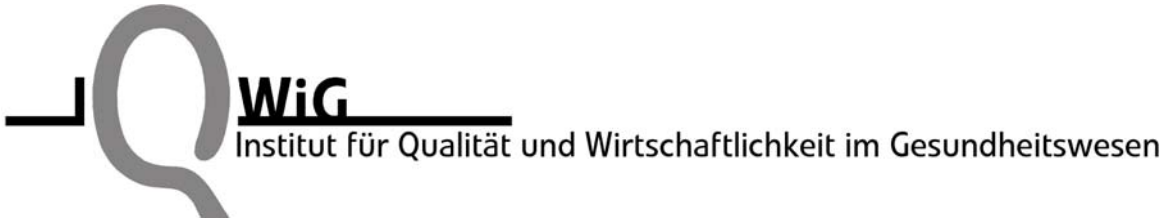


Methods[a]

9 November 2005

Produced by the Institute's Steering Committee[b]

Contact:

Institute for Quality and Efficiency in Health Care

(Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; IQWiG)

Dillenburger Str. 27

51105 Cologne

Germany

Tel.: +49 (0)221 – 35685-0

Fax: +49 (0)221 – 35685-1

E-mail: methoden@iqwig.de

---

[a]Translation based on Version 1.0 of "Methoden" (published on 1 March 2005).

[b]Bastian H, Bender R, Kaiser T, Kirchner H, Lange S, Müller de Cornejo G, Noelle G, Sawicki PT.

General comments:

The first draft of the Institute's methods paper was produced and published online 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. The present version (Version 1.0) was subsequently produced. This document will be reviewed annually, unless errors in the document or relevant developments necessitate prior updating. The valid version of the Institute's methods paper will be indicated for every product generated by the Institute. Any required modifications of methodology will be declared.

Preamble

With the introduction of the health care reform in 2003 (Health Care Modernisation Act; *Gesundheitssystem-Modernisierungsgesetz, GMG*), legislation determined the establishment of a new Institute, independent of the state, within the German health care system. The Federal Joint Committee *(Gemeinsamer Bundesausschuss, G-BA)* set up this scientific institution in the form of a private foundation. The sole purpose of the foundation is the creation and maintenance of the Institute for Quality and Efficiency in Health Care *(Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, IQWiG)*. The foundation's bodies include a Foundation Board and a five-member Board of Directors. The Institute is an establishment of the foundation and is under independent scientific management. The Institute's advisory committees are a 30-member Board of Trustees and a Scientific Advisory Board. The Scientific Advisory Board, with a minimum of six and a maximum of twelve members, is appointed by the Board of Directors. Until 2008, the seat of the Institute will be Cologne.

The Institute is responsible for the scientific evaluation of the effects, quality and efficiency of health care services. This includes the evaluation of clinical practice guidelines, the submission of recommendations on disease management programmes, the evaluation of the effects of pharmaceuticals, and the publication of health information for patients and consumers.

The Institute, as the foundation's professionally independent scientific establishment, will address fundamental issues relating to the quality and efficiency of services provided by statutory health insurance, taking into account specific factors such as age, gender and living conditions. This includes the following areas in particular:

- Investigation, description and evaluation of the current state of medical knowledge on diagnostic tests and therapeutic techniques for selected diseases;
- Production of scientific reports, expert opinions and statements on the quality and efficiency of services provided by statutory health insurance;
- Evaluation of evidence-based guidelines for the epidemiologically most important diseases;
- Submission of recommendations on disease management programmes;
- Evaluation of the effects of pharmaceuticals;
- Provision of easily understandable general information on the quality and efficiency of health care to the public.

The Institute's responsibility in these areas is to support the Federal Joint Committee in fulfilling its legislative duties by submitting recommendations and providing scientific advice.

The Institute's work should contribute to continuous improvement in the quality of health care for the public. The Institute's aim is to develop independent scientific capacity to answer the research questions posed, to evaluate medical issues and concepts relevant to health care, and to assess research requirements relevant to patients' needs. The information compiled will be relayed to the Federal Ministry of Health (and Social Security), the Federal Joint Committee and the public.

The Institute's duties include the production of reports on specific topics requested by the Federal Joint Committee or the Federal Ministry of Health, as well as the initiation, coordination and publication of scientific work in areas where care-related medical knowledge needs to be complemented. The Institute regularly screens and evaluates literature for care-related medical innovations and distributes this information in an understandable form. On the basis of international literature and its own literature searches, the Institute can provide proposals for research related to innovative health care, initiate and participate in research projects and publish its findings.

The Institute's Steering Committee includes the Institute's Director and the department heads. This committee produces and modifies this methods paper and develops and monitors the Institute's working procedures. The publication and ongoing discussion of the Institute's methods paper are explicitly desired, and should contribute both to the transparency of the Institute's work and to a continuously dynamic improvement of these methods.

The Institute is well aware that not all steps in an evaluation process can be presented in advance and in detail in every case. Individual procedures are, amongst other things, dependent on the respective research question, the scientific evidence available and expert opinions received. This document should therefore be regarded as a guideline when evaluating a medical technique. In the respective report plans, the organisation and description of the evaluation process will be case-related and based on this methods paper.

In order to use the available resources meaningfully and efficiently, the Institute will consider and, if applicable, use the preparatory work of other national and international health care institutions.

Table of contents

List of abbreviations

CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CMA	Cost-minimisation analysis
CPG	Clinical practice guideline
CUA	Cost-utility analysis
DARE	Database of Abstracts of Reviews of Effectiveness
DMP	Disease management programme
GCP	Good clinical practice
HTA	Health technology assessment
QALY	Quality-adjusted life year
RCT	Randomised controlled trial

List of figures

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. Scientific methods and statistics

1.1 Description of risks and effects

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 strength 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 (possibly after appropriate weighting). For categorical outcome variables, the usual effect and risk measures of 2x2 tables apply [1]. After specification of a primary effect measure for data analysis, both absolute measures (e.g. absolute risk reduction or number needed to treat) and relative measures (e.g. relative risk or odds ratio) should be used for the descriptive presentation of data. Chapter 8 of the Cochrane Reviewers' Handbook provides a well-structured summary of the advantages and disadvantages of typical effect measures [2]. Agresti describes the specific aspects to be considered for ordinal data [3,4].

It is mandatory to describe the degree of statistical uncertainty for every effect estimate. The calculation of the standard error and the confidence interval are frequently employed methods for this purpose. Whenever possible, an adequate confidence interval should be stated, 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. give an overview of the most common methods for calculation of confidence intervals [5].

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 respective data situation (e.g. very small samples) and the research question posed. Agresti provides an up-to-date discussion on exact methods [6].

References

[1] 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: 116-121.

[2] Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In: Alderson P, Green A, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.2 [updated March 2004]; Section 8. In: The Cochrane Library, Issue 1, 2004. Chichester: Wiley; 2004.

[3] Agresti A. Categorical Data Analysis. New York: Wiley; 1990.

[4] Agresti A. Modelling ordered categorical data: Recent advances and future challenges. Stat Med 1999; 18: 2191-2207.

[5] Altman DG, Machin D, Bryant TM, Gardner MJ, editors. Statistics with Confidence: Confidence Intervals and Statistical Guidelines, 2nd ed. London: BMJ Books; 2000.

[6] Agresti A. Dealing with discreteness: Making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. Stat Meth Med Res 2003; 12: 3-21.

1.2 Evaluation of statistical significance

Statistical significance tests enable the distinction between positive test results due to genuine effects and purely accidental findings attributable to the usual data variability. The convention of speaking of a “statistically significant result” if $p<0.05$ applies is often meaningful. One should not, however, insist on a significance level of $\alpha<0.05$. It may be necessary to demand a much smaller *p*-value, depending on the research questions posed and hypothesis made. Conversely, there are situations where a higher significance level can be used.

A range of aspects should be considered when interpreting *p*-values. It must be clear which research question and data situation the significance level refers to and exactly how the statistical hypothesis is formulated. In particular, it should be clear whether a one or two-sided hypothesis applies [1] and whether this hypothesis is to be regarded as part of a multiple hypothesis problem [2]. These two issues, whether a one or two-sided hypothesis is to be formulated, and whether adjustments for multiple testing need to be made, are the subject of continual controversy in scientific literature.

Regarding the hypothesis formulation, a two-sided test problem is traditionally assumed. Exceptions include non-inferiority studies (q.v. Section 1.20). 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 studies demand that the typical significance level should be halved from 5% to 2.5% [3]. A basic principle, which always applies, is the clear a priori determination of the hypothesis formulation (one or two-sided) and the significance level.

If the investigated hypothesis is clearly part of a multiple hypothesis problem, adequate adjustment for multiple testing is required. Bender and Lange [4] provide an overview of the situations where this case applies and describe the appropriate methods available.

A non-significant finding should not per se be evaluated as evidence of the absence of an effect [5]. For evidence of equivalence, methods for equivalence hypotheses need to be employed (q.v. Section 1.20).

References

[1] Bland JM, Altman DG. One and two sided tests of significance. BMJ 1994; 309: 248.

[2] Tukey JW. Some thoughts on clinical trials, especially problems of multiplicity. Science 1977; 198: 679-684.

[3] ICH E9 Expert Working Group. ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials. Stat Med 1999; 18: 1905-1942.

[4] Bender R, Lange S. Adjusting for multiple testing - when and how? J Clin Epidemiol 2001; 54: 343-349.

[5] Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995; 311: 485.

1.3 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 strength of a possible effect but also by data variability and sample size. When interpreting the relevance of *p*-values, in particular the sample size of the underlying study needs to be taken into account [1]. 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 [2,3]. Consequently, the clinical relevance of a study's results can by no means be derived from a *p*-value alone.

Formal evaluation of the clinical relevance of study results is a methodological problem that is still largely unsolved. Only a few guidelines contain details of the definition of clinically relevant or irrelevant differences between groups. A first approach to assess the clinical relevance of study findings is the evaluation of the effect estimate and of the corresponding confidence interval (q.v. Section 1.1) 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 employment of a statistical test with a shifting of the null hypothesis in order to detect clinically relevant effects. A further option is to formulate a relevance criterion individually, e.g. in terms of a responder definition. Moreover, the individual assessment of affected patients plays an important role; the explanation of patient-relevant endpoints may provide indications for this purpose (q.v. Section 1.5).

References

[1] Royall RM. The effect of sample size on the meaning of significance tests. Am Stat 1986; 40: 313-315.

[2] Feinstein AR. Invidious comparisons and unmet clinical challenges. Am J Med 1992; 92: 117-120.

[3] Hung HMJ, O'Neill RT, Bauer P, Köhne K. The behavior of the *P*-value when the alternative hypothesis is true. Biometrics 1997; 53: 11-22.

1.4 Subgroup analyses

Subgroup analyses are discussed very critically in the methodological literature [1,2]. The interpretation of their results is mainly complicated by three factors:

- Their post-hoc character (unless subgroup analyses were planned a priori, were well-founded and were an integral part of the study protocol [or its amendments]).
- A statistically significant result achieved by multiple hypothesis testing (q.v. Section 1.2), unless this problem was accounted for by adjusting the confidence level.
- A statistically non-significant result due to a (frequently existing) lack of power; this lack of power may result in an inability to detect moderate treatment differences (by means of inferential statistics), unless an additional sample size estimation for a subgroup analysis was performed with adequate power (and, if necessary, with a correspondingly increased sample size) [3].

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

Furthermore, subgroup analyses are not interpretable if the characteristic was defined (or can be defined) after initiation of treatment (after randomisation), e.g. so-called responder analyses.

The statistical analysis of subgroup effects should be conducted by means of an 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 noted above, for some research questions subgroup analyses may represent the best scientific evidence currently available for the evaluation of effects in subgroups in the foreseeable future [4], as factors such as ethical considerations may argue against the reproduction of findings of subgroup analyses in a validation study.

References

[1] Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992; 116: 78-84.

[2] Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000; 355: 1064-1069.

[3] Brookes ST, Whitely E, Egger M, Davey Smith G, 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: 229-236.

[4] 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: 989-991.

1.5 Determination of patient-relevant endpoints and measures

As a rule, clinical studies are designed to measure disease-related outcomes. These outcomes may or may not be relevant for patients. Moreover, the magnitude of a benefit (q.v. Section 2.1) or of a risk for the patient may be assessed differently, depending on the individual patient. For example, people who are experiencing (or have experienced) disease or injury perceive and assess benefits and risks differently from those who are not (or have not) [1].

Many studies measure surrogate outcomes as assumed indicators of a patient-relevant outcome. The direct inference that a change in a surrogate variable means a change in a patient-relevant outcome is, however, not possible without question [2-4]. The Institute will therefore only draw on such surrogate outcome measures for the evaluation of a medical procedure on the condition that a sufficiently sound assessment of the causal chain (with evidence of patient-relevant outcome) has been made.

In addition to clinically relevant events such as mortality, morbidity and adverse effects of medical interventions, other outcomes or events can also be of relevance to patients, e.g.:

- Disease-related quality of life, including effects on activities of daily living and personal well-being;
- Convenience, circumstances, and time and cost of treatment;
- Treatment satisfaction and treatment preferences;
- Effects on relatives.

Results that are important for patients should be taken into account before evaluatory conclusions on interventions are made. Data derived from such results can substantially alter the conclusions of a systematic review [5]. In addition, it should be considered whether results are transferable to pa-

tient subgroups (e.g. stratified by gender or age) and whether results from long-term studies are available.

If derivable, it should be shown whether the effect size of an intervention is relevant for patients [5,6]. If possible, these data should be obtained by literature searches; findings from qualitative research should also be evaluated, including consultations with patients and relatives (e.g. focus group interviews).

The external validity of many patient-relevant outcome measures can be influenced by cultural and social conditions and can vary in the course of time. Moreover, the assessment of an intervention by patients can be affected by the mere fact that it is a new development.

References

[1] Slevin ML, Stubbs L, Plant HJ et al. Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public. BMJ 1990; 300: 1458-1460.

[2] Gøtzsche PC et al. Beware of surrogate outcome measures. Int J Tech Assess Health Care 1996; 12: 238-246.

[3] Liberati A, Sheldon TA, Banta HD et al. Eur-Assess project subgroup report on methodology: Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care 1997; 13: 186-219.

[4] The CAST investigators. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. New Engl J Med 1989; 321: 406-412.

[5] Dixon-Woods M et al. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.

[6] Egger M, Davey Smith G, Phillips AN. Meta-analysis principles and procedures. BMJ 1997; 315: 1533-1537.

1.6 Aspects of the evaluation of study quality

When assessing the general quality of studies, a range of aspects, which cannot all be outlined in detail here, play a role. In principle, a recognised standardised concept should be followed in a study, from planning to conduct to evaluation and presentation. This includes a study protocol describing all the important methods and procedures. The standards for studies are defined by the basic principles of good clinical practice (GCP) for randomised clinical trials [1,2] and by the guidelines and recommendations to ensure good epidemiological practice (GEP) in epidemiological studies [3]. A key quality criterion in studies is whether the study data were actually analysed in the way planned. This cannot usually be reliably concluded from the respective publications. A section

on sample size estimation may, however, at least provide indications in this regard. Furthermore, a comparison with the (possibly previously published) study protocol or with the corresponding publication on the study design may be useful.

The following relevant statements were formulated to improve the quality of publications:

- The CONSORT[c] statement on randomised controlled trials (RCTs) [4] and the corresponding explanatory document [5];
- The CONSORT statement on cluster randomised trials [6];
- The QUORUM[d] statement on meta-analyses of randomised trials [7];
- The TREND[e] statement on non-randomised intervention trials [8];
- The STARD[f] statement on diagnostic studies [9] and the corresponding explanatory document [10].

If a publication fails to conform to these statements, it may be an indication of deficiencies in the respective study. Additional key papers on this issue describe fundamental aspects of the assessment of the quality of studies [11–13].

The following principles are key factors in the evaluation of RCTs: adequate concealment, i.e. the unforeseeability and concealment of allocation to groups (e.g. by external randomisation in non-[double]blindable trials); blinded evaluation of outcome measures in non-[double]blindable trials (q.v.1.15); adequate application of the "intention-to-treat" principle; determination of a clear primary endpoint; and appropriate consideration of possible multiple testing problems (q.v. Section 1.2).

The evaluation of formal criteria provides essential indications for the quality of studies. It is, however, always necessary to make an evaluation beyond purely formal aspects, e.g. in order to detect errors, contradictions and inconsistencies in publications and assess their relevance in the interpretation of results.

[c]Consolidated Standards of Reporting Trials.
[d]The Quality of Reporting of Meta-analyses.
[e]Transparent Reporting of Evaluations with Nonrandomized Designs.
[f]Standards for Reporting of Diagnostic Accuracy.

References

[1] Kolman J, Meng P, Scott G. Good Clinical Practice. Standard Operating Procedures for Clinical Researchers. Chichester: Wiley, 1998.

[2] ICH Steering Committee. Official web site for the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). http://www.ich.org/. Access on 20 October 2004.

[3] Arbeitsgruppe Epidemiologische Methoden der Deutschen Arbeitsgemeinschaft für Epidemiologie (DAE), Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP). *(Guidelines and recommendations to ensure Good Epidemiological Practice [GEP]. Epidemiological Methods Task Group of the German Working Group on Epidemiology).* http://medweb.uni-muenster.de/institute/epi/dae/Empfehlungen.doc. Access on 21 October 2004.

