766.
 100. Dixon-Woods M, Fitzpatrick R, Roberts K. Including qualitative research in systematic reviews: opportunities and problems. *J Eval Clin Pract* 2001; 7(2): 125-133.
 101. Dixon-Woods M, Sutton A, Shaw R, Miller T, Smith J, Young B et al. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *J Health Serv Res Policy* 2007; 12(1): 42-47.
 102. Donner A, Klar J. Design and analysis of cluster randomization trials in health research. London: Arnold; 2000.
 103. Draborg E, Gyrd-Hansen D, Poulsen PB, Horder M. International comparison of the definition and the practical application of health technology assessment. *Int J Technol Assess Health Care* 2005; 21(1): 89-95.
 104. Drazen JM. COX-2 inhibitors - a lesson in unexpected problems. *N Engl J Med* 2005; 352(11): 1131-1132.
 105. Eccles MP, Rousseau N, Freemantle N. Updating evidence-based clinical guidelines. *J Health Serv Res Policy* 2002; 7(2): 98-103.
 106. Edwards AG, Bastian H. Risk communication: making evidence part of patient choices? In: Edwards AG, Elwyn GJ (Ed). *Evidence-based patient choice: Inevitable or impossible?* Oxford: Oxford University Press; 2001: 144-160.
 107. Edwards AG, Elwyn GJ, Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002; 324(7341): 827-830.
 108. Edwards AG, Evans R, Dundon J, Haigh S, Hood K, Elwyn GJ. Personalised risk communication for informed decision making about taking screening tests. *Cochrane Database Syst Rev* 2006; 4: CD001865.
 109. Edwards JE, McQuay HJ, Moore RA, Collins SL. Reporting of adverse effects in clinical trials should be improved: lessons from acute postoperative pain. *J Pain Symptom Manage* 1999; 18(6): 427-437.
 110. Egger M, Davey Smith G, Altman DG. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group; 2001.
 111. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315(7109): 629-634.

112. Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003; 7(1): 1-76.
113. Elwyn GJ, O'Connor A, Stacey D, Volk R, Edwards AG, Coulter A et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ* 2006; 333(7565): 417.
114. Engels EA, Schmid C, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med* 2000; 19(13): 1707-1728.
115. Entwistle VA, Sheldon TA, Sowden AJ, Watt I. Evidence-informed patient choice. Practical issues of involving patients in decisions about health care technologies. *Int J Technol Assess Health Care* 1998; 14(2): 212-225.
116. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. *JAMA* 2004; 291(19): 2359-2366.
117. Europäisches Parlament, Rat der Europäischen Union. Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden. [*European Parliament Council of the European Union. Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products*]. *Amtsbl Eur Gemeinschaften* 2000; L18: L18-1-L18/5; 22. Jan. 2000.
118. Europarat. Entwicklung einer Methodik für die Ausarbeitung von Leitlinien für optimale medizinische Praxis: Empfehlung Rec(2001)13 des Europarates und Erläuterndes Memorandum [*Council of Europe. Developing a methodology for drawing up guidelines on best medical practice. Rec(2001)13 of the Council of Europe and explanatory memorandum*]. Strasbourg: Council of Europe; 2002.
119. European Medicines Agency. Note for guidance on clinical investigation of medicinal products in the treatment of diabetes mellitus. London: EMEA; 2002.
120. European Medicines Agency. Guideline on clinical trials in small populations. London: EMEA; 2005.
121. European Medicines Agency. Guideline on the choice of the non-inferiority margin. London: EMEA; 2005.
122. 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. London: EMEA; 2005.
123. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992; 268(17): 2420-2425.
124. Eysenbach G. Recent advances: consumer health informatics. *BMJ* 2000; 320(7251): 1713-1716.
125. Eysenbach G, Kohler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ* 2002; 324(7337): 573-577.
126. Eysenbach G, Tuische J, Diepgen TL. Evaluation of the usefulness of Internet searches to identify unpublished clinical trials for systematic reviews. *Med Inform Internet Med* 2001; 26(3): 203-218.
127. Feinstein AR. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: Saunders; 1985.
128. Feinstein AR. Invidious comparisons and unmet clinical challenges. *Am J Med* 1992; 92(2): 117-120.
129. Feldman-Stewart D, Brennenstuhl S, Brundage MD. A purpose-based evaluation of information for patients: an approach to measuring effectiveness. *Patient Educ Couns* 2007; 65(3): 311-319.
130. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004; 328(7437): 432-434.

131. Fergusson D, Laupacis A, Salmi LR, McAlister FA, Huet C. What should be included in meta-analyses? An exploration of methodological issues using the ISPOt meta-analyses. *Int J Technol Assess Health Care* 2000; 16(4): 1109-1119.
132. Fervers B, Burgers JS, Haugh MC, Latreille J, Mlika-Cabanne N, Paquet L et al. Adaptation of clinical guidelines: literature review and proposition for a framework and procedure. *Int J Qual Health Care* 2006; 18(3): 167-176.
133. Field MJ, Lohr KN. *Clinical practice guidelines: directions for a new program*. Washington, D. C.: National Academy Press; 1990.
134. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 1996; 125(7): 605-613.
135. Fletcher RH, Fletcher SW. The effectiveness of journal peer review. In: Godlee F, Jefferson T (Ed). *Peer review in health sciences*. London: BMJ Publishing Group; 2003: 62-75.