[4] Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med 2001; 134: 657-662.

[5] Altman DG, Schulz KF, 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: 663-694.

[6] Campbell MJ, Elbourne DR, Altman DG for the CONSORT Group. CONSORT statement: Extension to cluster randomised trials. Br Med J 2004; 328: 702-708.

[7] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF et al. Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. Lancet 2000; 354: 1896-1900.

[8] Des Jarlais DC, Lyles C, Crepaz N for the TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. Am J Public Health 2004; 94: 361-366.

[9] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med 2003; 138: 40-44.

[10] 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: W1-12.

[11] Egger M, Davey Smith G, Altman DG, editors. Systematic Reviews in Health Care: Meta-Analysis in Context. London: BMJ Books; 2001.

[12] Alderson P, Green S, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.2 (updated March 2004). In: The Cochrane Library, Issue 1, 2004. Chichester: Wiley; 2004.

[13] Guyatt G, Rennie D, editors. Users' Guide to the Medical Literature. Chicago, IL: AMA Press, 2002.

1.7 Determination of adverse effects

The application of a medical intervention, whatever its nature (pharmaceutical, non-pharmaceutical, surgical, diagnostic, preventive, etc.) always carries the risk of adverse effects. The

term "adverse effects" signifies all events and effects that are individually perceived or objectively identifiable physical or mental detriments. These detriments have the potential to cause mild to severe, short to long-term impairments in quality of life and/or a reduction in life expectancy.

The description of the potential risks of a medical intervention is an essential and important element in the production of a report describing the intervention. It ensures an informed, population-related, but also individual weighing of benefits and risks [1]. A prerequisite for this is that, analogously to the specification of the effect size of a medical intervention, the probabilities of potential adverse effects are estimated and reported using the available data.

The description of the risk profile of a medical intervention is, however, usually far more difficult to provide than a description of the benefits (q.v. Section 2.1) [2]. As a general rule, only a few patient-relevant effectiveness endpoints can be positively influenced by a medical intervention; in contrast, the occurrence of adverse effects is rather complex. Furthermore, the incidence of serious adverse events (e.g. organ failure, death) is usually low in clinical studies. It should also be noted that studies with the specific objective of detecting rare but serious adverse effects (including the description of a causal relationship to the medical intervention) are considerably underrepresented in medical research [3,5]. The reasons for this include the lack of adequately developed methods for clinical trials and the considerable resources needed to enable exclusion of a more than irrelevant increase in rare adverse events. Moreover, various conflicts of interest may possibly prevent a balanced description of benefits and risks.

The consequence of the above-mentioned obstacles is that in many cases, in spite of enormous efforts, the uncertainty of statements on adverse effects is greater than that of statements on positive effects [6].

It is necessary to find a meaningful balance between the completeness of the reporting and evaluation of adverse effects on the one hand and the amount of resources required on the other; consequently, it will be necessary to limit reporting and evaluation to relevant adverse effects.

In particular, adverse effects can be described as relevant wherever initial indications exist that these effects:

- May completely or almost completely eliminate the benefits of an intervention;
- May considerably differ from adverse effects occurring with (an) otherwise equivalent treatment option(s);
- May occur predominantly with (a) treatment option(s) (of several treatment options) that are particularly effective;

- May have a dose-effect relationship;
- May be regarded by patients as especially important (q.v. Section 1.5);
- May be accompanied by serious morbidity or even increased mortality or may be associated with substantial impairment of quality of life.

In the interests of patient safety, the Institute will observe the following principles when assessing and describing adverse effects:

- The basis for the selection of relevant adverse effects according to the above-mentioned criteria is a compilation of adverse effects and events that were frequently reported in connection with the respective medical intervention. In particular, this compilation is made based on data from controlled intervention trials in which the benefit of the intervention was specifically investigated. In addition, data are obtained from epidemiological (e.g. cohort or case-control studies) and uncontrolled studies (e.g. post-marketing surveillance studies, pharmacovigilance centres) as well as from animal trials and experiments to test pathophysiological constructs.
- If justified suspicion of the presence of an adverse effect emerges from the above-mentioned data sources, the occurrence of such an effect will be regarded as possible until it can be ruled out with sufficient certainty by the results of specific research. This principle applies in particular to serious adverse effects (the hierarchy of evidence corresponds to that of therapeutic studies; q.v. Sections 1.10 and 1.11). By "sufficient" it is meant that:
    - The corresponding study/studies, in their design and planning, were aimed primarily at showing the equivalence between a medical intervention and other treatment options, or placebo, or no intervention.
    - The (statistical) definition of equivalence is comparable with the one determined by the Institute, e.g. through patient interviews before the start of the evaluation procedure (q.v. Section 1.5).[g]

[g]When (statistically) defining equivalence, it should be noted that the width of the range accepted as equivalent usually correlates with the feasibility of the respective studies. For example, the demand that equivalence will only be accepted on exact agreement between event rates is not practicable.

References

[1] 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: 706-713.

[2] Derry S, Loke YK, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. BMC Medical Research Methodology 2001; 1: 7.

[3] 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: 427-437.

[4] Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials – an evaluation of seven medical areas. JAMA 2001; 285: 437-443.

[5] Bonhoeffer J, Zumbrunn B, Heininger U. Reporting of vaccine safety data in publications: systematic review. Pharmacoepidemiol Drug Saf 2004 Jun 16 [Epub ahead of print].

[6] Loke YK, Price D, Herxheimer A, on behalf of the Cochrane Adverse Effects Subgroup. Including adverse effects in your review. http://www.dsru.org/wwwboard/latestdraft.pdf. Access on 6 October 2004.

1.8 Evaluation of studies conducted with outdated methods

If an Institute's project includes the evaluation of older studies that do not satisfy current quality standards (q.v. Section 1.9) because they were planned and conducted at a time when these standards did not exist, then this will be taken into account in their evaluation. Their disadvantages and deficiencies will be shown and possible consequences discussed, but these deficiencies will not automatically lead to their exclusion from an evaluation.

1.9 Evaluation of different types of studies

Only the most relevant study designs in medical research are summarised here.

A distinction is made between observational studies and intervention studies. Observational studies often provide the first information on a topic via case reports or case series. These are of course susceptible to all kinds of bias, so that evidence on a specific research question can only be inferred to a limited extent (q.v. Section 1.10). The prevalence of diseases can be estimated from population-based cross-sectional studies, but no associations between variable exposures and diseases can be inferred. Other important epidemiological study types are case control studies [1], where exposures in cases and controls are assessed retrospectively, and cohort studies [2], where specific groups (cohorts) are observed over a period of time. Cohort studies are prospective in character, although retrospective cohort studies are also conducted in which past exposure is recorded (fre-

quently used in studies on occupational or pharmacological epidemiology). In principle, prospective designs are preferable to retrospective designs. However, case-control studies are frequently the only practicable way of gaining information about an association between exposures and rare diseases.

Intervention studies require a control group. In a design with dependent samples without a control group, the effect of an intervention cannot usually be inferred from a sole "before/after" comparison. Exceptions include diseases with a deterministic (or practically deterministic) course (e.g. ketoacidotic diabetic coma). Randomisation and blinding are quality criteria that increase the relevance of controlled studies (q.v. Sections 1.6 and 1.18). Parallel group studies [3], cross-over studies [4] and cluster randomised studies [5] are further common designs in clinical trials.

The use of adequate sequential designs should be considered if interim analyses are planned [6].

The choice of an adequate design in diagnostic and screening studies depends on their objectives, which may differ substantially (q.v. Sections 2.3 and 2.4).

In the last few years, the relatively new discipline of genetic epidemiology has emerged for the investigation of genetic factors that can cause the development and distribution of diseases [7]. In this field, there is a range of new, specific designs for genetic association and genetic coupling studies, which cannot be discussed in detail here.

References

[1] Breslow NE, Day NE. Statistical Methods in Cancer Research Vol. I: The Analysis of Case-Control Studies. Lyon: Int. Agency for Res. on Cancer; 1980.

[2] Breslow NE, Day NE. Statistical Methods in Cancer Research Vol. II: The Design and Analysis of Cohort Studies. Lyon: Int. Agency for Res. on Cancer; 1987.

[3] Pocock SJ. Clinical Trials: A Practical Approach. Chichester: Wiley; 1983.

[4] Jones B, Kenward MG. Design and Analysis of Cross-Over Trials. London: Chapman & Hall; 1989.

[5] Donner A, Klar J. Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold; 2000.

[6] Whitehead J. The Design and Analysis of Sequential Clinical Trials. Chichester: Ellis Horwood; 1983.

[7] Khoury MJ, Beaty TH, Cohen BH. Fundamentals of Genetic Epidemiology. New York: Oxford University Press; 1993.

1.10 Ranking of different study types/evidence levels

There are different approaches towards allocating particular evidence levels to specific study types to create ranking of evidence for the production of systematic reviews or development of guidelines [1,2]. However, no system currently exists that is generally accepted and universally applicable to all systematic reviews [3]. A rough and generally accepted hierarchy of study types can be adapted from the system created by the Scottish Intercollegiate Guidelines Network Grading Review Group [2]. According to this guideline, the highest evidence level for therapeutic studies is allocated to systematic reviews of RCTs, followed by single RCTs. Some guidelines further classify these trials according to quality. The following levels include non-randomised intervention studies, prospective observational studies, retrospective observational studies, non-experimental studies (e.g. case reports and case series) and, with the lowest evidence level, expert opinions not based on scientific rationale. This rough grading system needs to be adapted to the respective situation and research question and described in more detail.

References

[1] Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ et al. Users' guides to the medical literature. IX. A method for grading health care recommendations. JAMA 1995; 274: 1800-1804.

[2] Harbour R, Miller J, for the Scottish Intercollegiate Guidelines Network Grading Review Group. A new system for grading recommendations in evidence based guidelines. BMJ 2001; 323: 334-336.

[3] Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VSS, Grimme KA. A systematic review of the content of critical appraisal tools. BMC Med Res Meth 2004; 4: 22.

1.11 Relationship between study type and research question

The RCT is regarded as the study type of the highest quality. However, this must be seen in connection with the research questions posed. This design is only usually required when the objective of the study is to demonstrate the effectiveness of an intervention. Alternative study designs may be considered for other research questions; the most important ones are described below.

In many cases, a cross-sectional study is sufficient to investigate diagnostic tests, e.g. if the aim of the study is limited to the analysis of discrimination ability (q.v. Section 2.3). The optimum design to investigate prognostic factors is a prospective cohort study. Case-control studies are used to investigate the association between exposures and very rare diseases. If, however, diagnostic tests or prognostic factors are to be assessed as a strategy together with the consequences (e.g. initiation of

a therapy) resulting from the information gained, then an RCT is the design of choice (q.v. Sections 2.3, 2.4 and 2.10).

1.12 Evaluation of unpublished or partially published data

In practice, the problem that essential data or information is partially or totally missing for the evaluation of publications frequently arises. This mainly concerns so-called "grey" literature and abstracts, but also full publications. In these cases, the Institute's first step is to procure missing information, e.g. by contacting the authors. If it is only possible to obtain incomplete data sets, sensitivity analyses (best-case and worst-case scenarios) will be conducted, presented and discussed. Where relevant information is completely lacking, a publication cannot be evaluated and it will merely be stated that further data exist on a particular topic, but these are unavailable for evaluation.

1.13 Handling of raw data

For the scientific evaluation of medical services, one of the Institute's principal tasks is to collect and analyse published data from systematic literature searches. For certain research questions, the Institute may also evaluate raw data provided by external sources (e.g. health insurance funds) that have not previously been analysed. A prerequisite for a meaningful analysis of these data is that the framework within which these data were obtained is clear and that the plausibility and quality of the data can be reviewed. It is especially important to ensure that essential quality criteria are observed; e.g. for therapeutic studies the data should have been compiled according to GCP standards (q.v. Section 1.6). Furthermore, in most cases the provision of a study protocol will be necessary for adequate evaluation. Legal aspects of data protection will be taken into account when handling raw data (see also Section 1.16).

1.14 Evaluation of the uncertainty of results

In principle, every result of an empirical study is uncertain. The statistical uncertainty of a parameter estimate, which results from the limited sample size, can be quantified and assessed in the form of standard errors and confidence intervals (q.v. Section 1.1). These calculations are made on the assumption that the statistical method selected is correct and that no other systematic errors and biases exist. The uncertainties that arise because the actual conditions deviate more or less widely from the statistical model chosen remain unconsidered here [1]. Formal approaches exist that take

these general model uncertainties into account, e.g. Bayesian methods [2] or simulation techniques [3], but they have not been sufficiently developed and investigated to be routinely applied in practice [4,5]. In any case, a qualitative assessment of the general uncertainty of results based on the current literature addressing the respective topic should be demanded. Hill's classic causality criteria [6] are still a valid aid for this purpose.

References

[1] Chatfield C. Model uncertainty, data mining and statistical inference (with discussion). J R Stat Soc A 1995; 158: 419-466.

[2] Draper D. Assessment and propagation of model uncertainty (with discussion). J R Stat Soc B 1995; 57: 45-97.

[3] Phillips CV. Quantifying and reporting uncertainty from systematic errors. Epidemiology 2003; 14: 459-466.

[4] Hjort NL, Claeskens G. Frequentist model average estimators. J Am Stat Assoc 2003; 98: 879-899.

[5] Wang D, Zhang W, Bakhai A. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. Stat Med 2004; 23: 3451-3467.

[6] Hill AB. The environment and disease: Association or causation? Proc R Soc Med 1965; 58: 295-300.

1.15 Evaluation of non-blindable techniques

To avoid systematic bias when estimating effects (q.v. Section 1.18), controlled studies are if possible conducted in a randomised and double-blind way. However, in some situations, blinding of the intervention is not possible. Non-blinded studies can also lead to interpretable results; again, randomisation and the appropriate choice of outcome variables are important instruments to prevent bias. In non-(double)blindable studies, it is crucial to ensure adequate concealment (q.v. Section 1.6) and to select a "hard" objective endpoint as an outcome variable. At least it should be ensured that the outcome variable is independent of the (non-blinded) treating physician (investigator) or evaluated in a blinded manner independent of the treating physician (blinded determination of outcome measures).

1.16 Consideration of legal aspects of data protection/confidentiality

The processing of personal data within the Institute is conducted according to the respective federal data protection laws.[h] 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 originally were collected or are being collected by a third party, the corresponding declarations on compliance with legal regulations need to be supplied to the respective institution. Furthermore, for each individual case, the availability of the necessary legal requirements (informed consent, patient information, etc.) needs to be assessed beforehand.

A further aim is to receive personal data that is 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 coded this way for research purposes and individual research questions. In particular, possible reservations about transferring data to the Institute should thereby be dispelled.

With regard to the confidential handling of data from commercial enterprises, appropriate declarations guaranteeing the Institute's confidentiality will, if necessary, be made to third parties. Besides having the necessary technical infrastructure to ensure data safety, corresponding clauses obliging personnel to observe confidentiality are included in all the Institute's employment contracts. In individual cases, externally appointed persons or institutions must also make corresponding obligations towards the Institute.

References

[1] (German) Federal Data Protection Act of December 20, 1990 (BGBl. I 1990 S.2954), amended by law of September 14, 1994 (BGBl. I S. 2325).

1.17 Consideration of ethical aspects

The Institute's primary aim is to improve the health care of the population in Germany through its high-quality work. The Institute's main focus is the maximisation of overall as well as individual patient benefits and the strengthening of patient autonomy through health education and informa-

[h]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" (Federal Data Protection Act, [1]).

tion. The methods of evidence-based medicine are seen as essential and valuable tools for this purpose. The Institute will apply these tools conscientiously, taking their limitations into account.

Furthermore, the Institute is aware of its standing in the German health care system and especially of its responsibility towards the people and institutions using, performing, financing or developing health care services. Even though health care legislation requires strict separation of scientific evaluation on the one hand, and any decision for or against the inclusion of a medical intervention as a service provided by the statutory health insurance on the other, the Institute is well aware that its work may have a direct or indirect influence on health care. Consequently, the consideration of the possible or probable effects that the Institute's reports have or will have on individuals, population groups or occupational groups, as well as on institutions or commercial enterprises, constitutes an essential element in the Institute's work. The Institute will therefore ensure the involvement of individual representatives of groups and institutions affected by the Institute's projects. The Institute's central responsibility is the public interest. The Institute will do its utmost to resist any influences, whatever their origin, that seek to divert the Institute's work away from transparency and independence towards statements steered by specific interests.

The Institute will not ignore questions concerning the fairness of distribution of resources. Having only limited resources means that freedom of choice on the one hand is accompanied by limitation of the same on the other. The Institute will convey the message that the decision for or against a medical procedure must derive from a conscientious consideration of its benefits and drawbacks. In this regard, the Institute sees the consideration of impact on equity and on minorities and disadvantaged population groups as an important public responsibility.

Ethical issues also have priority in the Institute's own research projects. When producing a scientific report, it is necessary to consider the advantages and disadvantages for those affected. Furthermore, where necessary, advice on ethical issues should be sought during project planning and conduct.

1.18 Description of types of bias

Bias is the systematic distortion of the estimation of effects identified from study data. A range of possible causes may produce bias [1]. The following text only describes the most important types; a detailed overview of various types of bias in different situations is presented by Feinstein [2].

Selection bias is caused by a distorted allocation of study participants to groups to be compared. As a result, systematic differences of characteristics may be present between groups and may lead to an unequal distribution of important confounders. Randomisation is the best method to avoid selec-

tion bias, as the groups formed do not differ systematically with regard to known and unknown confounders. In non-randomised studies, the effect of known confounders can be taken into account by employing multi-factorial methods (q.v. Section 1.22). However, the risk of a systematic difference between the groups due to unknown confounders remains.

Besides the comparability of the groups with regard to potential prognostic factors, equality of care (apart from the intervention to be investigated) and the equality of observation of study participants play an important role. *Performance bias* is systematic distortion by the different forms of care provided. A breach of equality of observation can lead to *detection bias*. Double-blinding is an effective protection against both performance and detection bias. In epidemiological studies, performance and detection bias are often summarised as information bias.

Protocol violations and study withdrawals can cause a systematic distortion of study results (*attrition bias*). To avoid attrition bias, the "intention-to-treat" principle can be applied; all randomised study participants are evaluated within the group to which they were assigned, independently of protocol deviations.

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 can lead to *spectrum bias* (q.v. also Section 2.3). 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 in only an apparent increase in survival times, due to non-comparable starting points (*lead time bias*). Increased survival times can also be feigned for diseases with a long pre-clinical phase (*length bias*). The conduct of a randomised trial to investigate the effectiveness of a screening technique can protect against these bias mechanisms (q.v. Section 2.4).

A common problem arising from the estimation of effects is a bias of results by measurement errors and misclassifications in the study data collected [3,4]. In practice, measurement errors can hardly be avoided, and it is known that non-differential measurement errors can also lead to bias when estimating effects. In the case of a simple linear regression model with a random classical measurement error in the explanatory variable, *dilution bias* occurs, i.e. an attenuation of the estimate towards the zero effect. In other models and more complex situations, bias in all directions is possible. For this reason, the size of potential measurement errors should be discussed in all studies and, if necessary, methods to adjust bias should be employed.

Missing values present a similar problem. Missing values not due to a random mechanism can also cause bias in a finding [5]. The possible causes and effects of missing values should therefore be

discussed on a case-by-case basis and, if necessary, statistical methods employed to account or compensate for bias.

Publication bias plays an important role in systematic reviews [6]. As significant results are more frequently published than non-significant results, a systematic bias of the common effect estimate occurs when published results are summarised. Graphical methods such as the funnel plot and/or statistical methods such as meta-regression are techniques for identifying and considering publication bias [7].

References

[1] Sackett DL. Bias in analytic research. J Chron Dis 1979; 32: 51-63.

[2] Feinstein AR. Clinical Epidemiology. The Architecture of Clinical Research. Philadelphia: WB Saunders Co.; 1985.

[3] Carroll RJ, Ruppert D, Stefanski LA. Measurement Error in Nonlinear Models. London: Chapman & Hall; 1995.

[4] Cheng C-L, van Ness JW. Statistical Regression with Measurement Error. London: Arnold; 1999.

[5] Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York: Wiley; 1987.

[6] Begg CB, Berlin JA. Publication bias: A problem in interpreting medical data (with discussion). J R Stat Soc A 1988; 151: 419-463.

[7] Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. BMJ 2001; 323: 101-105.

1.19 Evaluation of a difference

Various aspects need to be considered when presenting empirical evidence that certain groups differ with regard to a certain characteristic. It should be noted that the term “evidence” should not be understood in the mathematical sense of the term “proof”. 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 provide statistical evidence in support of a hypothesis. Significance tests are the typical methods used to provide this type of statistical evidence in medical research (q.v. Section 1.2). This level of argumentation should be distinguished from the evaluation of the clinical relevance of a difference (q.v. Section 1.3). The combination of both arguments in practice provides an adequate description of a difference on the basis of empirical data.

When employing a significance test to show a difference, there should be an a priori determination of the research question posed and, based on this question, a determination of the outcome variable,

the effect measure and the formulation of the statistical hypothesis. It is necessary to estimate the sample size before the start of study, so that the size of the study population is sufficient to detect a treatment difference. In addition to the above information, a statement on the clinically relevant difference as well as an estimate of the variability of the outcome measure should be provided for simple situations. For more complex designs and/or research questions, further information, for example on the correlation structure, recruitment scheme and estimation of drop-out numbers, is required [1,2].

Finally, the description of results should include the following details: a statement on the significance level and on the adequate confidence interval for the effect measure chosen (q.v. Section 1.1); a descriptive statement on further effect measures to expound different aspects of the results; and a discussion on the clinical relevance of the results based on the determination of patient-relevant target criteria (q.v. Sections 1.3 and 1.5).

References

[1] Desu MM, Raghavarao D. Sample Size Methodology. Boston: Academic Press, 1990.

[2] Bock J, Toutenburg H. Sample size determination in clinical research. In: Rao CR, Chakraborty R, editors. Handbook of Statistics Vol. 8. Amsterdam: Elsevier, 1991: 515-538.

1.20 Evaluation 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 [1]. To demonstrate equivalence, methods to investigate the equivalence hypothesis need to be employed [2]. It is important to understand that showing “exact equivalence”, e.g. the statement that the difference in mean values between two groups is exactly zero, is not possible with the aid of statistical methods. In practice, it is not evidence of exact equivalence that is required, but rather evidence 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. the determination of an equivalence range is necessary. To draw meaningful conclusions on equivalence, the research question and the resulting outcome variable, effect measure and statistical hypothesis formulation need to be determined a priori (similar to methods to demonstrate evidence of a difference; q.v. Section 1.19).

The equivalence range must be clearly defined in equivalence studies. 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 being referred to as a “non-inferiority hypothesis” [3].

As in superiority studies, it is necessary to calculate the required sample size in equivalence studies before the start of the study. The appropriate method depends on the exact hypothesis and the method of analysis chosen [4]. For this purpose, specifically developed methods should be applied to analyse data in equivalence studies. The "confidence interval inclusion method" is a frequently employed technique. If the confidence interval calculated lies completely within the previously defined equivalence range, then this is classified as evidence of equivalence. To maintain the level of α=0.05, it is sufficient to calculate a 90% confidence interval [5].

In comparison with superiority studies, equivalence studies have specific methodological problems. On the one hand, it is often difficult to provide meaningful definitions of equivalence ranges [6]; on the other, the usual criteria for study designs, such as randomisation and double-blinding, no longer offer sufficient protection from bias [7]. Even without knowledge of the treatment group, it is possible, for example, to shift the treatment difference between groups to zero and hence in the direction of the desired alternative hypothesis. Moreover, "the intention-to-treat" principle should be applied carefully, as its inadequate application may feign false equivalence [2]. For this reason, particular caution is necessary when assessing equivalence studies.

References

[1] Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995; 311: 485.

[2] Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996; 313: 36-39.

[3] D'Agostino RB, Massaro JM, Sullivan KM. Non-inferiority trials: Design concepts and issues - the encounters of academic consultants in statistics. Stat Med 2003; 22: 169-186.

[4] Roebruck P, Elze M, Hauschke D, Leverkus F, Kieser M. Literaturübersicht zur Fallzahlplanung für Äquivalenzprobleme *(Literature review of sample size estimations for equivalence problems)*. Inform Biom Epidemiol Med Biol 1997; 28: 51-63.

[5] ICH E9 Expert Working Group. ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials. Stat Med 1999; 18: 1905-1942.

[6] Lange S, Freitag G. Choice of delta: requirements and reality - results of a systematic review. Biometrical J 2005; 47 (in press).

[7] Senn S. Inherent difficulties with active control equivalence studies. Stat Med 1993; 12: 2367-2375.

1.21 Meta-analyses

Meta-analysis is the statistical summary of the results of several studies in a systematic review (q.v. Section 2.7) [1]. Before performing a meta-analysis, it should first be considered whether the sum-

mary of the studies in question is at all meaningful. Comparability of the studies with regard to the research questions posed should exist and the heterogeneity of study results should be investigated [2]. For this purpose, specific new statistical methods are available, such as the calculation of the I^2-measure [3,4]. If the heterogeneity of the studies is too great, a statistical summary of study results may not be meaningful [5]. The choice of effect measure also plays a role in this context. The choice of a particular measure may lead to a great heterogeneity of studies, whereas another measure may not. For binary data, relative effect measures are often more stable than absolute ones, as they do not depend so heavily on the baseline risk [6]. In such cases, the data analysis should be conducted with relative effect measures. Absolute measures should be derived from relative ones for the descriptive presentation of data.

References

[1] Cook DJ, Sackett DL, Spitzer WO. Methodological guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. J Clin Epidemiol 1995; 48: 167-171.

[2] Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. Stat Med 2002; 21: 1503-1511.

[3] Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21: 1539-1558.

[4] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327: 557-560.

[5] Deeks JJ, Higgins JPT, Altman DG. Analysing and presenting results. In: Alderson P, Green A, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.2 [updated March 2004]; Section 8. In: The Cochrane Library, Issue 1, 2004. Chichester: Wiley; 2004.

[6] 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: 72-76.

1.22 Adjustment principles and multi-factorial methods

Multi-factorial methods that enable the effect of confounders to be compensated play a key role in non-randomised studies (q.v. Section 1.18) [1]. Studies with several treatment groups are a further important field of application for these methods [2]. The description of results obtained with multi-factorial methods is unfortunately often insufficient in the medical literature [3,4]. To be able to assess the quality of such an analysis, the description of essential aspects of the statistical model formation is necessary [5,6], as well as details on the quality of the model (goodness of fit) [7].

The most relevant information for this purpose is:

- A clear description and an a priori determination of the outcome variable and all explanatory variables;
- Information on the measurement scale and on 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 for all explanatory variables.

Inadequate description of the results obtained with multi-factorial methods is especially critical if, as a result of the (unclearly described) statistical modelling, a shift in effects to the "desired" range occurs that 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 [1,8,9].

References

[1] Katz MH. Multivariable analysis: A primer for readers of medical research. N Engl J Med 2003; 138: 644-650.

[2] McAlister A, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: A systematic review. JAMA 2003; 289: 2545-2553.

[3] Bender R, Grouven U. Logistic regression models used in medical research are poorly presented (Letter). BMJ 1996; 313: 628.

[4] Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: A cross-sectional survey. Ann Intern Med 2002; 136: 122-126.

[5] Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15: 361-387.

[6] Royston P. A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. Stat Med 2000; 19: 1831-1847.

[7] Hosmer DW, Lemeshow S. The importance of assessing the fit of logistic regression models: A case study. Am J Public Health 1991; 81: 1630-1635.

[8] Bender R, Grouven U. Ordinal logistic regression in medical research. J R Coll Physic London 1997; 31: 546-551.

[9] 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: 979-985.

1.23 Evaluation of qualitative studies

Qualitative research methods are used to explore and understand subjective experiences, individual actions and the social world [1-4]. The Institute aims to use qualitative studies to generate hypotheses and to assist in the interpretation of data.

Quantitative research works primarily with numbers of different dimensions and is standardised by strong recommendations, although personal and social experiences may also be taken into account. Conversely, in qualitative research the emphasis is on subjective data [1].

The Institute's main task in the evaluation of social science studies is to determine whether the study design, study quality and reliability of results were appropriate for the research question posed. The Institute will expect the same standards of both quantitative and qualitative studies. The assessment of the studies included follows the respective scientific criteria. A weaker general consensus exists on the validity of criteria for the conduct, evaluation and synthesis of qualitative studies than for other research fields [1-5].

The Institute will use defined criteria for its assessment of qualitative studies using evidence-based research criteria to evaluate these studies within the framework of systematic reviews and health technology assessments (HTAs) [1].

References

[1] Dixon-Woods M et al. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.

[2] Murphy E et al. Qualitative research methods in health technology assessment: a review of the literature. NHS R&D HTA Programme. Health Technology Assessment, 1998; Vol 2: No 16.

[3] McClelland S. Qualitative research methods: Their role in health services research. NHS Management Briefing, National Electronic Library for Health. 00&04, February 2001. http://libraries.nelh.nhs.uk/healthmanagement. Access on 22 October 2004.

[4] Harden A et al. Applying systematic review methods to studies of people's views: an example from public health research. J Epidemiol Community Health 2004; 58: 794-800.

[5] Thomas J et al. Integrating qualitative research with trials in systematic reviews. BMJ 2004; 328: 1010-1012.

1.24 Consultation techniques

According to the research questions posed and tasks assigned, it may be necessary to use a variety of consultation techniques.

Against the background of time limitations, rapid appraisal methods (e.g. focus groups) will generally be employed. However, these methods have various strengths and weaknesses; they can vary strongly in their representativeness and validity. When employing these methods, the Institute will therefore follow existing quality standards for their selection, application and analysis. Ethical aspects will particularly be taken into account where techniques are applied that may have detrimental effects on participants (e.g. focus groups).

Consultation techniques will be preceded by a literature search for relevant qualitative data. The results of rapid appraisal methods will be interpreted in the context of the available results of relevant and more detailed qualitative studies.

The Institute may apply the following consultation techniques:

- Key informant interviews [1];
- Focus groups [2-4];
- Group interviews, group meetings and consultations [5-7];
- Surveys and polling (including online polling and feedback mechanisms);
- Occasional use of consensus techniques, e.g. Delphi techniques [8] and participatory evaluation [9].

The Institute may also develop health impact assessments (HIAs) using both qualitative and quantitative methods [10]. The basis for conducting these assessments is a clear and transparent procedure, also with regard to the potential impact of decisions on equity and social justice.

The techniques for gaining people's views and identifying attitudes vary greatly in their representativeness. The Institute will therefore need to take care to ensure that the views of disadvantaged groups are adequately considered in this process.

One of the objectives of the Department of Patient Information and Research is to promote health and scientific literacy in the population. On the one hand, the aim is to enhance understanding of scientific terminology pertaining to health issues and evidence-based aspects of the health care system, and on the other, to arouse public interest in the Institute's work. The public is to be actively involved in this process.

For this purpose, the Department may use and further develop the methodology of consultation techniques, consensus building and public decision-making [11], using the Institute's website and, where appropriate, additional multi-media strategies.

Population level techniques, such as online surveys, citizens' juries [11] and public discussions are widely used in resource allocation decisions in the health care system [11,12]. Citizens' juries have been found to be particularly effective in the investigation of complex issues. Some of these techniques may be adapted by the Institute to achieve the aims described above.

References

[1] USAID Center for Development Information and Evaluation. Conducting key informant interviews. Washington: USAID Center for Development Information and Evaluation; 1996 No 2. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs541.pdf. Access on 22 October 2004.

[2] USAID Center for Development Information and Evaluation. Conducting focus group interviews. Washington: USAID Center for Development Information and Evaluation; 1996 No 10. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnaby233.pdf. Access on 22 October 2004.

[3] Aylward P. Conducting research with focus groups. Staff Development Session. Flinders University of South Australia. http://www.flinders.edu.au/staffdev/courses/research/resources/Focusgroups.pdf. Access on 22 October 2004.

[4] Dixon-Woods M et al. Integrative approaches to qualitative and quantitative evidence. London: NHS Health Development Agency; 2004.

[5] National Resource Centre for Consumer Participation in Health. Feedback, participation and diversity: A literature review. Canberra: Commonwealth of Australia, 2000. http://www.participateinhealth.org.au. Access on 22 October 2004.

[6] National Health and Medical Research Council, Consumers Health Forum of Australia. Statement on consumer and community participation in research. Canberra: Commonwealth of Australia, 2002. http://www.nhmrc.gov.au/publications/pdf/r22.pdf. Access on 22 October 2004.

[7] National Resource Centre for Consumer Participation in Health. Methods and models of consumer participation. Melbourne: National Resource Centre for Consumer Participation in Health, 2004. http://www.participateinhealth.org.au. Access on 22 October 2004.

[8] Liberati A, Sheldon TA, Banta HD et al. Eur-Assess project subgroup report on methodology: Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care, 1997; 13: 186-219.

[9] USAID Center for Development Information and Evaluation. Conducting a participatory evaluation. Washington: USAID Center for Development Information and Evaluation; 1996 No 1. http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs539.pdf. Access on 22 October 2004.

[10] Barnes R, Scott-Samuel A. Health Impact Assessment: A ten minute guide. Liverpool: International Health Impact Assessment Consortium, 2000. http://www.ihia.org.uk/hiaguide.html. Access on 25 October 2004.

[11] Ryan M et al. Eliciting public preferences for healthcare: A systematic review of techniques. Health Technology Assessment, 2001; Vol 5 No 5.

[12] Hicks N, Harford J. Summary report on consumer participation in resource allocation. Melbourne: National Resource Centre for Consumer Participation in Health, 2000.

1.25 Peer review

To ensure the quality of the Institute's reports and other products, statements from experts, consumers, patients and relatives will be obtained, depending on the respective research question and product.

The evidence base for peer reviews has increased. However, only a few meaningful studies are available that investigate the effectiveness of specific methods used in peer reviews [1,2]. In particular, there is a lack of sufficiently valid intervention studies. According to the studies available [3], the conventional procedures employed in medical journals [2,3], including review by consumers and patients [4], may have little demonstrable effect.