136. Food and Drug Administration. *Guidance for industry: developing medical imaging drug and biological products. Part 2: clinical indications*. Rockville: FDA; 2004.
137. Forde OH. Is imposing risk awareness cultural imperialism? *Soc Sci Med* 1998; 47(9): 1155-1159.
138. Fransen GA, van Marrewijk CJ, Mujakovic S, Muris JW, Laheij RJ, Numans ME et al. Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. *BMC Med Res Methodol* 2007; 7: 16.
139. Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? *BMJ* 2001; 322(7292): 989-991.
140. Freemantle N, Blonde L, Bolinder B, Gerber RA, Hobbs FD, Martinez L et al. Real-world trials to answer real-world questions. *Pharmacoeconomics* 2005; 23(8): 747-754.
141. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2): 88-94.
142. Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol* 2002; 31(1): 72-76.
143. Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *BMJ* 2002; 324(7337): 569-573.
144. Garcia-Castillo D, Feters MD. Quality in medical translations: a review. *J Health Care Poor Underserved* 2007; 18(1): 74-84.
145. Gemeinsamer Bundesausschuss. *Verfahrensordnung des Gemeinsamen Bundesausschusses vom 20. September 2005 [Federal Joint Committee. Code of Procedure of the Federal Joint Committee of 20 September 2005]*. Siegburg: G-BA; 2006.
146. Gemoets D, Rosemblat G, Tse T, Logan R. Assessing readability of consumer health information: an exploratory study. *Medinfo* 2004; 11(Pt 2): 869-873.
147. Glasziou PP, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334(7589): 349-351.
148. Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; 21(11): 1503-1511.
149. Glasziou PP, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004; 328(7430): 39-41.
150. Glenton C, Nilsen ES, Carlsen B. Lay perceptions of evidence-based information - a qualitative evaluation of a website for back pain sufferers. *BMC Health Serv Res* 2006; 6: 34.
151. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003; 3: 28.

152. Golder S, McIntosh HM, Duffy S, Glanville J. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Info Libr J* 2006; 23(1): 3-12.
153. Gøtzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *Int J Technol Assess Health Care* 1996; 12(2): 238-246.
154. GRADE Working Group. The GRADE Working Group [Online]. 2007 [Accessed on 06.05.2007]. URL: <http://www.gradeworkinggroup.org/index.htm>.
155. Gray JAM. How to get better value healthcare. Oxford: Offox Press; 2007.
156. Greenhalgh T, Hurwitz B. Narrative based medicine: why study narrative? *BMJ* 1999; 318(7175): 48-50.
157. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol* 1989; 18(1): 269-274.
158. Gregoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol* 1995; 48(1): 159-163.
159. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. *J Med Internet Res* 2006; 8(2): e10.
160. Grimes DA, Schulz K. An overview of clinical research: the lay of the land. *Lancet* 2002; 359(9300): 57-61.
161. Grimes DA, Schulz K. Surrogate end points in clinical research: hazardous to your health. *Obstet Gynecol* 2005; 105(5 Pt 1): 1114-1118.
162. Grimshaw JM, Eccles MP, Russell I. Developing clinically valid practice guidelines. *J Eval Clin Pract* 1995; 1(1): 37-48.
163. Grol R, Dalhuijsen J, Thomas S, Veld C, Rutten G, Mokkink H. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ* 1998; 317(7162): 858-861.
164. Grouin JM, Coste M, Lewis J. Subgroup analyses in randomized clinical trials: statistical and regulatory issues. *J Biopharm Stat* 2005; 15(5): 869-882.
165. Guyatt GH. Evidence-based medicine. *ACP J Club* 1991; 14 (Suppl 2): A-16.
166. Guyatt GH, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B et al. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an american college of chest physicians task force. *Chest* 2006; 129(1): 174-181.
167. Guyatt GH, Jaeschke R, Roberts R. N-of-1 randomized clinical trials in pharmacoepidemiology. In: Strom BL (Ed). *Pharmacoepidemiology*. Chichester: Wiley; 2005: 665-680.
168. Guyatt GH, Rennie D. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago: American Medical Association; 2002.
169. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. *Users' guides to the medical literature. IX: a method for grading health care recommendations*. *JAMA* 1995; 274(22): 1800-1804.
170. Guyatt GH, Sackett DL, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy - randomized trials in individual patients. *N Engl J Med* 1986; 314(14): 889-892.
171. Guyatt GH, Tugwell P, Feeny DH, Drummond MF, Haynes RB. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J Chronic Dis* 1986; 39(4): 295-304.
172. Guyatt GH, Vist GE, Falck-Ytter Y, Kunz R, Magrini N, Schünemann HJ. An emerging consensus on grading recommendations? *ACP J Club* 2006; 144(1): A8-A9.
173. Harbour RT, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001; 323(7308): 334-336.

-
174. Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G et al. Applying systematic review methods to studies of people's views: an example from public health research. *J Epidemiol Community Health* 2004; 58(9): 794-800.
 175. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
 176. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15(4): 361-387.
 177. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001; 20(3 Suppl): 21-35.
 178. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006; 144(6): 427-437.
 179. Haynes RB, Cotoi C, Holland J, Walters L, Wilczynski N, Jedraszewski D et al. Second-order peer review of the medical literature for clinical practitioners. *JAMA* 2006; 295(15): 1801-1808.
 180. Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *Evid Based Med* 2002; 7: 36-38.