There is insufficient sound scientific knowledge of the best procedure to be followed in a peer review. This includes the question of the number of reviewers required and whether a peer review should be open (naming authors and reviewers) or closed. The respective project managers or department heads will decide whether one or more reviewers are to be involved in a report. In accordance with its claim to transparency, the Institute will work with an open peer-review system and develop a questionnaire for reviewers to gather information on reviewers' interests and on how many reports they can review. The data obtained will be stored in a database [4].

Any medical publisher who releases many articles per year needs to develop a system for editorial rating that enables the quality of the peer reviewers' contributions to be evaluated [5]. The Institute will develop a scoring system for evaluations based on validated systems [5].

All peer reviewers will be sent copies of the other peer review reports as well as the final published report.

The Institute will evaluate and update its peer-review system with regard to new developments in the field of peer-review methodology and evidence. Besides employing the traditional methods used in journals (individual assessments of manuscripts), the departments may also use a variety of rapid appraisal methods (q.v. Section 1.24).

References

[1] Rennie D. Editorial peer review: Its development and rationale. In Godlee F, Jefferson T, editors. Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.

[2] Fletcher RH, Fletcher SW. The effectiveness of journal peer review. In Godlee F, Jefferson T, editors. Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.

[3] Jefferson TO, Alderson P, Davidoff F, Wager E. Editorial peer-review for improving the quality of reports of biomedical studies. The Cochrane Database of Methodology Reviews 2001, Issue 3.

[4] Bastian H. Non-peer review: consumer involvement in research review. In Godlee F, Jefferson T, editors. Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 248-262.

[5] Smith J. How to set up a peer review system. In Godlee F, Jefferson T, editors. Peer Review in Health Sciences (2nd ed). London: BMJ Books, 2003; 151-163.

2. Specific evaluation of medical and health care issues

The extraction and evaluation of data from relevant publications found in literature searches (q.v. Section 4.8) will be conducted and documented in a structured manner. For the documentation process, it is planned to use data extraction forms, which have yet to be developed. If these forms are not used, this must be justified and agreed to by the project manager. Alternative forms will then be developed, agreed to and applied.

Data extraction forms will be used, amongst other things, for:

- HTA reports,
- Systematic reviews,
- Guidelines,
- Intervention studies,
- Diagnostic studies to evaluate test quality criteria,
- Prognostic studies.

2.1 Evaluation of effects in medicine and health care

To date, no uniform and generally accepted definition of the term “benefit” or “utility” (of medical measures) exists on a national or international level.

The general assessment as to which medical measures are necessary and beneficial for the population depends not only on their objective effectiveness, but also on socially accepted values and economic conditions, i.e. the resources of a state. If one assumes that health care resources are in principle not limited by law and that the term “necessary” refers both to healing a disease and to preventing potential future undesired events in the sick and the healthy, then the decision on employing a measure is focused on weighing its potentially beneficial and harmful results. Furthermore, the uncertainty of assumptions regarding the occurrence of these results needs to be taken into account. When assessing potential beneficial and harmful effects, patient-relevant endpoints and not their surrogates (i.e. disease-relevant aspects) should primarily be taken into consideration. Moreover, it is essential that meaningful clinical studies estimating the effect size of these events are available. In these studies, the probability of a positive effect on health or disease is contrasted with the diagnostic or therapeutic investment required and the potential damage that any medical measure may entail.

Decisions are therefore dependent on subjective and individual issues, which mean that evaluations must be made on a case-by-case basis. Each of the Institute's report plans will reflect the meaning of the term "benefit" in relation to the specific research question under consideration.

2.2 Pharmaceutical and non-pharmaceutical interventions

The objective of the evaluation of a study on a pharmaceutical or non-pharmaceutical intervention is to show with what certainty an effect or the absence of an effect can be derived from the study findings (certainty of results). Moreover, it is necessary to describe whether and to what extent the study results are transferable to local conditions (e.g. the population affected, type of care provided) and to assess local particularities that could have an influence on the results or on their interpretation. From this point of view, studies in which actual health care conditions are portrayed as accurately as possible are particularly relevant; however, the criteria described below on the certainty of results should not be disregarded.

Certainty of results is essentially influenced by four components:

- The study design;
- The internal validity (dependent on the study design);
- The consistency of the results of several studies of equal quality;
- The disease towards which a therapeutic or preventive medical intervention is targeted.

The criteria for the assessment of internal validity are described in detail in Section 1 and are applied correspondingly in the evaluation of studies on pharmaceutical and non-pharmaceutical interventions.

The study design has considerable influence on the certainty of results. For example, a causal relationship between intervention and effect cannot be shown with prospective or retrospective epidemiological studies, whereas an experimental study design is, in principle, suitable for this purpose [1]; that is, if factors influencing the results can be eliminated completely or almost completely. For this reason, an RCT represents the gold standard for the evaluation of pharmaceutical and non-pharmaceutical interventions [2].

To assess effectiveness, the Institute will therefore use non-randomised intervention studies or epidemiological studies only in exceptional cases. These exceptions must be justified. Reasons for exception are e.g. the non-feasibility of an RCT or the fact that other study types may also provide sufficient certainty of results for the research question posed. In this regard, the aforementioned

point (consistency of results in various studies of equal quality) is relevant. For diseases that without intervention would be fatal within a short period, the availability of several consistent case reports can provide sufficient certainty of results as to whether a particular intervention may prevent this otherwise inevitable course [3].

As part of the report plan (q.v. Section 4.4), the Institute will therefore determine beforehand which study types can, on the basis of the research questions posed, theoretically be regarded as providing sufficient certainty of results (with high internal validity). Studies not complying with these quality standards (q.v. Sections 1.9 and 1.10) will not be primarily considered in the evaluation process.

Finally, the transferability of study results must be verified in a separate process that is independent of the study design and quality.

References

[1] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet 2002; 359: 57-61.

[2] Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. The Cochrane Database of Methodology Reviews 2002, Issue 4. Art. No.: MR000012. DOI: 10.1002/14651858.MR000012.

[3] Liberati A, Sheldon TA, Banta HD et al. Eur-Assess project subgroup report on methodology: Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care 1997; 13: 186-219.

2.3 Diagnostic tests

To date, the methodology for evaluation of diagnostic tests has not achieved a comparable penetration level to that for evaluation of therapeutic procedures [1]. For this reason, it can be assumed that the information on the evaluation of diagnostic tests presented below (q.v. Sections 2. 3.1 to 2.3.3) will only be available completely in exceptional cases.

As studies on diagnostic tests are conducted with different objectives and, depending on the objectives set, not all information is relevant, the essential basis of evaluation is the precise formulation of a research question. The formulation of a research question or the process leading to the final formulation follows the principles described in Section 4.5.

2.3.1 General aspects

The following information in particular is relevant to evaluate a diagnostic test:

- Clear definition of the disease(s) to be diagnosed/detected or, more generally, the health status to be detected (e.g. physical and mental fitness).
- Information on the prevalence of the disease(s) to be diagnosed/detected in the population and sub-populations to be investigated.
- Unambiguous definition of the "gold standard", i.e. of the method by which the disease to be detected (or the health status to be detected), can be unambiguously identified in a generally accepted manner.
- Exact description of the diagnostic test, including details on the material and human resources needed in training for the test, as well as for its execution and evaluation.
- Description of the risks involved in the application of the diagnostic test, and the acceptance and reasonability both for patients and medical personnel, as well as for the general public (e.g. environmental risks).
- Information on any further consequences to be expected from the respective findings (e.g. further diagnostics, therapy, non-therapeutic interventions, monitoring, lifestyle, basis for informed decisions) describing to what extent these consequences constitute a benefit for the patient.
- Details on alternative diagnostic tests and, if necessary, description of the advantages of the new tests over the conventional ones.

2.3.2 Test quality criteria and test characteristics

The evaluation criteria for diagnostic tests are as follows:

I. Technical prerequisites

- Information on diagnostic accuracy.
- Information on diagnostic sensitivity and specificity.
- Information on reproducibility (reliability).
- Variability:

- Intra-test/rater variability,
- Inter-test/rater variability,
- Intra-patient variability (short and long-term).

- If necessary, information on consistency of measurement values with already established standard tests.
- Information on possible confounders (most notably, systematic distortion effects).

II. Discrimination ability

- Information on diagnostic sensitivity and specificity or, alternatively, information on likelihood ratios. For quantitative methods: presentation of ROC[i] curves with specification of (an) appropriate cutpoint(s) and the rationale for its/their selection (weighting of sensitivity and specificity).

III. Prediction

- Information on predictive values (for quantitative methods: information on predictive values for [a] selected cutpoint[s]).

2.3.3 Evidence of effectiveness

- Evidence that the use of a diagnostic test leads to an improvement in outcome for the patient.

For studies on technical prerequisites (q.v. Section 2.3.2.I), it should be ensured that the corresponding parameters are determined under everyday clinical conditions and in the situation of application. For example, information on the reproducibility of a diagnostic test is usually insufficient if this test has only been tested in healthy persons.

In principle, there are two types of procedures for studies on discrimination ability (q.v. Section 2.3.2.II). Firstly, the application of the diagnostic test in selected persons with a known disease status; secondly, the application of the test in unselected persons with an unknown disease status (2-4). If the required results have been achieved, the first procedure is generally the prerequisite for the conduct of a further (usually more extensive) study following the second procedure. Studies based on the first procedure typically provide an overoptimistic estimate of discriminatory ability [5,6]; this factor should be taken into account in any evaluation.

[i]Receiver operating characteristics.

The following basic methodological principles should be considered in the assessment of studies evaluating discrimination ability (q.v. Section 2.3.2II) and predictability (q.v. Section 2.3.2III) of diagnostic tests:

- Clear formulation of a research question and of the respective study planning; this includes sample size estimations, which can be based on the desired precision of an estimate (width of the confidence interval) and/or the evidence (by means of inferential statistics) of exceeding a minimum level.
- Conduct of the diagnostic test in a population of patients or clinically healthy persons to whom the test is to be applied in the future (suitable patient spectrum and prevention of spectrum bias).
- Blinded assessments (independent of each other) of the diagnostic test to be evaluated and of the gold standard (mutual blinding).
- Confirmation of diagnosis (gold standard) or the type of confirmation of diagnosis (existence of different gold standards) should not be made dependent on the result of the diagnostic test to be evaluated (danger of verification bias). If confirmation of diagnosis cannot be performed in all patients, the selection of patients should be random.[j]
- Appropriate consideration of patients with unclear, non-interpretable or intermediate test results (not just simple exclusion of patients).
- If the diagnostic test to be evaluated is (or is to be) embedded in a diagnostic strategy, an isolated assessment of this test is often not meaningful (problem of the dependence of test quality criteria on the combination of diagnostic tests applied).
- If the diagnostic test to be assessed is a constituent of the gold standard, particular methodological problems requiring detailed discussion and consideration may arise.

Experience shows that the above principles are frequently lacking in published diagnostic studies. It is therefore necessary to describe exactly the methodological deficits of the respective studies, and consequently their results, in order to take this situation into account and be able to make any kind of statement at all. Caution is also necessary if a (statistical) summary of individual results (in terms of a meta-analysis) is planned.

[j]However, this does not solve the problem entirely, especially in situations with low a priori probabilities [7].

Similar recommendations for studies on diagnostic tests have been published in analogy to the CONSORT statement on therapeutic studies, thereby aiming to achieve a uniform and comprehensive presentation of the topic [8].

Whiting et al. have compiled a checklist for the quality assessment of diagnostic studies in systematic reviews [9].

Studies on the evidence of the effectiveness (Section 2.3.3) of diagnostic tests[k] can be planned as a comparison between patients to whom the diagnostic test is applied and patients to whom this test is not applied. The same requirements as formulated in Section 2.4.4 apply in essence to the evaluation of such studies. One disadvantage of such studies is that the value of the diagnostic findings cannot be separated from the resulting consequences, i.e. for a negative outcome it cannot be distinguished whether the diagnostic information is insufficient or whether, for example, the therapy (for those with a pathological test result) is ineffective.

As an alternative to assessing the conduct of the test, the disclosure of the test results can be investigated, i.e. persons for whom the test result is known can be compared with those for whom the result remains blinded [4]. Such a procedure offers the advantage of enabling the evaluation of the natural course of the disease in persons with a positive test result.

In another design option, the diagnostic test to be evaluated is applied to all patients in a therapeutic setting (independent of the treatment group), and the result remains blinded for all patients throughout the whole trial. In this type of study, the interaction between diagnostic information and therapeutic effectiveness can be assessed, i.e. whether patients experience different therapeutic benefits, depending on the result of the diagnostic test [3].

References

[1] Knottnerus JA, van Weel C, Muris JWM. Evidence base of clinical diagnosis. Evaluation of diagnostic procedures. BMJ 2002; 324: 477-480.

[2] Köbberling J, Trampisch HJ, Windeler J. Memorandum for the Evaluation of Diagnostic Measures. J Clin Chem Clin Biochem 1990; 28: 873-879.

[3] Richter K, Lange S. Methoden der Diagnoseevaluierung *(Methods of diagnostic evaluation)*. Internist 1997; 38: 325-336.

[4] Sackett DL, Haynes RB. Evidence base of clinical diagnosis. The architecture of diagnostic research. BMJ 2002; 324: 539-541.

[k]It should be noted that studies on evidence of effectiveness of diagnostic techniques as described here are seldom found in the literature.

[5] Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JHP, Bossuyt PMM. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999; 282: 1061-1066.

[6] Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess 2004; 8: iii, 1-234.

[7] Rückmann A, Windeler J. Selektionsbias bei der Schätzung der Sensitivität von Screeningmaßnahmen *(Selection bias in the estimation of the sensitivity of screening measures)*. In: Trampisch HJ, Lange S (Hrsg.). Medizinische Forschung – Ärztliches Handeln. München: MMV Medizin Verlag; 1995. p. 227-231.

[8] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC; Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. Radiology 2003; 226: 24-28.

[9] 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.

2.4 Screening

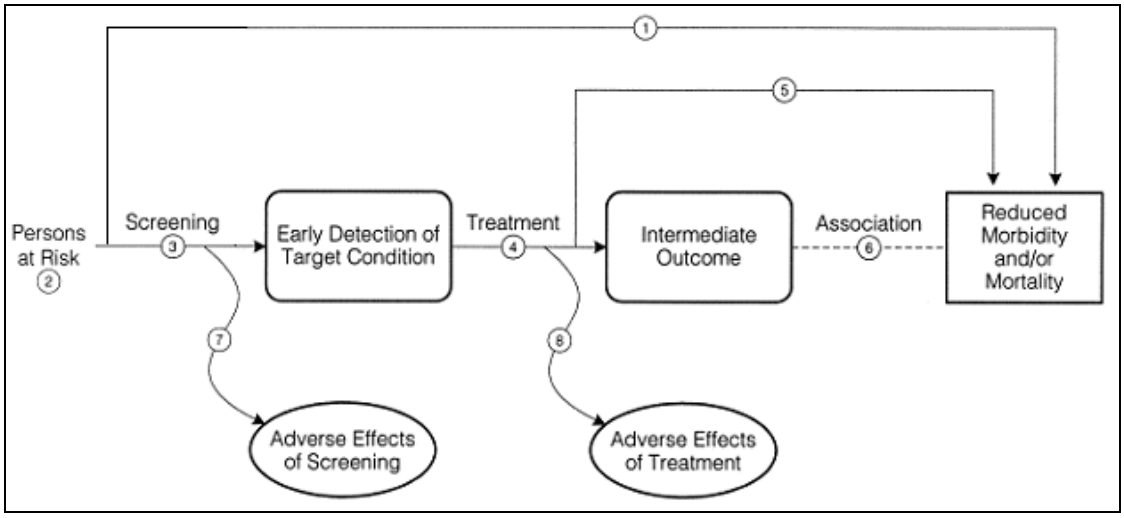
The essential basis for the evaluation of a screening test is to formulate a precise research question, as screening programmes are conducted with different objectives and are composed of different modules, which can be examined either as a whole or in part. The formulation of a research question or the process leading to the final formulation of a question follows the principles stipulated in Section 4.5.

The evaluation of a screening test follows criteria that have already been established and published, for example by the UK National Screening Committee (UK NSC, [1]) and the US Preventive Services Task Force (US PSTF, [2]).

These evaluation criteria comprise: (1) the disease to be detected; (2) the (diagnostic) screening test to be employed; (3) the type of therapy in the case of a positive (pathological) result or a different consequence derived from a positive result; and (4) the screening programme as a whole.

When positive results are present, the evaluation should discriminate between programmes resulting in therapeutic measures and those resulting in other, non-therapeutic measures. Furthermore, a distinction should be made between (a) situations in which direct evidence for the effectiveness of the screening programme exists and (b) situations in which the evidence is derived indirectly through conclusions by analogy ([a] comparison of persons with and without screening with a patient-relevant target criterion in a study, Arrow 1 in Figure 1; [b] several screening modules are assessed in different studies, e.g. Arrows 3, 4 and 6 in Figure 1).

Figure 1: Screening chain



Modified according to [2].

2.4.1 Disease to be evaluated

It is to be evaluated whether the disease in question is an important health problem, whereby the evaluation can refer to different indicators, e.g. to the frequency, severity or cost of a disease (or to different levels, e.g. population or individual levels).