 181. Health on the Net (HON). HON Code of Conduct (HONcode) für medizinische und Gesundheits-Websites [*HONcode for medical and health information websites*]. [Online]. 1997 [Accessed on 25.09.2007]. URL: <http://www.hon.ch/HONcode/German>.
 182. Helou A, Ollenschläger G. Ziele, Möglichkeiten und Grenzen der Qualitätsbewertung von Leitlinien: ein Hintergrundbericht zum Nutzermanual der Checkliste "Methodische Qualität von Leitlinien" [*Aims, possibilities, and limits of the quality evaluation of guidelines. A background report on the user manual for the checklist "Methodological quality of guidelines"*]. *Z Ärztl Fortbild Qualitätssich* 1998; 92(5): 361-365.
 183. Herxheimer A, McPherson A, Miller R, Shepperd S, Yaphe J, Ziebland S. Database of patients' experiences (DIPEX): a multi-media approach to sharing experiences and information. *Lancet* 2000; 355(9214): 1540-1543.
 184. Herxheimer A, Ziebland S. DIPEX: fresh insights for medical practice. *J R Soc Med* 2003; 96(5): 209-210.
 185. Higgins JPT, Green S, Eds. *Cochrane handbook for systematic reviews of interventions*, Version 5.0.0 (updated February 2008) [Online]. 2008 [Accessed on 13.05.2008]. URL: <http://www.cochrane-handbook.org>.
 186. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21(11): 1539-1558.
 187. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
 188. Hill AB, Ed. *Controlled clinical trials*. Oxford: Blackwell; 1960.
 189. Holmes-Rovner M. International patient decision aid standards (IPDAS): Beyond decision aids to usual design of patient education materials. *Health Expect* 2007; 10(2): 103-107.
 190. Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. [Cochrane Review]. *Cochrane Database Syst Rev* 2007; 2: MR000001.
 191. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. [Cochrane Review]. *Cochrane Database Syst Rev* 2007; 2: MR000010.
 192. Hosmer DW, Taber S, Lemeshow S. The importance of assessing the fit of logistic regression models: a case study. *Am J Public Health* 1991; 81(12): 1630-1635.

193. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient Educ Couns* 2006; 61(2): 173-190.
194. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005; 5(1): 13.
195. Hung HM, O'Neill RT, Bauer P, Kohne K. The behavior of the P-value when the alternative hypothesis is true. *Biometrics* 1997; 53(1): 11-22.
196. Inan H. Measuring the success of your website: a customer-centric approach to website management. Frenchs Forest: Pearson Education Australia; 2002.
197. Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e.V.(IVW). Definitionen und technische Erläuterungen Version 1.9, Anlage 1 zu den IVW-Richtlinien für Online-Angebote [*Society for the Documentation of the Dissemination of Advertising Media (IVW). Definitions and technical explanations. Version 1.9. Appendix 1 on the IVW guidelines for online offers*]. Berlin: IVW; 2007.
198. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Vereinbarung über die vertrauliche Behandlung von Unterlagen [*Institute for Quality and Efficiency in Health Care. Agreement on the confidential handling of documents*]. [Online]. 2005 [Accessed on 30.11.2006]. URL: http://www.iqwig.de/download/VertraulichkeitIQWIG-VFA_19-8-05_final_2.pdf.
199. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). [Datenschutzerklärung [*Institute for Quality and Efficiency in Health Care. Data protection declaration*]]. [Online]. 2007 [Accessed on 25.09.2007]. URL: <http://www.gesundheitsinformation.de/nutzungsbedingungen.51.de.html>.
200. International Conference on Harmonisation E9 Expert Working Group. ICH harmonised tripartite guideline. Statistical principles for clinical trials. *Stat Med* 1999; 18(15): 1905-1942.
201. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonised tripartite guideline. The extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life-threatening conditions. Geneva: ICH; 1994.
202. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Official web site for ICH [Online]. 2004 [Accessed on 20.10.2004]. URL: <http://www.ich.org>.
203. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; 294(2): 218-228.
204. Ioannidis JPA, Evans S, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004; 141(10): 781-788.
205. Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001; 285(4): 437-443.
206. Ioannidis JPA, Mulrow CD, Goodman SN. Adverse events: the more you search, the more you find. *Ann Intern Med* 2006; 144(4): 298-300.
207. Irani J, Brown CT, Van der Meulen JH, Emberton M. A review of guidelines on benign prostatic hyperplasia and lower urinary tract symptoms: are all guidelines the same? *BJU Int* 2003; 92(9): 937-942.
208. Jackson D. The power of the standard test for the presence of heterogeneity in meta-analysis. *Stat Med* 2006; 25(15): 2688-2699.
209. Jackson N, Waters E. Criteria for the systematic review of health promotion and public health interventions. *Health Promot Int* 2005; 20(4): 367-374.
210. Jadad AR. Randomised controlled trials: a user's guide. London: BMJ Publishing Group; 1998.

-
211. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ* 1997; 156(10): 1411-1416.
 212. Jadad AR, Gagliardi A. Rating health information on the internet: navigating to knowledge or to Babel? *JAMA* 1998; 279(8): 611-614.
 213. James Lind Library. Editorial commentary: Avoiding biased comparisons [Online]. 2007 [Accessed on 01.05.2007]. URL: <http://www.jameslindlibrary.org/essays/bias/avoiding-biased-comparisons.pdf>.