This requires an exact knowledge of the epidemiology and natural course of the disease. Typical data sources include epidemiological cross-sectional, register and cohort studies. In exceptional cases, data from case series and economic studies are used.

2.4.2 Screening tests

The general requirements for the assessment of a diagnostic test apply as formulated in Sections 2.3.1 and 2.3.2. However, due to the special ethical implications, higher demands are as a rule to be made on the test quality criteria and the quality of the underlying studies [3]. Moreover, the test should be easy to handle, and it is to be assessed whether, in the case of a positive test result, a generally accepted strategy is available for further diagnostic clarification (gold standard) and for other available alternatives.

2.4.3 Therapy

For patients with a positive test result and, if available, with confirmation of diagnosis following further diagnostic tests (gold standard), it needs to be investigated whether an effective treatment or intervention exists (q.v. Section 2.2 for the respective evaluation criteria). In addition, it needs to be assessed whether evidence is available showing that early treatment leads to better results than late treatment.

For screening programmes not resulting in direct therapeutic measures following a positive test result, it should be evaluated whether the information gained from the positive result is associated with a different (non-therapeutic) benefit, e.g. of the kind that allows affected persons to make better informed decisions (e.g. prenatal screening for Down syndrome, screening for genetic carriers of incurable diseases). In these cases it may be meaningful to apply decision analysis methods.

2.4.4 Screening Programmes

Ideally, there is available evidence that the screening programme as a whole is effective in reducing morbidity and/or mortality. The criteria formulated in Section 2.2 are used to evaluate the respective studies. In particular, evidence from non-randomised studies on the assessment of screening programmes needs to be evaluated critically, as specific bias mechanisms such as lead time bias or length bias may occur (q.v. Section 1.18)

If direct evidence for the effectiveness of a screening programme is only available for individual screening modules, then in addition to the assessment of individual modules, an evaluation of their coherence and consistency should be made. Coherence in this regard means that the modules form a comprehensible model; consistency means that different study findings contribute to coherence under different conditions [2].

The screening programme should achieve a net benefit, i.e. the benefit gained from a screening programme should exceed the potential physical or mental damage caused by the screening test or by the subsequent diagnostic measures and/or therapy (q.v. Section 1.7 for the evaluation of adverse effects of an intervention).

If the screening programme or its modules were assessed in the setting in which the programme is to be implemented, then it needs to be reviewed whether evidence is available showing that the results can be generalised or transferred.

References

[1] UK National Screening Committee. Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. http://www.nsc.nhs.uk/pdfs/criteria.pdf. Access on 28 October 2004.

[2] Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, Atkins D; Methods Work Group, Third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med 2001; 20(Suppl3): 21-35.

[3] Ewart RM. Primum non nocere and the quality of evidence: rethinking the ethics of screening. J Am Board Fam Pract 2000; 13: 188-196.

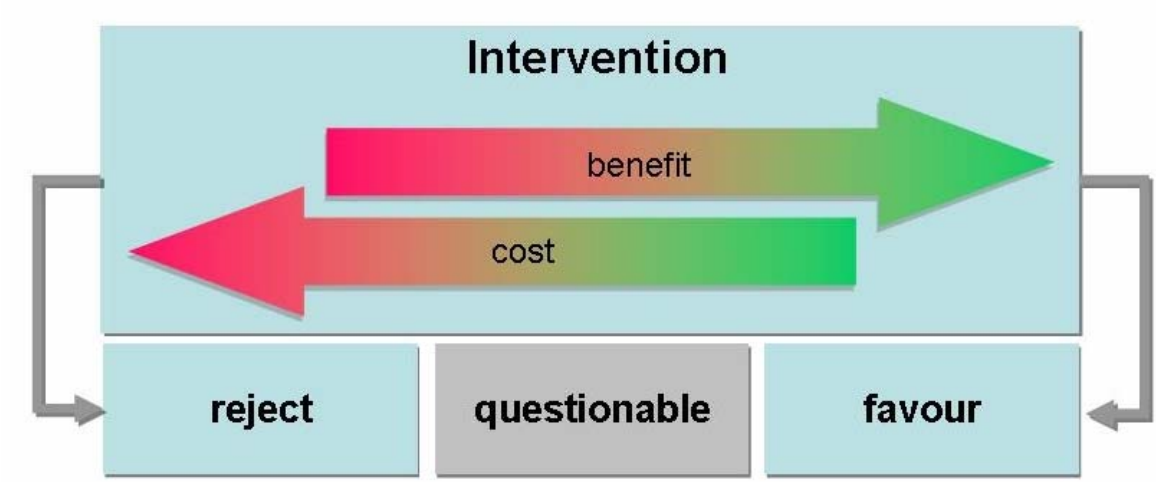
2.5 Health economics

2.5.1 Contents

The Health Economics Department fulfils direct and indirect assignments within the Institute; direct assignments with regard to direct commissioning by the Federal Joint Committee or the Federal Ministry of Health, and indirect assignments with regard to research questions from areas related to economic aspects covered by other departments. Furthermore, the department can develop its own research questions for scientific projects to be conducted within the department or Institute after consultation with the Steering Committee.

In order to meet health care needs, health economics aims to prevent inappropriate or under- or overprovision of care. These shortcomings are important aspects of quality deficits in terms of an imbalance in cost-benefit allocations. Cost-benefit evaluations can hereby refer to different aspects in the evaluation of current and future medical or health policy interventions.

Figure 2: Cost-benefit and decision spectrum



In situations of low benefit and high cost or those of high benefit and low cost, a clear evaluation is possible. However, this is not usually the case [1] and further conditions must be considered, which makes health-related decision-making more complex.

In contrast to other economic sectors, ethical questions are especially relevant in this context, for example with regard to weighing the principle of social solidarity against the needs of the individual. It is internationally accepted that, ultimately, medical need and effectiveness are the primary consideration rather than cost [2-6]. The prevailing social norms and their ethical foundation according to which resources are allocated in the health care system are therefore more relevant than expenditure.

2.5.2 Outcome measures

The assessment of cost-benefit ratios is dependent on the social objectives pursued, such as the maximisation of average life expectancy in the population as a whole versus maximisation of average life expectancy among individual social groups with a lower (or presumed lower) life expectancy. The same applies to the pursuit of strategies aimed at improving the quality of life or the comparison of life expectancy with quality of life.

First of all, the respective outcome measure, towards which further assessments can be targeted, needs to be determined based on the research question posed. If several outcome measures are defined, the effects on the respective outcome measures need to be described in order to enable the consideration of different aspects of outcome measures, especially with regard to ethical issues.

It has to be noted that various proposals published in the 1990s suggesting the measurement of adequate cost-benefit ratios by quality-adjusted life years (QALY) have not become generally accepted and have not led to subsequent standardised recommendations. However, €50 000 per QALY gained are still frequently viewed as acceptable or cost-effective. The UK health authorities have presented a matrix model that describes a cost-benefit evaluation for different evidence levels, underlining the fact that a pure cost-benefit analysis reflects only one dimension of the quality of care. A combined view of evidence-based medicine and cost-benefit analysis therefore seems appropriate in the allocation of resources and the prioritisation of medical measures, if further factors such as the principles of solidarity and subsidiarity are to be considered in an appropriate and balanced way.

2.5.3 Methodology

For the consultation and valid interpretation of study results and publications in the field of health economics, the defined standards should suffice that are generally all directed towards transparency, comparability and quality [7]. In recent years, different evaluation systems for the planning and evaluation of studies have been established and published by:

- The Commonwealth of Australia (1995);
- The Ontario Ministry of Health (1994);
- The Canadian Co-ordinating Office for Health Technology Assessment (1994);
- The Task Force on Principles of Economic Analysis of Health Care Technology (1996);
- The US Public Health Service Panel on Cost-Effectiveness in Health and Medicine [8];
- The Hanover Consensus Group recommendations for evaluations in health economics [9], which closely follow the British Medical Journal checklist compiled by Drummond and Jefferson [10].

One of the main initial tasks of the Health Economics Department will be to evaluate the various instruments available in order to formulate, publish and bindingly apply the Institute's own standards.

When conducting cost analyses, the different types of estimated intervention costs are to be transparently presented. A distinction is made between direct and indirect costs, fixed and variable costs, and overhead and marginal costs. Conversely, the calculation and comparison of opportunity costs is, in practice, difficult, or costly to realise and describe [11]. The same applies to the calcula-

tion of indirect costs, which needs to be transparent and comprehensible and can be based on the human capital approach or the friction costs method. In addition to direct and indirect costs, so-called intangible costs describe the drawbacks of illness such as pain, depression or loss of quality of life, and cannot or only indirectly be quantified in monetary terms.

In an analysis of benefits, health-related quality of life is assessed. This shows how satisfied people are with their physical, mental and social health status and how they evaluate these three aspects of their health [12]. So-called utility-theoretic measurement methods or psychometric techniques are used for this purpose; whereas the former represent a rather general approach, the latter are more concrete with regard to perceptions as to what quality of life consists of. By and large, the discussion as to which method is best suited for measurement of quality of life is not yet concluded. This also applies to the methodological evaluation of validity and reliability where it is important to assess future empirical research critically.

The following analysis methods are regarded as most relevant in the evaluation of costs and benefits: cost-minimisation analysis (CMA), cost-effectiveness analysis (CEA), cost-utility analysis (CUA) and, in the narrower sense, cost-benefit analysis (CBA).

Cost analysis is conducted in a similar fashion for all methods; the evaluation of benefits is, however, made differently.

In a CMA it is assumed that different cost situations are produced by equivalent benefits (of alternative treatments); therefore this type of analysis should naturally be assessed critically.

In a CEA, the benefits are measured in natural units, resulting in a cost-benefit ratio. Comparability with other interventions is only possible if uniform benefit or effectiveness parameters are used. The choice of appropriate effectiveness parameters is often difficult and should be orientated towards defined patient-relevant endpoints; this normally requires large RCTs and epidemiological long-term studies. As a comparison of incremental costs and benefits of different interventions is often required, incremental analysis is performed to determine the ratio between these costs and benefits.

In a CBA, the benefits are quantified in monetary units, for example as a savings effect. The frequently applied conversion of quality of life and life expectancy into monetary units, for example by way of “willingness to pay” (WTP), should in principle be viewed as problematic.

In a CUA [13], costs are mostly contrasted with QALYs gained. This analysis enables the comparison of different intervention strategies and is especially suitable for questions referring to quality of life, mortality and morbidity. According to the recommendations of the US Panel on Cost Effec-

tiveness in Health and Medicine, a health economics study should as a matter of principle be conducted as a CUA [8].

References

[1] Haycox A, Bagust A, Walley T. Clinical guidelines – the hidden costs. BMJ 1999; 318: 391-393.

[2] Pellegrino ED. The commodification of medical and health care: the moral consequences of a paradigm shift from a professional to a market ethic. J Med Phil 1999; 3: 243-266.

[3] Schwartz P. Medical ethics under managed care. Int J Fert 1996; 2: 124-128.

[4] Ulsenheimer K. Qualitätssicherung und Risk-Management im Spannungsverhältnis zwischen Kostendruck und medizinischem Standard *(Quality control and risk management in the area of conflict between cost pressures and medical standards)*. MedR 1995; 11: 438-442.

[5] World Medical Association. Statement on professional responsibility for standards of medical care. October 1996.

[6] Westhofen M. Operative Hochleistungsmedizin. Handlungszwang zwischen ärztlicher Ethik, wirtschaftlichem Erfolg und Qualitätskontrolle. *(Operational high-performance medicine. The enforcement of action regarding medical ethics, economic success and quality control).* In: Brudermüller G (Hrsg). Angewandte Ethik und Medizin. Würzburg: Schriften des Instituts für Angewandte Ethik e.V. 1999, Bd. 1; 171-184.

[7] Kurscheid T, Schrappe M, Lauterbach KW. Kritische Bewertung gesundheitsökonomischer Studien *(Critical evaluation of health economics studies).* In: Lauterbach, Schrappe (Hrsg). Gesundheitsökonomie, Qualitätsmanagement und Evidence-based Medicine. 2 Aufl. Stuttgart, New York: Schattauer 2004; 114-126.

[8] Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russel LB. Recommendation of the Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996; 276: 1253-1258.

[9] Hannoveraner Konsensus Gruppe. Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation – revidierte Fassung des Hannoveraner Konsens *(German recommendations on health economics evaluation - revised version of the Hanover Consensus).* Gesundheitsökonomie und Qualitätsmanagement 1999; 4: A 62-65.

[10] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ Economic Evaluation Working Party. BMJ 1996; 313: 275-283.

[11] Schumann J. Mikroökonomie *(Microeconomics).*Heidelberg New York: Springer; 1992.

[12] Konerding U. Gesundheitsbezogene Lebensqualität *(Health-related quality of life).* In: Lauterbach, Schrappe (Hrsg). Gesundheitsökonomie, Qualitätsmanagement und Evidence-based Medicine. 2 Aufl. Stuttgart, New York: Schattauer 2004; 160-182.

[13] Sinclair C, Frankel M. The effect of quality assurance activities on the quality of mental health services. Qual Rev Bull 1982; 8: 7-15.

2.6 Clinical practice guidelines and disease management programmes

2.6.1 Background to the evaluation of clinical practice guidelines

Clinical practice guidelines (CPGs) are seen as key instruments in the improvement and assurance of medical quality in the provision of health care for patients [1]. Their objective is to reduce inappropriate differences in patient care and improve care by means of the formulation of concrete recommendations for decisions. Furthermore, in Germany they are used as a basis for decisions on steering procedures in the health care system, e.g. for the formulation of requirements for disease management programmes (DMPs), in acc. with § 137f Social Code Book V. Consequently, CPGs are increasingly influencing decisions affecting the structural level of the health care system.

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

However, in many cases the connection to current scientific evidence is lacking [2,3], and CPGs on identical topics in part reveal considerable differences with regard to the content of their recommendations [4,5].

One main reason for this is that the internationally stipulated quality standards for CPG development are not consistently followed [6-8].

2.6.2 Aim of evaluation of clinical practice guidelines

The evaluation of CPGs aims to improve care through greater transparency in the health care system. It is therefore particularly important to:

- Discriminate between CPGs of good or bad methodological quality and quality of content;
- Make clear and specialist statements on the reasonability and effectiveness of the implementation of different medical recommendations;
- Offer the Federal Joint Committee or its panels a basis for decisions in discussions on DMPs;
- Ensure that only verified (quality-assured) CPGs, where evidence of an improvement in outcome exists, 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 the above procedures provide the users of CPGs (physicians, health facilities, health policy committees, decision-makers in the health care system and patients) with orientation on meaningful and adequate recommendations on high-priority health care problems.

For the specific evaluation of content of CPGs, the available methodological competence and expertise of external institutes, facilities or organisations are to be used and taken into account as far as possible.

2.6.3 Methods to evaluate clinical practice guidelines

Essential aspects in the assessment and review of CPGs are

- Internal validity;
- Relevance, appropriateness and practicality of health care recommendations;
- External validity.

Validity is the key quality criterion of a CPG.

A CPG is considered valid if the expected benefit (medical and/or economic outcome) can be achieved by its application [1,9,10]. Strictly speaking, the validity of a CPG can only be tested by a rigorous evaluation of effects [2]. Due to the high expenditure in financial and human resources, this cannot be realised for every existing CPG; indeed, pilot studies have been conducted for only a few CPGs before their publication [11,12].

For the final evaluation, a multi-stage process combining various aspects of the assessment of form and content is therefore conducted and summarised in a full report. The methodology applied by the Institute will be regularly reviewed and, if necessary, updated taking current scientific publications and national and international experience into account.

I. Formal assessment

An approach to the question of CPG validity can be made by formal assessment following methodological criteria that have been shown to have a great influence on validity [5,6,13]. Formal CPG evaluation is conducted in a structured manner based on CPG clearing procedures, and in accordance with evaluation criteria developed by the German Medical Association *(Bundesärztekammer)* and the Association of Statutory Health Insurance Physicians *(Kassenärztliche Vereinigung)* [14,15]. The evaluation is performed by two independent assessors. Where conflicting assessments are made, the issues will be discussed and evaluated once again. If the dissent continues, the mat-

ters of dissent will be separately documented. A number of CPGs, often differing greatly in method and content, exist worldwide on specific medical issues [4]. The formal assessment of CPGs has an important function as a filter for further steps in the evaluation of single recommendations.

II. Comparison of clinical practice guidelines and evaluation of content for key recommendations

The evaluation of the CPG content is of special relevance. The criteria applied so far with conventional instruments (ÄZQ check list[l], AGREE instrument[m], DELBI[n]) for the identification and interpretation of evidence and for the formulation of CPG recommendations are essentially transparency criteria in which only the description of the process, e.g. for literature searches, is evaluated, without providing an assessment of the completeness and topicality of the search. For the essential key recommendations included in a CPG, their derivation from the underlying evidence must therefore be individually assessed. Besides the evaluation of completeness and topicality of the literature consulted, the assessment of content includes the evaluation and interpretation of study results. As this procedure involves much time and effort, for pragmatic reasons the assessment of content must be limited to the research questions posed by the Federal Joint Committee or to the CPG's key recommendations. Identification of the key recommendations is made within the context of each particular assignment in consultation with the department heads and external experts concerned. If information on individual research questions is lacking in CPGs, these questions will be processed in consultation with the Institute's other departments.