 214. James A, Horton R, Collingridge D, McConnell J, Butcher J. The Lancet's policy on conflicts of interest. *Lancet* 2003; 363(9402): 8-9.
 215. Jefferson T, Rudin M, Brodny FS, Davidoff F. Editorial peer-review for improving the quality of reports of biomedical studies [Cochrane Methodology Review]. *Cochrane Database Methodol Rev* 2006; 1: MR000016.
 216. Jones B, Jarvis P, Lewis J, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996; 313(7048): 36-39.
 217. Jones B, Kenward MG. Design and analysis of cross-over trials. London: Chapman & Hall; 1989.
 218. Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: Systematic review. *BMJ* 2006; 333(7572): 782.
 219. Jull A, Bennett D. Do n-of-1 trials really tailor treatment? *Lancet* 2005; 365(9476): 1992-1994.
 220. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002; 31(1): 115-123.
 221. Kaltenthaler E, Brazier J, De Nigris E, Tumor I, Ferriter M, Beverley C et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. *Health Technol Assess* 2006; 10(33): iii, xi-iii, 168.
 222. Kaptchuk TJ, Stason WB, Davis RB, Legedza AR, Schnyer RN, Kerr CE et al. Sham device v inert pill: randomised controlled trial of two placebo treatments. *BMJ* 2006; 332(7538): 391-397.
 223. Kass PH, Gold EB. Modern epidemiologic study designs. In: Ahrens W, Pigeot I (Ed). *Handbook of Epidemiology*. Berlin: Springer; 2005: 321-344.
 224. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar S, Grimmer KA. A systematic review of the content of critical appraisal tools. *BMC Med Res Methodol* 2004; 4: 22.
 225. Katz MH. Multivariable analysis: a primer for readers of medical research. *Ann Intern Med* 2003; 138(8): 644-650.
 226. Kerr CE, Murray E, Stevenson F, Gore C, Nazareth I. Internet interventions for long-term conditions: patient and caregiver quality criteria. *J Med Internet Res* 2006; 8(3): e13.
 227. Kettunen T, Liimatainen L, Villberg J, Perko U. Developing empowering health counseling measurement. Preliminary results. *Patient Educ Couns* 2006; 64(1-3): 159-166.
 228. Kickbusch IS. Health literacy: addressing the health and education divide. *Health Promot Int* 2001; 16(3): 289-297.
 229. Kieser M, Röhm J, Friede T. Power and sample size determination when assessing the clinical relevance of trial results by 'responder analyses'. *Stat Med* 2004; 23(21): 3287-3305.
 230. Kim SY, Millard RW, Nisbet P, Cox C, Caine ED. Potential research participants' views regarding researcher and institutional financial conflicts of interest. *J Med Ethics* 2004; 30(1): 73-79.
 231. Köbberling J. Der Zweifel als Triebkraft des Erkenntnisgewinns in der Medizin [*Doubt as the impetus for gaining knowledge in medicine*]. In: Kunz R, Ollenschläger G, Raspe H, Jonitz G, Donner-Banzhoff N (Ed). *Lehrbuch evidenz-basierte Medizin in Klinik und Praxis*. Köln: Deutscher Ärzteverlag; 2007: 3-14.
 232. Köbberling J, Trampisch HJ, Windeler J. Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem* 1990; 28(12): 873-879.
-

-
233. Koch G. No Improvement - still less than half of the Cochrane reviews are up to date. In: Cochrane Collaboration (Ed). XIV Cochrane Colloquium. Programme and Abstract Book. Dublin: Cochrane Collaboration; 2006: 104.
234. Kolman J, Meng P, Scott G. Good clinical practice: standard operating procedures for clinical researchers. Chichester: Wiley; 1998.
235. Kommission der Europäischen Gemeinschaften. Richtlinie 2003/63/EG der Kommission vom 25. Juni 2003 zur Änderung der Richtlinie 2001/83/EG des Europäischen Parlaments und des Rates zur Schaffung eines Gemeinschaftskodexes für Humanarzneimittel, Abschnitt 5.2.5.1: Berichte über kontrollierte klinische Studien zur angegebenen Indikation [*Commission of the European Communities. Directive 2003/63/EC of the Commission of 25 June 2003 amending the Directive 2001/83/EC of the European Parliament and the Council on the community code relating to the medicinal products for human use. Section 5.2.5.1: Study reports of controlled clinical studies pertinent to the claimed indication*]. Amtsbl Eur Gemeinschaften 2005; L 159: 76.
236. Kools M, Van de Wiel MW, Ruiter RA, Kok G. Pictures and text in instructions for medical devices: effects on recall and actual performance. Patient Educ Couns 2006; 64(1-3): 104-111.
237. Köpke S, Berger B, Steckelberg A, Meyer G. In Deutschland gebräuchliche Bewertungsinstrumente für Patienteninformationen: eine kritische Analyse [*A critical analysis of assessment instruments for patient information commonly used in Germany*]. Z Ärztl Fortbild Qual Gesundh wes 2005; 99(6): 353-357.
238. Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. JAMA 2006; 296(10): 1286-1289.
239. Kranich C. Patientenkompetenz: was müssen Patienten wissen und können? [*Patient competence: What must patients know and be able to do?*]. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz 2004; 47(10): 950-956.
240. Krug S. Don't make me think: a common sense approach to web usability. Indianapolis: New Riders Publishing; 2000.
241. Kunz R, Vist GE, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2007; 2: MR000012.
242. Laaser U, Hurrelmann K. Gesundheitsförderung und Krankheitsprävention [*Health promotion and disease prevention*]. In: Hurrelmann K, Laaser U (Ed). Handbuch Gesundheitswissenschaften. Weinheim: Juventa; 1998: 395-424.
243. Lange S, Freitag G. Choice of delta: requirements and reality. Results of a systematic review. Biom J 2005; 47(1): 12-27.
244. Laupacis A, Wells G, Richardson WS, Tugwell P, Evidence-Based Medicine Working Group. Users' guides to the medical literature. V: how to use an article about prognosis. JAMA 1994; 272(3): 234-237.
245. Lehoux P, Tailliez S, Denis JL, Hivon M. Redefining health technology assessment in Canada: diversification of products and contextualization of findings. Int J Technol Assess Health Care 2004; 20(3): 325-336.
246. Lewis D. Computer-based approaches to patient education: a review of the literature. J Am Med Inform Assoc 1999; 6(4): 272-282.
247. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001; 322(7300): 1479-1480.
248. Leys M. Health care policy: qualitative evidence and health technology assessment. Health Policy 2003; 65(3): 217-226.
249. Liberati A, Sheldon TA, Banta HD. EUR-ASSESS project subgroup report on methodology: methodological guidance for the conduct of health technology assessment. Int J Technol Assess Health Care 1997; 13(2): 186-219.

-
250. Lipkus IM, Hollands JG. The visual communication of risk. *J Natl Cancer Inst Monogr* 1999; (25): 149-163.
 251. Little RJA, Rubin DB. *Statistical analysis with missing data*. Hoboken: Wiley; 2002.
 252. Loke YK, Price D, Herxheimer A, Eds. Chapter 14: Adverse effects. In: Higgins JPT, Green S, Eds. *Cochrane handbook for systematic reviews of interventions*, Version 5.0.0 (updated February 2008) [Online]. 2008 [Accessed on 13.05.2008]. URL: <http://www.cochrane-handbook.org>.
 253. Lord SJ, Irwig LM, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med* 2006; 144(11): 850-855.
 254. Macaskill P, Walter SD, Irwig LM. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001; 20(4): 641-654.
 255. MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG, Southern California Evidence-Based Practice Center. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol* 2003; 56(1): 44-51.
 256. Malterud K. The art and science of clinical knowledge: evidence beyond measures and numbers. *Lancet* 2001; 358(9279): 397-400.
 257. March JS, Silva SG, Compton S, Shapiro M, Califf R, Krishnan R. The case for practical clinical trials in psychiatry. *Am J Psychiatry* 2005; 162(5): 836-846.
 258. Martini P. *Methodenlehre der therapeutischen Untersuchung [Methodology of therapeutic investigations]*. Berlin: Julius Springer; 1932.
 259. McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA* 2003; 289(19): 2545-2553.
 260. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000; 356(9237): 1228-1231.
 261. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ* 2002; 324(7351): 1448-1451.
 262. Medical Services Advisory Committee. *Guidelines for the assessment of diagnostic technologies*. 2005.
 263. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; 354(9193): 1896-1900.
 264. Moher D, Pham B, Klassen TP, Schulz K, Berlin JA, Jadad AR et al. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000; 53(9): 964-972.
 265. Moher D, Schulz K, Altman DG, CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001; 134(8): 657-662.
 266. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007; 4(3): e78.
 267. Molenaar S, Sprangers M, Oort F, Rutgers E, Luiten E, Mulder J et al. Exploring the black box of a decision aid: what information do patients select from an interactive cd-rom on treatment options in breast cancer? *Patient Educ Couns* 2007; 65(1): 122-130.
 268. Montori VM, Wilczynski N, Morgan D, Haynes RB. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005; 330(7482): 68.
 269. Moulton B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. *Health Expect* 2004; 7(2): 165-175.

-
270. Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. *Ann Intern Med* 2002; 136(2): 122-126.
271. Munday J. *Introducing translation studies*. London: Routledge; 2001.
272. Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess* 1998; 2(16)
273. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease [Cochrane Review]. *Cochrane Database Syst Rev* 2005; 4: CD004274.
274. National Advisory Committee on Health and Disability. *Screening to improve health in New Zealand. Criteria to assess screening*. 2003.
275. National Health and Medical Research Council. *Statement on consumer and community participation in health and medical research*. Canberra: 2002.
276. National Institute of Clinical Studies. *The impact of the internet on consumers health behaviour*. Melbourne: NICS; 2003.
277. National Resource Centre for Consumer Participation in Health. *Feedback, participation and diversity: a literature review*. Canberra: Brown & Wilton Integrated Publishing Services; 2000.
278. National Resource Centre for Consumer Participation in Health. *Methods and models of consumer participation*. Canberra: 2004.
279. NHS Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. CRD Report 4 (2nd edition) [Online]*. 2001 [Accessed on 07.05.2007]. URL: <http://www.york.ac.uk/inst/crd/report4.htm>.
280. NHS Centre for Reviews and Dissemination. *Effectiveness matters. The database of abstracts of reviews of effects (DARE)*. 2002; 6(2): 1-4.
281. Nielsen J. *Designing web usability*. Indianapolis: New Riders Publishing; 2000.
282. Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. *Cochrane Database Syst Rev* 2006; 3: CD004563.