A synoptic comparison of the content of the CPGs can be helpful to identify key recommendations. In particular, questions that are the subject of scientific dissent can be identified. Methodologically, the synoptic comparison only facilitates the evaluation process. An evaluation of the underlying evidence is also useful in procedures that are consistently recommended.

A comparison of CPG recommendations with procedures generally applied in routine health care is particularly useful. If complex amendments to the CPGs are required, implementation is more difficult and must be accompanied by supportive tools and measures [5].

---

[l]Ärztliches Zentrum für Qualität in der Medizin *(Agency for Quality in Medicine).* In AWMF (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften; *Association of the Scientific Medical Professional Societies*), 2001, Leitlinien-Manual *(Guideline manual).*

[m]Appraisal of Guidelines Research and Evaluation.

[n]Deutsches Leitlinien-Bewertungs-Instrument *(German Guideline Evaluation Instrument).*

To identify current CPGs for the respective research questions, comprehensive CPG searches will be conducted in the relevant specialist databases (CPG and literature databases), taking into account the procedures mentioned in Section 4.8 (literature search). The search strategy (search words, choice of databases, etc.) as well as inclusion and exclusion criteria will be determined and documented beforehand with the optional support of an expert.

III. Evaluation of the relevance, appropriateness and practicality of recommendations

What is scientifically correct might not necessarily be meaningful, practicable and appropriate.

Individual recommendations/key points of CPGs are assessed very differently by those affected with regard to their relevance, appropriateness and practicality. This cannot be captured by a formal assessment but should be evaluated by way of professional expertise (peer review), and by patients. To this end, independent focus groups comprising providers in various health care fields [11,16,17] as well as patients can be formed to assess the relevance of content of recommendations and, if necessary, identify important missing elements relevant to health care.

Focus groups will be formed by the Institute.

If the focus group concludes that essential research questions on improvement in health care are not covered by available CPGs, or that evidence currently available has not been taken into account, the missing evidence will be sought and evaluated.

IV. Improvement of outcomes

The central question in the preparation and implementation of CPGs is whether their implementation leads to a measurable improvement in health care. If results of pilot studies or projects testing CPGs are available, these are to be included in the overall evaluation (e.g. by describing methods, quality indicators, results and consequences). CPGs from other countries in particular must be assessed as to the transferability of their conclusions to the German health care system and/or to the structural prerequisites needed for their successful implementation.

The Institute can be commissioned by the Federal Joint Committee to evaluate CPGs.

2.6.4 Presentation of quality assessment

A structured report will be prepared from the evaluation results available to provide the Federal Joint Committee with a basis for further consultations. Where DMPs are concerned, concrete proposals for the structuring of DMPs will be developed on the basis of these evaluations.

The reports can also serve as the basis to produce topic-related information for physicians and patients.

2.6.5 Submission of recommendations on disease management programmes

The Federal Joint Committee makes recommendations, in accordance with § 91 Social Code Book V, on the choice of chronic diseases to be included in DMPs. This choice is based on:

1. The number of insured persons affected by the disease;
2. The possibilities of improving the quality of health care;
3. The availability of evidence-based CPGs;
4. The need for treatment that overlaps health care sectors;
5. The possibility of influencing the course of the disease through the personal initiative of the insured person;
6. The cost of treatment.

For the chronic diseases selected, the Federal Joint Committee makes recommendations to the Federal Ministry of Health on the structure of DMPs.

This especially refers to:

1. The treatment according to current medical knowledge with consideration of evidence-based CPGs or according to the best available evidence, taking the respective health care sectors into account;
2. The quality assurance measures to be conducted;
3. The prerequisites and procedures for registration of insured persons in a programme, including the duration of participation;
4. Training of care providers and insured persons;
5. Documentation;

6. Evaluation of the effectiveness and costs, of the time span between the evaluations of a programme, and of the duration of its approval in accordance with §137g Social Code Book V.

The Institute's aims and role in the submission of recommendations on disease management programmes

The aim is to prepare scientific information for the Federal Joint Committee on individual aspects of the DMPs as a basis for its decisions. As the requirements not only involve denoting evidence-based CPGs, but also the specification of the participating providers' obligation to supply services, these requirements can form a component of current evidence-based CPGs. Furthermore, they can contain recommendations made on the basis of current findings from systematic literature searches and evaluations. Consequently, quality indicators can be developed that depict the improvements in health care.

Further possibilities of supporting the Federal Joint Committee and its panels in the development, assessment or evaluation of DMPs will be assessed in consultation with the responsible committees.

References

[1] Council of Europe. Methodology for drawing up guidelines on best medical practice – recommendation No R(01)13., Council of Europe, Strasburg, own publication. http://www.coe.int.

[2] Helou A, Ollenschläger G. Ziele, Möglichkeiten und Grenzen der Qualitätsbewertung von Leitlinien. Ein Hintergrundbericht zum Nutzermanual der Checkliste „Methodische Qualität von Leitlinien". *(Possibilities and limits of quality evaluation of guidelines. A background report on the user manual for the checklist "Methodological quality of guidelines").* Z ärztl Fortbild Qualsich 1998; 92: 361-365.

[3] Savoie I, Kazanjian A, Bassett K. Do clinical practice guidelines reflect research evidence? J Health Serv Res Policy 2000; 5: 76-82.

[4] Irani J, Brown CT, van der Meulen J, Emberton M. A review of guidelines on benign prostatic hyperplasia and lower urinary tract symptoms: are all guidelines the same? BJU Int 2003; 92: 937-942.

[5] Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R, and the AGREE Collaboration. 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: 148-157.

[6] Cluzeau F, Littlejohns P, Grimshaw J, Feder G, Moran S. Development and application of a generic methodology to assess the quality of clinical guidelines. Int J Qual Health Care 1999; 11: 21-28.

[7] Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed literature. JAMA 1999; 281: 1900-1905.

[8] Bandolier Forum, Independent evidence-based health care, On Care Pathways, July 2003, p. 1-12. http://www.ebandolier.com; Access on 10 October 2004.

[9] Grol R, Dalhuijsen J, Thomas S, in t'Veeld C, Rutten G, Mokkink H. Attributes of clinical guidelines in general practice: observational study. BMJ 1998; 317: 858-861.

[10] Worrall G, Chaulk P, Freake D. The effects of clinical practice guidelines on patient outcomes in primary care: a systematic review. CMAJ 1997; 156: 1705-1712.

[11] Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „COPD" *(Agency for Quality in Medicine. Guideline clearing report "COPD").* ÄZQ-Schriftenreihe Band 14; Verlag Videel, Niebüll; 2003.

[12] Ärztliches Zentrum für Qualität in der Medizin. Leitlinien-Clearingbericht „Depression". *(Agency for Quality in Medicine. Guideline clearing report "Depression").* ÄZQ-Schriftenreihe Band 12; Verlag Videel, Niebüll; 2003.

[13] Field MJ, Lohr KN. Clinical practice guidelines. Directions from a new program. Institute of Medicine, Washington D.C. 1990.

[14] Ärztliche Zentralstelle Qualitätssicherung. Checkliste „Methodische Qualität von Leitlinien". *(Medical Centre for Quality Assurance. Checklist "Methodological Quality of Guidelines").* Dtsch Ärztebl 1998; 95: A2576-A2578, C1838-C1840.

[15] Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, Ärztliche Zentralstelle Qualitätssicherung. Das Leitlinien-Manual. *(Association of the Scientific Medical Societies, Medical Centre for Quality Assurance, Guideline Manual).* Z ärztl Fortbild Qualsich 2001; 95 (Suppl. I): 1-84.

[16] Kitzinger J. Introducing focus groups. BMJ 1995; 311: 299-302.

[17] Kean S. Focus Group Interviews: Ein qualitativer Forschungsansatz in der Pflege. *(A qualitative research approach to health care).* Pflege 2000; 13: 145-151.

2.7 Systematic reviews and HTA reports

Systematic reviews and HTA reports provide a descriptive summary of the current status of studies and are a valuable basis for decisions to be made by physicians and patients, as well as for decision-makers in the health care system. The Institute will use available systematic reviews and HTA reports for the production of its reports and patient information materials. The prerequisite for this is that the methodology of these papers corresponds to the Institute's requirements. In particular, a transparent description of the literature search and selection strategy is necessary; the strategy and inclusion and exclusion criteria used for selection of the relevant literature should also comply with the principles set out in the Institute's respective report plan.

Like any other scientific publication, a systematic review can not only lead to incorrect findings, but its conclusions can also be affected by systematic bias [1]. In principle, it needs to be taken into

account that considerable differences in quality may exist in systematic reviews and therefore may not justify general classification of these reviews into a high evidence level.

The number of systematic reviews has increased substantially in recent years [2]. The increase in published systematic reviews also raises problems of contradictory conclusions within reviews, of differences between the results of reviews [3], and of differences between the results of RCTs with regard to comparable research questions [4].

A necessary but insufficient prerequisite for the inclusion of systematic reviews in the Institute's reports is a methodological assessment based on the QUORUM statement [1] and the methods of the Centre for Reviews and Dissemination in York, according to which systematic reviews are assessed before they are entered into the DARE[o] database (http:www.York.ac.uk/inst/crd/). Furthermore, assessments of content follow these assessments of form.

An essential aspect of evaluation, besides the description of the methodology applied in the production of the review, is the issue whether the studies included were subjected to a quality assessment, what the results of the assessment were, and whether or how it had any influence on a potential synthesis of the individual study results. A systematic review cannot be better than the individual studies upon which it is founded; the inclusion of poor quality studies can lead to a substantial bias of results [5].

Critical evaluation of HTA Reports

In contrast to systematic reviews, HTA reports are not written primarily as an aid for daily clinical work. Their aim is to provide relevant information on decisions to be made at various management levels of the health care system [6]; for example, decisions on investments, on cost coverage of services and on the structure of the statutory health insurance catalogue. Furthermore, HTA reports are often consulted as an important reference for the development of CPGs and production of patient information materials.

A standardised format for HTA reports as suggested by the EUR-ASSESS[p] Group [7] has not yet found general acceptance, which indicates that local conditions always have a substantial influence on the structure of reports. Quality criteria for the assessment of HTA reports are therefore primarily related to orderly, transparent and reproducible execution and interpretation of the single constituents (e.g. of health economics reviews or analyses) [8].

[o]Database of Abstracts of Reviews of Effectiveness.

[p]The "Coordination and Development of Health Care Technology Assessment in Europe" programme created by the European Union.

For the assessment of HTA reports, the Institute will follow the internationally recognised standards (e.g. INAHTA[q] or ECHTA[r]) for their production and evaluation [9,10].

In particular, the following factors will be considered [6]:

- The description of the background to the evaluation of the technology;
- The formulation of the specific research question;
- Details on the technological status quo;
- Technical characteristics;
- The systematic evaluation of safety and clinical effectiveness;
- The evaluation of health economics;
- The connection between organisational structures and procedures and the technology used;
- A discussion on generalisability and transferability;
- The assessment of ethical, social and legal implications.

The practice by the former Federal Committee of Statutory Health Insurance Physicians and Sickness Funds[s] *(ehemaliger Bundesausschuss der Ärzte und Krankenkassen)* of seeking the opinions of interested parties has proven useful for the acceptance of HTA reports in the German health care system [8]. The Institute will consider whether such opinions will be taken into account in the evaluation of HTA reports.

The inclusion or exclusion of available HTA reports will be documented in a comprehensive manner. The respective criteria depend on the topic to be assessed. They are to be agreed upon by the project management beforehand and to be documented in the report plan (q.v. Section 4.4).

References

[1] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomized controlled trials: the QUORUM statement. Lancet 1999; 354: 1896-1900.

[2] Chalmers I, Haynes RB. Reporting, updating, and correcting systematic reviews of the effects of health care. In: Chalmers I, Altman DG, eds. Systematic reviews. London: BMJ Publishing Group 1995: 86-95.

[q] International Network of Agencies for Health Technology Assessment.
[r] European Collaboration on Health Technology Assessment.
[s] A sickness fund is a statutory health insurance fund.

[3] Jadad AR, Cook DJ, Browman G. A guide to interpreting discordant systematic reviews. CMAJ 1997; 156: 1411-1416.

[4] LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials . N Engl J Med 1997; 337: 536-542.

[5] Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials JAMA 1995; 273: 408-412.

[6] Perleth M. Kritische Bewertung von HTA-Berichten. *(Critical evaluation of HTA reports).* In: Kunz R, Ollenschläger G, Raspe H, Jonitz G, Kolkmann FW (Hrsg.). Lehrbuch Evidenzbasierte Medizin in Klinik und Praxis. Dt Ärzteverl 2000: 147-155.

[7] Liberati A, Sheldon TA, Banta HD. EUR-Assess project Subgroup report on methodology: methodological guidance for the conduct of health technology assessment. Int J Techn Assess Health Care 1997; 13: 186-219.

[8] Gibis B, Rheinberger P. Erfahrungen mit Health Technology Assessment im Bundesausschuss der Ärzte und Krankenkassen *(Experiences of the Federal Committee of Statutory Health Insurance Physicians and Sickness Funds with Health Technology Assessment).* Z ärztl Fortbild Qualsich 2002; 96: 82-90

[9] Busse R, Velasco M, Perleth M, Orvain J. Best practice in undertaking and reporting HTA. ECHTA working group 4 report: 1-104.

[10] Hailey D. Toward transparency in health technology assessment. A checklist for HTA reports. Int J Techn Assess Health Care. 2003; 19: 1-7.

2.8 Prognosis

The basis for the evaluation of prognostic studies is the precise formulation of a research question, as studies conducted on the evaluation of prognostic characteristics have different objectives (evaluation of risk factors, score development/validation, so-called responder analyses, etc.). Moreover, discrimination between diagnosis and/or screening studies can be difficult and, depending on the study objectives, different evaluation principles are applied. The formulation of a research question or the procedure leading to the final research question follows the principles described in Section 4.5. A prognostic characteristic provides information that should not be an end in itself but should lead to a consequence that constitutes a verifiable benefit for the patient. In this respect, the (general) requirements applying to a prognostic procedure are similar to those that apply to a diagnostic test (q.v. Sections 2.3.1 and 2.3.2).

If a prognostic characteristic is to be used within a screening or prevention programme, then the principles formulated in Sections 2.4 and 2.13 need to be considered in its assessment.

The following points are especially relevant in the evaluation of studies on prognostic characteristics:[t]

- Clear determination of one characteristic to be evaluated (usually described as an exposition or risk [or protective] factor in epidemiological studies). If a number of characteristics are to be assessed simultaneously in a study, the problem of multiple testing should be taken into account (q.v. Section 1.2).
- Clear definition of the characteristic to be evaluated,[u] including its statistical handling (e.g. dichotomisation or assessment of terziles or quartiles for a quantitative characteristic) and justification of the procedure selected.
- Clear determination and definition of potential confounders and effect modifiers, including their statistical handling.
- Clear definition of the outcome on which the prognostic assessment is to be based.
- Clear formulation of a research question and the study design related to it. This includes sample size estimation, which can for example be based on the desired precision of the estimate (width of the confidence interval) and requires an estimate of both the prevalence and incidence of exposition with regard to the outcome 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 and recruitment procedure of study participants (random selection, representativeness issues).
- Homogeneity of the population investigated. If the population is heterogeneous, it needs to be ensured that a prognostic statement can be made as constantly as possible on the subgroups causing heterogeneity (e.g. existence of different baseline risks for the outcome in question).
- In cohort studies, completeness of the 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. development within a so-called learning sam-

[t] It should be noted here that no generally accepted quality criteria exist for the evaluation of prognostic studies [2].

[u] Even if, as described above, several characteristics are simultaneously assessed in a study, for the sake of better readability only the singular case will be used here and in following sections, unless otherwise indicated.

ple 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 as risk factors include case-control and cohort studies and, in exceptional cases (when assessing constant characteristics), also cross-sectional studies. The underlying principles for the evaluation of such studies beyond the aspects mentioned above are described in Section 1.6.

The methodological quality of studies (or their publications) on prognostic characteristics is frequently insufficient [1]. Therefore meta-analyses (not systematic reviews) of prognostic studies are often inappropriate and their findings should be used with some reservation [2]. The literature search for the evaluation of prognostic characteristics (in a systematic review) is more difficult than, for example, for therapeutic studies, and no generally accepted optimum search strategy exists to date. Furthermore, it can be assumed that this research field is especially susceptible to publication bias [2].

References

[1] Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, Kasten LE, McCormack VA. Issues in the reporting of epidemiological studies: a survey of recent practice. BMJ 2004; 329: 883-887.

[2] Altman D. Systematic reviews of evaluations of prognostic variables. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context, 2nd ed. London: BMJ Books; 2001.

2.9 Individual risk assessment

Besides using the results of studies investigating single or, in the majority of studies, multiple prognostic characteristics, risk charts (also called risk engines) are being increasingly employed to assess the individual risk of patients (or clinically healthy persons) of experiencing an adverse event. Multi-factorial estimates for the concurrence of numerous risk factors are made in these charts (e.g. the Sheffield Table [1] or Joint British Chart [2]).