283. Nutbeam D. Health promotion glossary. *Health Promot Int* 1998; 13(4): 349-364.
284. Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J et al. Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. *Health Technol Assess* 2004; 8(15): 1-IV.
285. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *CMAJ* 1988; 138(8): 697-703.
286. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44(11): 1271-1278.
287. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992; 116(1): 78-84.
288. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; 44(1): 91-98.
289. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998; 17(24): 2815-2834.
290. Pennekamp PH, Diedrich O, Schmitt O, Kraft CN. Prävalenz und Stellenwert der Internetnutzung orthopädischer Patienten [*Prevalence and relevance of Internet use in orthopaedic patients*]. *Z Orthop* 2006; 144(5): 459-463.
291. Perleth M, Jakubowski E, Busse R. What is 'best practice' in health care? State of the art and perspectives in improving the effectiveness and efficiency of the European health care systems. *Health Policy* 2001; 56(3): 235-250.

-
292. Peters JL, Sutton A, Jones D, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006; 295(6): 676-680.
 293. Petticrew M, Song F, Wilson P, Wright K, NHS CRD Review DaIT. Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (DARE). *Int J Technol Assess Health Care* 1999; 15(4): 671-678.
 294. Pham B, Platt R, McAuley L, Klassen TP, Moher D. Is there a "best" way to detect and minimize publication bias? An empirical evaluation. *Eval Health Prof* 2001; 24(2): 109-125.
 295. Pocock SJ. *Clinical trials: a practical approach*. Chichester: Wiley; 1983.
 296. Pocock SJ, Collier TJ, Dandreo KJ, De Stavola BL, Goldman MB, Kalish LA et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 2004; 329(7471): 883-887.
 297. Poensgen A, Larsson S. *Patients, physicians and the internet: myth, reality, and implications*. Boston: Boston Consulting Group; 2001.
 298. Poynard T, Munteanu M, Ratziu V, Benhamou Y, Di Martino V, Taieb J et al. Truth survival in clinical research: an evidence-based requiem? *Ann Intern Med* 2002; 136(12): 888-895.
 299. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989; 8(4): 431-440.
 300. Rennie D. Editorial peer review: its development and rationale. In: Godlee F, Jefferson T (Ed). *Peer review in health sciences*. London: BMJ Publishing Group; 2003: 1-13.
 301. Richter JG, Becker A, Specker C, Monser R, Schneider M. Krankheitsbezogene Internetnutzung bei Patienten mit entzündlichen-rheumatischen Systemerkrankungen [*Disease-oriented Internet use in outpatients with inflammatory rheumatic diseases*]. *Z Rheumatol* 2004; 63(3): 216-222.
 302. Roebruck P, Elze M, Hauschke D, Leverkus F, Kieser M. Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme [*Literature review on sample size planning for equivalence problems*]. *Inform Biom Epidemiol Med Biol* 1997; 28(2): 51-63.
 303. Röhm J, Hauschke D, Koch A, Pigeot I. Biometrische Verfahren zum Wirksamkeitsnachweis im Zulassungsverfahren: Nicht-Unterlegenheit in klinischen Studien. [*Biometric approaches to demonstrate in approval procedures: Non-inferiority in clinical trials*]. *Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz* 2005; 48: 562-571.
 304. Rossouw JE. Estrogens for prevention of coronary heart disease. Putting the brakes on the bandwagon. *Circulation* 1996; 94(11): 2982-2985.
 305. Rothwell PM. Treating individuals 2: subgroup analysis in randomised controlled trials. Importance, indications, and interpretation. *Lancet* 2005; 365(9454): 176-186.
 306. Royal Society. *Science and the public interest: communicating the results of new scientific research to the public*. London: Royal Society; 2006.
 307. Royall RM. The effect of sample size on the meaning of significance tests. *Am Stat* 1986; 40(4): 313-315.
 308. Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. *Health Technol Assess* 2003; 7(34): iii, ix-51.
 309. Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. *Stat Med* 2000; 19(14): 1831-1847.
 310. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Appl Stat* 1994; 43(3): 429-467.
 311. Rutten LJ, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). *Patient Educ Couns* 2005; 57(3): 250-261.
 312. Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; 32(1-2): 51-63.
-

313. Sackett DL. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ* 2004; 328(7448): 1136.
314. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. London: Churchill-Livingstone; 2000.
315. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R et al. Should meta-analysts search Embase in addition to Medline? *J Clin Epidemiol* 2003; 56(10): 943-955.
316. Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. *J Clin Epidemiol* 2006; 59(10): 1057-1063.
317. Sanger S, Lang B, Klemperer D, Thomeczek C, Dierks ML. Manual Patienteninformation - Empfehlungen zur Erstellung evidenzbasierter Patienteninformationen. [*Patient information manual: Recommendation on the preparation of evidence-based patient information*]. Berlin: AZQ; 2006.
318. Santo A, Laizner AM, Shohet L. Exploring the value of audiotapes for health literacy: a systematic review. *Patient Educ Couns* 2005; 58(3): 235-243.
319. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* 2005; 23(9): 2020-2027.
320. Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *J R Stat Soc [Ser A]* 1999; 162(1): 71-94.
321. Savoie I, Kazanjian A, Bassett K. Do clinical practice guidelines reflect research evidence? *J Health Serv Res Policy* 2000; 5(2): 76-82.
322. Sawicki PT. Qualitat der Gesundheitsversorgung in Deutschland: ein randomisierter Sechs-Lander-Vergleich aus Patientensicht [*Quality of health care in Germany. A randomised six-country comparison from patients' point of view*]. *Med Klin* 2005; 100(11): 755-768.
323. Schaeffer D, Dierks ML, Hurrelmann K, Keller A, Krause H, Schmidt-Kaehler S et al. Evaluation der Modellprojekte zur Patienten- und Verbraucherberatung nach § 65b Sozialgesetzbuch V - Abschlussbericht der wissenschaftlichen Begleitforschung fur die Spitzenverbande der GKV. [*Evaluation of the model project on patient and consumer counselling according to §65b Social Code Book V: Final report of the accompanying scientific research of the National Confederation of Regional Associations of the Statutory Health Insurance Funds*]. Bielefeld: 2004.
324. Schluter PJ, Ware RS. Single patient (n-of-1) trials with binary treatment preference. *Stat Med* 2005; 24(17): 2625-2636.
325. Schmidt-Kaehler S. Patienteninformation Online: Theoretische Grundlagen, Planung und Entwicklung eines Konzeptes fur die Patientenschulung im Internet [*Patient information online: Theoretical principles, planning, and development of a concept for online patient education*]. Bern: Hans Huber; 2004.
326. Schneider N, Dierks ML, Seidel G, Schwartz FW. The federal government commissioner for patient issues in Germany: initial analysis of the user inquiries. *BMC Health Serv Res* 2007; 7: 24.
327. Schunemann HJ, Best D, Vist GE, Oxman AD. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003; 169(7): 677-680.
328. Schunemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A et al. An official ATS statement: Grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006; 174(5): 605-614.
329. Scott JT, Harmsen M, Prictor MJ, Entwistle VA, Sowden AJ, Watt I. Recordings or summaries of consultations for people with cancer [Cochrane Review]. *Cochrane Database Syst Rev* 2003; 2: CD001539.
330. Senn S. The many modes of meta. *Drug Inf J* 2000; 34: 535-549.
331. Senn S. Trying to be precise about vagueness. *Stat Med* 2007; 26(7): 1417-1430.

-
332. Senn SJ. Inherent difficulties with active control equivalence studies. *Stat Med* 1993; 12(24): 2367-2375.
333. Senn SJ. Turning a blind eye: authors have blinkered view of blinding. *BMJ* 2004; 328(7448): 1135-1136.
334. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999; 281(20): 1900-1905.
335. Shekelle PG, Ortiz E, Rhodes S, Morton SC, Eccles MP, Grimshaw JM et al. Validity of the agency for healthcare research and quality clinical practice guidelines: how quickly do guidelines become outdated? *JAMA* 2001; 286(12): 1461-1467.
336. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med* 2007; 147(4): 224-233.
337. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? *Soc Sci Med* 2007; 64(9): 1853-1862.
338. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2007; 26(15): 2982-2999.
339. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994; 69(6): 979-985.
340. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton A. Publication and related biases. *Health Technol Assess* 2000; 4(10): 1-115.
341. Spek V, Cuijpers P, Nyklicek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. *Psychol Med* 2007; 37(3): 319-328.
342. Spiegelhalter DJ, Freedman LS, Parmar MK. Bayesian approaches to randomized trials. *J R Stat Soc [Ser A]* 1994; 157(3): 357-416.
343. Spiegelhalter DJ, Myles JP, Jones D, Abrams KR. Methods in health service research: an introduction to bayesian methods in health technology assessment. *BMJ* 1999; 319(7208): 508-512.
344. Statistisches Bundesamt. *Leben in Deutschland: Haushalte, Familien und Gesundheit - Ergebnisse des Mikrozensus 2005*. Wiesbaden: Statistisches Bundesamt; 2005.
345. Steckelberg A, Berger B, Kopke S, Heesen C, Muhlhauser I. Kriterien für evidenzbasierte Patienteninformationen. *Z Ärztl Fortbild Qual Gesundh wes* 2005; 99(6): 343-351.
346. Steckelberg A, Kasper J, Redegeld M, Muhlhauser I. Risk information - barrier to informed choice? a focus group study. *Soz Präventivmed* 2004; 49(6): 375-380.
347. Steiner JF. The use of stories in clinical research and health policy. *JAMA* 2005; 294(22): 2901-2904.
348. Sterne J, Egger M, Smith GD. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; 323(7304): 101-105.
349. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000; 283(15): 2008-2012.
350. Sullivan LM, Massaro JM, D'Agostino RB, Sr. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Stat Med* 2004; 23(10): 1631-1660.
351. Swift TL, Dieppe PA. Using expert patients' narratives as an educational resource. *Patient Educ Couns* 2005; 57(1): 115-121.
352. Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R et al. Integrating qualitative research with trials in systematic reviews. *BMJ* 2004; 328(7446): 1010-1012.
353. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21(11): 1559-1573.
-

-
354. Thompson WG. The placebo effect and health: combining science and compassionate care. Amherst: Prometheus Books; 2005.
 355. Thurow S. Search engine visibility. Indianapolis: New Riders; 2003.
 356. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc [Ser B]* 1996; 58(1): 267-288.
 357. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ* 2001; 322(7282): 355-357.