The following factors should be considered when evaluating such instruments:

- What type of study the underlying data originate from;
- Whether the variables included in the multi-factorial analysis were also analysed together in the underlying studies;

- Whether, and if so, how a multi-factorial statistical analysis was conducted in the underlying studies;
- If a multi-factorial analysis in the underlying studies was performed, whether adequate statistical models (e.g. log-linear models) were employed, taking interactions into account;
- Whether these instruments were ever validated in subsequent studies (test samples).

References

[1] Wallis EJ, Ramsay LE, Ul Haq I, Ghahramani P, Jackson PR, Rowland-Yeo K, Yeo WW. Coronary and cardiovascular risk estimation for primary prevention: validation of a new Sheffield table in the 1995 Scottish health survey population. BMJ 2000; 320: 671-676.

[2] British Cardiac Society, British Hyperlipidaemia Association, British Hypertension Society, endorsed by the British Diabetic Association. Joint British recommendations on prevention of coronary heart disease in clinical practice. Heart 1998; 80(Suppl2): S1-S29.

2.10 Evaluation methods depending on prevalence and type of disease

Every generalising statement in a study contains an error and can therefore lead to incorrect estimations. Amongst other things, the potential size of error is dependent on the study design, its conduct, the quality and completeness of the data evaluation and the quality of the scientific report. The possible size of this error therefore needs to be considered in the interpretation of study results and in their transferability to real-life health care conditions. The possibility of an incorrect estimation needs to be set in relation, for example, to the type and severity of disease, to the size and nature of the diagnostic or therapeutic effect and to the quality of the evidence on already existing diagnostic tests and therapeutic techniques with similar objectives.

For very rare diseases, it can sometimes be extremely difficult or even impossible to include enough patients in a clinical study to statistically detect even moderate effects with sufficient power. This does not however justify the renunciation of essential methodological instruments for the design of clinical studies, such as randomisation, double blinding and the intention-to-treat principle. To confront the power problem, it can be meaningful, when evaluating studies on such indications, to accept a priori – i.e. in ignorance of the data situation – a higher error level than that typically applied in medical research, namely the (usually two-sided) 5% level (q.v. Section 1.2). However, care should be taken that the criterion “very rare disease” is not created artificially by inadequate subgroup formation involving diseases that are not actually very rare (so-called disease slicing).

For very serious diseases, an (upward) adjustment of the (statistical) error level can be considered, for example, in diseases that have recently been described and lead to irreversible morbidity or even death in many patients within a very short time (e.g. AIDS in the 1980s, or SARS).

For diseases with an almost deterministic course (e.g. glioblastoma multiforme), studies of a lower evidence level (e.g. case series) can also be sufficient to conclude a positive evaluation of an effect. The prerequisite is that these studies fulfil the respective quality standards for the respective study type and that the study results show a dramatic effect running counter to the natural course of the disease. Conversely, the same of course applies to the observation and evaluation of serious (unexpected) adverse events in conditions of a known benign course.

2.11 Evaluation of complaints not directly associated with a defined disease

"Complaints not directly associated with a disease" are syndromes (e.g. chronic fatigue syndrome) that cannot be clearly classified into current disease entities mainly founded on pathophysiological mechanisms. In principle, therapeutic interventions for these syndromes can be evaluated with the same methods used for other types of diseases.

When evaluating diagnostic tests and prognostic techniques for the identification of such syndromes, special care should be taken that these tests and techniques are not part of the syndrome pattern, as this normally excludes an evaluation with regard to discrimination ability and prognostic accuracy.

2.12 Evaluation of complementary techniques

Complementary techniques, no matter what their nature (essentially diagnostic or therapeutic), can in principle be evaluated with the same methods used to evaluate scientifically founded techniques. This has been demonstrated by numerous examples (e.g. homeopathy [1] and acupuncture [2]). A reference to the alleged non-assessability of such techniques by scientific methods due to specific characteristics such as a special physician (or therapist) – patient relationship, or special disease (or health) constructs is unacceptable.

It would be useful, on the basis of the Institute's methodology, to include protagonists of these techniques both in protocol planning for the generation of scientific reports and in the report reviewing process.

References

[1] Jonas WB, Anderson RL, Crawford CC, Lyons JS. A systematic review of the quality of homeopathic clinical trials. BMC Complement Altern Med 2001; 1 :12.

[2] Haake M, Muller HH, Schade-Brittinger C et al. The German multicenter, randomized, partially blinded, prospective trial of acupuncture for chronic low-back pain: a preliminary report on the rationale and design of the trial. J Altern Complement Med 2003; 9: 763-770.

2.13 Evaluation of population-wide prevention and intervention measures

A methodological assessment of available studies on population-wide interventions is made according to the previously mentioned criteria on the assessment of individual study types.

Furthermore, an additional crucial requirement for preventive and health promotional measures is that they contain effective strategies to maintain public health and thereby have a lasting positive effect on the quality of life, mobility and performance capacity of the population.

The abundance of possible preventive measures on the one hand, and limited resources on the other, necessitate prioritisation of measures, taking the general principles on prioritisation of health issues into account (q.v. Section 4.10).

Previous prevention programmes have focused on a younger target population, emphasising the possibility of influencing deficits at this age. However, shortcomings of these programmes are evident in older age groups [1], in occupational health promotion schemes and in schools, health care facilities and areas with social problems [2].

Preventive measures must be clearly aimed at specific target groups. A detailed analysis of the needs, expectations and habits of the target group needs to be conducted beforehand. Provided that a meaningful intervention is planned, it must be clear which conditions are to be created so as to enable equal access to the intervention to all members of the target group.

An evaluation of effects has temporal (medium or long-term improvement in health) dimensions, as well as dimensions of perspective (from the individual's point of view, social and economic aspects). Clear objectives concerning the effectiveness of the intervention should be defined for such an evaluation.

Programmes designed to achieve long-term effects should lead to a reduction in premature morbidity and/or mortality and to an improvement in and/or maintenance of quality of life.

The following parameters are particularly useful to evaluate medium-term effects and the estimation of the success of such programmes:

- Health-related outcome parameters;
- Health-related quality of life [3];
- Evaluation of the development of competences (e.g. knowledge, self-confidence, attitude);
- Evaluation of access paths;
- The target group's satisfaction with the structure and conduct of the programme;
- Capacity building [4].

Sensitivity analyses should be employed to assess whether and under what circumstances the benefits achieved are stable.

For the evaluation of net benefits for the individual or society, health economic analyses are required (q.v. Section 2.5), if necessary, taking study results from occupational or communal projects into account.

References

[1] World Health Organization. The Jakarta Declaration on Leading Health Promotion into the 21st Century. WHO 1997, Geneva.

[2] Badura B. Gesundheitsförderung und Prävention – Schritte in eine „gesunde Gesellschaft". *(Health promotion and prevention – Steps towards a "healthy society").* Public Health Forum 2000; 28: 5-7.

[3] Bullinger M. Erfassung der gesundheitsbezogenen Lebensqualität mit dem SF-36 *(Determining health-related quality of life with SF-36).* Health Survey. Rehabilitation 1996; 35: 17-27.

[4] Hawe P, Noort M, King L, Jordaens C. Multiplying health gains: the critical role of capacity-building within health promotion programs. Health policy 1997; 39: 29-42.

2.14 Quality management in health care

Numerous Social Code Book V clauses and many in- and outpatient initiatives aim at improving the quality of health care. In the same way, many new health care models concentrate on improving quality and achieving greater cost-effectiveness. The Institute may, in consultation with the Federal Joint Committee, other panels and its own Health Economics Department and Statistics Department, determine and evaluate these various quality aspects and models.

2.15 Description of the type and size of the placebo effect

Studies can be biased by poorly controlled influences which may alter the study results. These influences mainly occur when study participants in the treatment group and placebo control groups know who is assigned to which group [1]. The blinding of a study and the use of a substance showing no pharmacological effects have the aim of minimising these influences. Furthermore, placebos possibly have inherent effects on medical parameters, mainly on those involving subjective assessments, such as pain [2]. When interpreting placebo-controlled studies, it is therefore necessary to evaluate whether blinding was actually ensured by the use of a placebo. The placebo effect appears to decrease with the size of the study [2]. The misinterpretation of findings of non-placebo-controlled studies is therefore more likely in studies with small sample sizes. In some interventions blinding may be impossible, i.e. in therapeutic interventions with desired or undesired effects that are clearly identifiable by the study participant. It is therefore meaningful to determine how many patients and physicians correctly guessed the group to which the patients were allocated. If the size of the placebo effect is known in the intervention, it may be possible to determine the degree of bias of results caused by missing or impossible blinding. The absence of a placebo group can result from the fact that an effective therapy for a disorder or disease already exists and the use of a non-active therapy in the control group would be incompatible with the Helsinki Convention [3]. In such a case, blinding of treatment without using a placebo group may be necessary.

References

[1] Gillespie R. Manufacturing knowledge: a history of the Hawthorne experiments. Cambridge, England: Cambridge University Press; 1991.

[2] Hróbjartsson A and Gøtzsche PC. Is the placebo powerless? N Engl J Med 2001; 344: 1594-1602.

[3] Emanuel EJ and Miller FG. The ethics of placebo-controlled trials. N Engl J Med 2001; 345: 915-919.

3. Evidence-based health information for consumers and patients

3.1 Goal

The Institute aims to become an effective, reliable, trusted and popular provider of evidence-based health information education for consumers and patients. Relevant information from the Institute's reports will be communicated to the public in a comprehensive and integrated approach. The aim is to provide health care information that is outcome-orientated, objective and commonly used. Furthermore, this information should be adapted to patients' psychological needs and be easily understandable, without the need for specific medical knowledge.

The goal of advancing health and scientific literacy is to:

- Improve understanding of physical, mental and emotional health;
- Improve understanding of medical and scientific information, including the concept of evidence-based medicine;
- Promote health-related behaviour;
- Enable support by relatives and friends;
- Promote the critical use of health care services;
- Support active decision-making about health issues (e.g. participatory doctor-patient relationships) that is responsive to individual needs and values.

It is not intended that the Institute should provide advice directly to individual consumers and patients. It is the Institute's intention to enhance independent and responsible informed choices on health issues, giving priority to consumer and patient autonomy [1].

3.2 Information system

Internet-based [2,3] and offline computer-based [4] health information can positively affect the consumers' and patients' state of knowledge, choices, and physical, mental and emotional well-being. However, information and education interventions can also be ineffective or harmful, and some techniques are more effective than others [3,5-8]. The Institute's first medium for the dissemination of evidence-based health information will be its website.

The website will be developed into a comprehensive and versatile reference work, aimed at meeting a variety of individual health information needs of consumers and patients at different levels [5,6]. Various information products, which are outlined in the following Section 3.2.1, will be employed for this purpose. These products will be interlinked, and supported by definitions, explanations and supplementary information.

The website will include an electronic newsletter, downloadable texts and other files on health-related topics as media for health information. Downloadable print versions will also be available. In addition, it will be possible to reproduce electronic information on other websites.

3.2.1 Information products

Information products include comprehensive health information topics, short summaries and other products.

Comprehensive health information topics

The comprehensive information topics will form the main focus of the reference work to be developed, comprehensively covering a wide variety of health-related issues. These topics can also be developed from the Institute's scientific reports.

The comprehensive information topics should consider:

- Education about the disease or condition, including:
  - Anatomy,
  - Physiology,
  - Disease aetiology,
  - Recognition of symptoms,
  - Normal course of diseases,
  - Prognosis,
  - Potential complications,
  - Recognition of complications,
  - Recovery,
  - Possible recurrence of disease,

  - Recognition of recurrent disease,
  - Risk groups (including relatives).
- Preventive measures and health promotion, including
  - Nutrition,
  - Physical activity,
  - Screening techniques,
  - Information.
- Diagnostic measures, including complementary diagnostic tests.
- Therapeutic measures, including
  - Pharmaceuticals,
  - Surgery,
  - Other non-pharmaceutical procedures,
  - Complementary therapies.
- Rehabilitation measures.
- Other health care services.
- Psychosocial aspects, including testimonies of patients suffering from different diseases as well as those of their relatives.

Short summaries

A much larger range of short summaries will be produced than comprehensive information topics. These short summaries will complement the reference work and make evidence-based information, currently only available in English, accessible to the general public.

The short summaries consist of easily understandable texts detailing important, interesting and/or current health topics. Short summaries will also be developed from the Institute's scientific reports.

Other products

Other products include visual and interactive tools such as diagrams, online calculators (e.g. for cigarette costs) and glossaries (e.g. online dictionaries).

These tools are designed to:

- Promote general understanding of health and medical issues;
- Improve understanding of diseases, e.g. develop knowledge of the normal course of diseases, symptom recognition, possible complications, recovery and possible recurrence of the disease;
- Increase ability to understand and weigh potential risks;
- Support self-management strategies, e.g. for chronic diseases.

3.2.2 Editorial system

The Department of Patient Information and Research, in consultation with the Institute's Director and the affected departments, will ensure that the contents of the health information website and other health information products are:

- Evidence-based and consistent with the current state of scientific knowledge;
- Orientated towards consumers' and patients' information needs and adapted to their psychological and emotional needs as far as possible;
- Consistent with all other information products published by the Institute.

The Head of the Department of Patient Information and Research will be the editor of the health information products.

Internal and external criticism on contents and quality of the health information provided will be referred to the editor, who will be responsible for initiating appropriate action and reporting regularly to the Steering Committee. Major or urgent problems will be reported immediately to the Institute's Director.

Close liaison between the editor, the Institute's Director, the Department of Public Relations and Quality Management and all other departments concerned will ensure that all publications are consistent in content. The final decision to publish will be the responsibility of the Steering Committee.

3.2.3 Multilingualism

The Institute will aim to publish health information in both German and English and keep both versions up to date. A broad international exchange and the best possible quality assurance are only

to be achieved through publications in English. This will enable the quality of the health information products to profit from feedback from international scientists and reviewers (e.g. authors of systematic reviews).

The Institute will also cooperate with external partners to enable at least some of its health information to be translated into the most widely spoken languages in Germany.

It is very difficult to assess the quality of translations according to objective criteria. No specific standard can be defined. Translations can be literal or can aim to capture the intent of the original in the target language. The Institute will frequently apply the latter method and the quality of samples of translations will also be checked.

The Department of Patient Information and Research will rely on in-house bilingual staff with expertise in translation. In addition, the foreign language competence of its own staff will be promoted. Computer-assisted translation may also be used. These programmes include a range of different tools, which compile and standardise terminology and phrases, and thereby facilitate and standardise the work of human translators.

3.3 Development of information products

Both internal and external advice will be essential factors in quality assurance of information products. Where the term “advisory group” is used in the following text, this refers to the patient representatives on the Board of Trustees as well as the Patient Information Panel in the Federal Joint Committee.

3.3.1 Selection of topics

The selection of health information topics is to reflect public interest as far as possible. It should be balanced, impartial and transparent, and supported by well-researched material to inform decision-making.

As the generation and maintenance of comprehensive health information topics involves significant investment of resources, rigorous methods of priority-setting will be required [9-11].

Priority-setting for comprehensive health information topics will be undertaken according to the following process:

a. Development of filter criteria (main and supplementary criteria):

The main filter criteria include the quantity and quality of scientific research on a topic, as well as public interest in it. Additional criteria include the size of the target population affected, general information and educational aspects, the current state of knowledge of the population, as well as assumptions about effects on health status and possible risks (at both the individual and population level).

b. Internal and external consultations on filter criteria.

c. Determination of filter criteria.

d. Evaluation.

On the basis of these filter criteria, the systematic selection of topics will be undertaken as follows:

a. Application of the filter criteria to a comprehensive number of topics using a three-stage model:

   Stage 1: Application of the main criteria (coarse filter),

   Stage 2: Application of supplementary criteria to topics selected in Stage 1,

   Stage 3: Internal and external consultations on the topics selected in Stage 2.

b. Final selection of topics by the Steering Committee.

c. Evaluation of the selected topics.

The complete process of health information production is summarised in Figure 3 in Section 3.3.3.

3.3.2 Scope and contents

A range of methods can be used to help identify questions which the health information needs to address. These methods vary in their cost and practicality, as well as in the transferability of their results to other health information issues [12,13]. As far as possible, the Institute will use high-quality data, surveys and studies. These materials may be supplemented by telephone interviews with experts (key informants) and/or focus groups and advisory groups (q.v. first paragraph, Section 3.3 and Section 1.24). The Institute's decisions are always made under the premise that the research questions posed are in the public interest. Special attention will be paid to the needs of disadvantaged population groups.

Internal project groups, consisting of members from the Department of Patient Information and Research and from other departments, will be formed to work on individual comprehensive health information topics. Formation of the project group and internal project coordination will be the responsibility of the Department of Patient Information and Research.

a. In the first project group meeting, a scoping exercise will determine the range of subjects to be incorporated in the health information topic.

b. A review of health care information available to consumers and patients on the subject (e.g. diagnostic tests and therapies) will then be carried out. A literature search and a search of key websites (the most important websites with information on health issues) will be undertaken. Telephone interviews of key informants (e.g. patient representatives and clinical experts) may be undertaken.

c. Consumers' and patients' needs, their current level of knowledge and their potential interest in the subject will be analysed as far as possible. Here again, a literature search including a search for websites containing information on patients' experiences [14] and telephone interviews with key informants will be carried out. Focus groups may also be formed on some occasions.

d. A draft of the proposed health information will be discussed at the second project meeting. The draft will cover the central question of the topic. The results of the literature search will also be tabled.