 358. Trevena LJ, Davey HM, Barratt A, Butow P, Caldwell P. A systematic review on communicating with patients about evidence. *J Eval Clin Pract* 2006; 12(1): 13-23.
 359. Tsay MY, Yang YH. Bibliometric analysis of the literature of randomized controlled trials. *J Med Libr Assoc* 2005; 93(4): 450-458.
 360. Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. *Science* 1977; 198(4318): 679-684.
 361. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003; 290(12): 1624-1632.
 362. UK National Screening Committee. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. Edinburgh: NSC; 2003.
 363. US Agency for International Development. Conducting focus group interviews. Washington: U.S. AID; 1996.
 364. US Agency for International Development. Conducting key informant interviews. Washington: U.S. AID; 1996.
 365. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Patient-reported outcome measures: use in medical product development to support labeling claims. Draft guidance. Rockville: CDER; 2006.
 366. Van den Brink-Muinen A, Verhaak PF, Bensing JM, Bahrs O, Deveugele M, Gask L et al. Doctor-patient communication in different European health care systems: relevance and performance from the patients' perspective. *Patient Educ Couns* 2000; 39(1): 115-127.
 367. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002; 21(4): 589-624.
 368. van Tinteren H, Hoekstra OS, Smit EF, van den Bergh JH, Schreurs AJ, Stallaert RA et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet* 2002; 359(9315): 1388-1393.
 369. Vandenbroucke JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Ann Intern Med* 2007; 147(8): W163-W194.
 370. Vidanapathirana J, Abramson MJ, Forbes A, Fairley C. Mass media interventions for promoting HIV testing [Cochrane Review]. *Cochrane Database Syst Rev* 2005; 3: CD004775.
 371. Villar J, Mackey ME, Carroli G, Donner A. Meta-analyses in systematic reviews of randomized controlled trials in perinatal medicine: comparison of fixed and random effects models. *Stat Med* 2001; 20(23): 3635-3647.
 372. Virtanen H, Leino-Kilpi H, Salanterä S. Empowering discourse in patient education. *Patient Educ Couns* 2007; 66(2): 140-146.
 373. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.

-
374. W3C Web Accessibility Initiative. Web Accessibility Initiative (WAI) [Online]. 2006 [Accessed on 21.09.2006]. URL: <http://www.w3.org/WAI/>.
375. Wallis EJ, Ramsay LE, Ul Haq I, Ghahramani P, Jackson PR, Rowland-Yeo K et al. Coronary and cardiovascular risk estimation for primary prevention: validation of a new sheffield table in the 1995 Scottish health survey population. *BMJ* 2000; 320(7236): 671-676.
376. Walter U, Schwartz FW. Prävention. In: Schwartz FW (Ed). *Das Public Health Buch: Gesundheit und Gesundheitswesen*. München: Urban und Fischer; 2003: 189-214.
377. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs. non-web-based interventions: a meta-analysis of behavioral change outcomes. *J Med Internet Res* 2004; 6(4): e40.
378. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006; 25(2): 183-203.
379. Whitehead J. *The design and analysis of sequential clinical trials*. Chichester: Ellis Horwood; 1983.
380. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3: 25.
381. Windeler J. Bedeutung randomisierter klinischer Studien mit relevanten Endpunkten für die Nutzenbewertung. In: Gesundheitsforschungsrat (GFR) des Bundesministeriums für Bildung und Forschung (BMBF) (Ed). *Begriffsdefinitionen und Einführung - Dokumentation des ersten gemeinsamen Workshops von GFR und IQWiG am 4. September 2007 in Berlin* [*The importance of randomised controlled trials with relevant endpoints for the benefit assessment of interventions. In: The Health Research Council of the Federal Ministry of Education and Research (Ed). Definition of terms and introduction - Documentation of the first conjoint workshop of the Health Research Council and IQWiG on 4 September 2007 in Berlin*]. Bonn: Gesundheitsforschungsrat (GFR) des Bundesministeriums für Bildung und Forschung (BMBF); 2007: 26-31.
382. Windeler J, Conradt C. Wie können "Signifikanz" und "Relevanz" verbunden werden? [*How can "significance" and "relevance" be connected?*]. *Med Klin* 1999; 94(11): 648-651.
383. Windeler J, Lange S. Nutzenbewertung in besonderen Situationen – Seltene Erkrankungen [*Benefit assessment of interventions in special situations – Rare diseases*]. *Z Evid Fortb Qual Gesundh wes* 2008; 102(1): 25-30.
384. Windeler J, Ziegler S. EBM-Splitter: Evidenzklassifizierungen [*EBM notes - Evidence classifications*]. *Z Ärztl Fortbild Qualitätssich* 2003; 97: 513-514.
385. Wofford JL, Smith ED, Miller DP. The multimedia computer for office-based patient education: a systematic review. *Patient Educ Couns* 2005; 59(2): 148-157.
386. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336(7644): 601-605.
387. Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: a systematic review. *CMAJ* 1997; 156(12): 1705-1712.
388. Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. *BMJ* 2007; 335(7631): 1202-1205.
389. Ziebland S, McPherson A. Making sense of qualitative data analysis: an introduction with illustrations from DIPEX (personal experiences of health and illness). *Med Educ* 2006; 40(5): 405-414.
390. Ziegler DK, Mosier MC, Buenaver M, Okuyemi K. How much information about adverse effects of medication do patients want from physicians? *Arch Intern Med* 2001; 161(5): 706-713.