3.3.3 Production

In principle, the same evidence-based methods are applied to the production of health information as to those applied in the Institute as a whole. The individual information products will be developed as described below.

Comprehensive patient information topics

The production of comprehensive patient information topics is carried out as follows:

a. Literature searches for published systematic reviews: The validity and topicality of the reviews found (as well as any subsequently published studies) will be discussed and evaluated. If no systematic reviews but only single studies are available, these will be the basis for evaluation and use.

b. Production of a preliminary version of the comprehensive information topic from the results of the literature search.

c. Internal peer review:
   Multiple checking of the preliminary version. The resulting feedback or criticism will be discussed and, if necessary, a further analysis and review will be made.

d. External peer review:
   Expert opinions will be sought from patient representatives, representatives working in health promotion fields, clinical experts and the advisory group (q.v. first paragraph in Section 3.3). Lead authors of major systematic reviews will be given the opportunity to comment on the Institute's draft (usually on the English version). If required, relevant public agencies will be consulted.

e. Where possible, the readability and comprehensibility of the comprehensive information topics (German version) will be tested by representatives of the target group. Feedback and any corrections required will be discussed with the corresponding departments in the Institute, as well as the Director.

f. The revised version of the comprehensive information topic will be written in German and English and forwarded to the Steering Committee. The German version will then be marked up for the Internet as a test version. The usability of this online version will be tested with 3-5 volunteers, including at least one patient or patient representative [15-17]. The English version will be subsequently marked up for the Internet as an (offline) test version. Finally, all other communication tools (e.g. downloadable versions) will be developed, tested and completed.

g. When multiple information products are to be produced from the source material and major parts of their content have been amended, it will be necessary for these products to undergo a comparable quality assurance procedure as described above.

h. Before publication, the comprehensive information topics will be discussed with the Institute's Director and the affected department heads. The Steering Committee can either release the final version or alternatively suggest further discussions or revisions (q.v. Section 3.2.2).

Short summaries

These summaries will be:

- Short single summaries abstracting a single systematic review or an important study, or several reviews or studies.
- Summaries based on a set of guidelines to be developed specifically for the production of these summaries.
- Summaries produced by the Department of Patient Information and Research itself; or by the Department of Patient Information and Research and other relevant departments and agreed upon with the Steering Committee; or by other relevant departments with editorial oversight on content and communication standards from the Department of Patient Information and Research.
- Summaries that are either reviewed by the same procedure applied to the review of comprehensive information topics, or summaries reviewed only internally in cooperation with the authors of reviews or of papers on single trials considered by the Institute (offering authors the opportunity of commenting on the interpretation of their work).

The Department of Patient Information and Research is responsible for the maintenance and updating of the short summaries.

Other products

Visual and interactive tools will either be developed by the Institute or purchased or commissioned. Tools of foreign origin will be adapted for Germany. All tools are to comply with the communication standards of the Department of Patient Information and Research as well as to health promotion standards.

3.3.4 Evidence-based communication standards

In its communication of health information, the Institute's goals are to:

- Communicate respectfully and effectively with the population in Germany so that it trusts the Institute as a reliable and easily understandable source of information;
- Produce health information that is easy and interesting to read without compromising scientific accuracy;

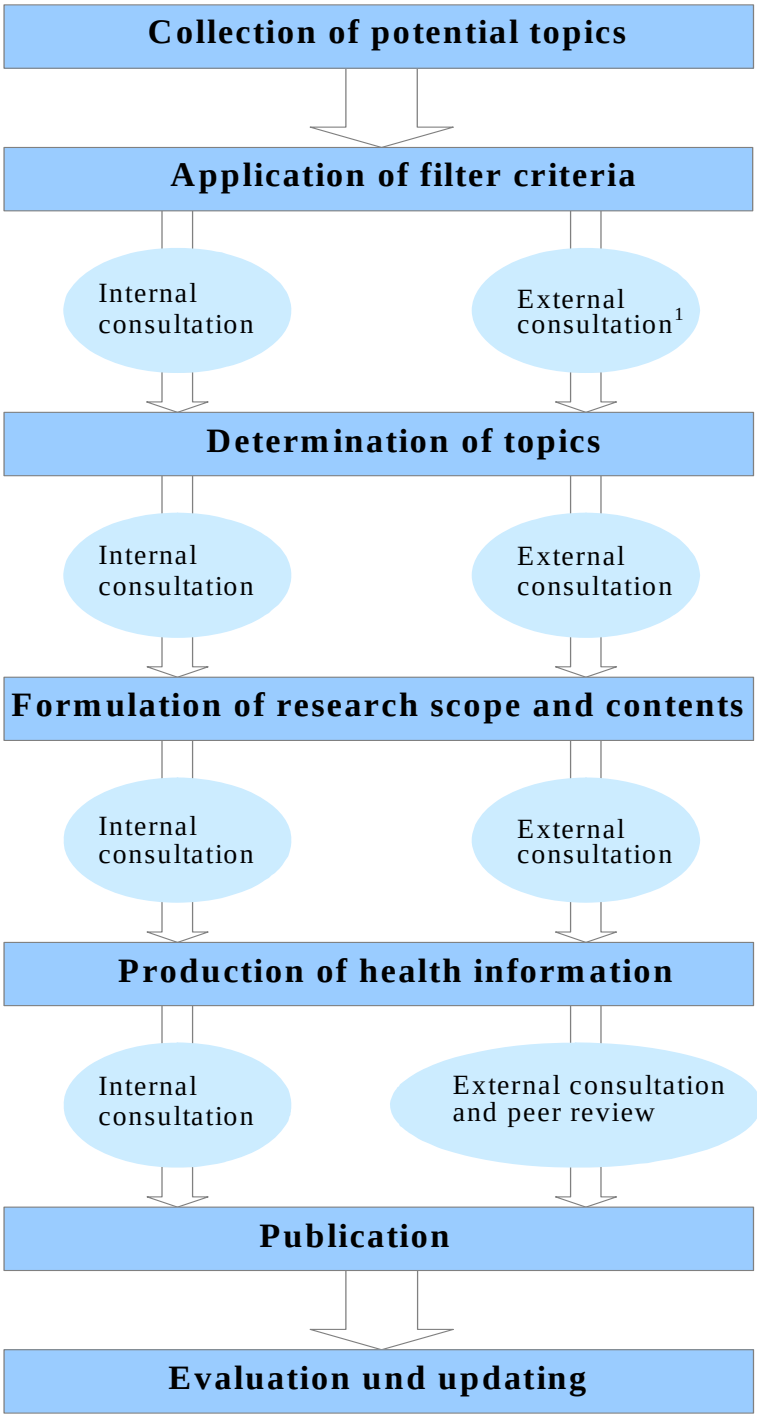
- Maintain a style of communication that is as neutral and unambiguous as possible;
- Demonstrate sensitivity and respect for patient knowledge, values and concerns, patient autonomy, and cultural differences;
- Support patient empowerment;
- Promote health and scientific literacy;
- Help the individual to relate evidence to his or her own personal situation;
- Respect readers' time**.**

The Institute will ensure that the website meets disability accessibility standards [18].

The Institute will clearly demonstrate its values through its professionalism and the high quality of the health information products it generates, as well as through the language and visual representations it employs. A house style guide will be developed for each type of information product created. This guide will be based as far as possible on evidence of effective communication.

The Institute will aim to be a popular source of information in Germany.

Figure 3: Development of comprehensive health information topics



[1] q.v. First paragraph in Section 3.3.

There are particular communication challenges in providing health information without exaggerating what is scientifically known and not telling people what they 'should' do. The effects of making people aware of the scientific uncertainty of much health care is largely unknown. The general

public is also accustomed to a more directive style of health information, often aimed at directly altering opinions and attitudes. The Institute aims to present information in a variety of ways, in order to enable as many people as possible to gain access to information [7].

As a general rule, health information detailing relative risks should be avoided. However, the inclusion of relative risk data may occasionally be necessary to enable the individual to compare treatments. These data will then not be presented on their own, but together with the absolute risk or the number needed to treat (the number of patients needed to be treated before a patient benefits from treatment or suffers harm through it).

There is evidence that the presentation of personalised or individualised risk estimates is an effective form of communicating health information [7,19]. The Department of Patient Information and Research can develop or adapt tools with which consumers and patients can estimate their personal risk, providing that reliable data on the development of tools are available.

Patient decision aids have been shown to be an effective means of communicating health information [20]. The Institute may develop or adapt decision aids and will seek to incorporate effective elements of these aids into its health information products.

In addition, the Department of Patient Information and Research will:

- Present health information in consistent formats;
- Explain the degree of uncertainty associated with this information;
- Indicate to which sub-population(s) the scientific evidence applies;
- Aim to achieve the highest possible standards of web usability (including web navigation);
- Be very careful to clearly distinguish between “absence of evidence” and “evidence of no effect”;
- Avoid biasing information in respect of the products of any particular company; the generic names of products will generally be used, supported by brand names of products only where relevant;
- Increase the understandability of the health information provided on the website by producing an online dictionary, integrated by way of hyperlinks.

3.4 Publications

The dissemination of health information will be discussed with the Institute’s Director, the affected department heads and the Department of Public Relations and Quality Management. Proposals will

be passed on to the Steering Committee. The Institute's Director, the Head of the Department of Patient Information and Research and the Institute's spokesperson will ensure the consistency of content for external communication.

Health information will mainly be published on the health information website as comprehensive information topics, short summaries or as other products (see also Section 3.2).

3.5 Evaluation and updating

There are many instruments and guidelines for the quality evaluation and assurance of health information in the Internet. However, there are no reliable data on the validity of these instruments and guidelines [21-24].It is unclear whether they in fact measure what they claim to measure. No reliable indicator of quality for the production of health information is therefore available [22,24]. There is also no evidence on the cost-effectiveness of production options.

Studies on research with patients, including patients in Germany [25,26], indicate that some issues suggested as being important in the evaluation of health information may not be important to patients. Some recommendations common in instruments and guidelines on the evaluation of information may actually reduce its scientific quality.

Most evaluation instruments focus on information about treatment and do not address the full range of a health issue (including aetiology, prognosis, screening and diagnostic tests). Yet health information on diagnosis and screening involves more complex decision-making and communication issues than information on treatment [6,7,19].

The health information produced by the Institute will therefore not rely on any currently available evaluation instruments, nor will it develop one. The Institute will, however, rely on evidence on specific aspects that could influence patients' decision-making and demonstrably affect the quality of information produced.

The Department of Patient Information and Research, in communication of health information, will stay up to date with the new scientific evidence on the options of quality and, if necessary, adapt its methods accordingly. Health information products will be tested as far as possible with representative population groups**.**

To ensure that information is kept up to date, the literature for individual health information products will be coded, enabling monitoring for publications of important new evidence and for changes in Cochrane reviews. The topicality of the health information products will also be ensured by the ongoing exchange of information between key personnel at the Cochrane Collaboration, the

Centre for Reviews and Dissemination and *Evidence-Based Medicine*. The currency will also be assured by a revision of the health information products, which will be carried out at least every two years.

The health information website will include a feedback mechanism for readers. Any form of internal and external feedback may result in immediate revision of health information already published.

Close cooperation between the Department of Patient Information and Research and other relevant departments ensures that the health information provided accords with current evidence. The Department of Patient Information and Research will rely on internal expertise and consultation with key informants, and offer the authors of systematic reviews the opportunity to comment on the Institute's interpretation of their work.

References

[1] Hope T. Evidence-Based Patient Choice. London: King's Fund; 1996.

[2] Bessell TL, McDonald S, Silagy CA et al. Do internet interventions for consumers cause more harm than good? A systematic review. Health Expectations 2002; 5: 28-37.

[3] National Institute of Clinical Studies. The impact of the internet on consumers' health behaviour. Prepared by the Centre for General Practice and the Centre for Evidence Based Practice, University of Queensland. NICS, Melbourne: 2003.

[4] Lewis D. Computer-based approaches to patient education: a review of the literature. J Am Informatics Association 1999; 6: 272-282.

[5] Coulter A, Entwistle V and Gilbert D. Informing Patients: An Assessment of the Quality of Patient Information Materials. London: King's Fund Publishing; 1998.

[6] Entwistle VA, Watt IS, Davis H et al. Developing information materials to present the findings of technology assessments to consumers: The experience of the NHS Centre for Reviews and Dissemination. Int J Tech Assess Health Care 1998; 14: 47-70.

[7] Edwards A, Bastian H. Risk communication – making evidence part of patient choices? In Edwards A, Elwyn G, editors. Evidence-Based Patient Choice: Inevitable or Impossible? Oxford: Oxford University Press, 2001; 144-160.

[8] Eysenbach G, Jadad AR. Consumer health informatics in the Internet age. In Edwards A, Elwyn G, editors. Evidence-Based Patient Choice: Inevitable or Impossible? Oxford: Oxford University Press, 2001; 289-307.

[9] Henshall C, Oortwijn W, Stevens A et al. Priority setting for health technology assessment. Theoretical considerations and practical approaches. A paper produced for the EUR-ASSESS project. Int J Technol Assess Health Care 1997; 13: 144-185.

[10] Townsend J, Buxton M, Harper G. Prioritisation of health technology assessment. The PATHS model: methods and case studies. Health Technology Assessment 2003; Vol 7: No 20. NHS R&D HTA Programme.

[11] Ghaffar A, de Francisco A, Matlin S. The Combined Approach Matrix: A priority-setting tool for health research. Geneva: Global Forum for Health Research, June 2004.

[12] Liberati A, Sheldon TA, Banta HD et al. Eur-Assess project subgroup report on methodology: Methodological guidance for the conduct of health technology assessment. Int J Tech Assess Health Care 1997; 13: 186-219.

[13] Sixma HJ, Kerssens JJ, Campben CV, Peters L. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. Health Expectations 1998; 1: 82-95.

[14] DiPEX. The Database of Patients' Experiences. http://www.dipex.org. Access on 18 October 2004.

[15] Krug S. Don't make me think: A common sense approach to web usability. Indiana: New Riders; 2000.

[16] Nielsen J. Designing web usability. Indiana: New Riders; 2000.

[17] Inan H. Measuring the success of your website: A customer-centric approach to website management. Sydney: Pearson Educational Australia, 2002.

[18] W3C. Web Accessibility Initiative. http://www.w3.org [Access on 26 January 2005]

[19] Edwards A, Unigwe S, Elwyn G, Hood K. Personalised risk communication for informed decision making about entering screening programs (Cochrane Review). In: The Cochrane Library, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd.

[20] O'Connor AM, Stacey D, Rovner D et al. Decision aids for people facing health treatment or screening decisions (Cochrane Review) (last updated October 2003). In: The Cochrane Library, Issue 3, 2004. Chichester, UK: John Wiley & Sons, Ltd.

[21] Jadad AR, Gagliardi A. Rating health information on the internet: navigating to knowledge or to Babel? JAMA 1998; 279: 611-614.

[22] Eysenbach G. Consumer health informatics. BMJ 2000; 320: 1713-1716.

[23] 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: 569-573.

[24] Moult 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 information. Health Expectations 2004; 7: 165-175.

[25] van den Brink-Muinen A, Verhaak PF, Bensing JM et al. Doctor-patient communication in different European health care systems: Relevance and performance from the patients' perspective. Patient Education Counseling 2000; 39: 115-127.

[26] 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: 573-577.

4. Production of reports

4.1 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 (q.v. Preamble). The generation of these products will generally not diverge from the methodology set out in this document, except for the potential addition of specific procedures. Scientific reports follow the steps presented in Section 4.5. In exceptional situations, it may be necessary to react without delay to topical issues. In agreement with the Federal Joint Committee and, if necessary, other responsible public institutions such as the Federal Institute for Drugs and Medical Devices (*Bundesinstitut für Arzneimittel und Medizinprodukte*), an accelerated process will take place (q.v. Section 4.6).

The same methodological principles apply to the production of easily understandable health information for consumers and patients as to that of scientific reports, unless additional methodological requirements have been formulated in Section 3.

4.2 Selection of external experts

The Institute will regularly provide information via its website on upcoming assignments to be accomplished or on projects in planning. In accordance with its legal remit, the Institute will include external experts in its work, e.g. in the production or review of reports.

In relation to the research question posed, the following is required of persons or working groups interested in participating in the production of scientific reports:

- Firstly, disclosure of potential conflicts of interest;
- Secondly, demonstration of their professional medical and methodological expertise (presenting evidence of practical application, research and teaching experience) including any preliminary work;
- Thirdly, a credible guarantee that they possess sufficient human resources to complete the work to be done in an appropriate period of time.

A form for the disclosure of potential conflicts of interest will be available on the Institute's website. This form must be completed by every individual involved in the scientific work concerned and submitted to the Institute. The declaration on medical and methodological expertise can be made once for all participants involved (e.g. as a working group).

The Steering Committee will select cooperation partners to participate in the respective research project (or part of the project) from the circle of interested persons, with due consideration of the above criteria (potential conflicts of interests, medical and methodological expertise and adequate human resources).

External experts intending to employ third parties to complete all or part of the work agreed upon are obliged to declare this and to name the said parties. The same requirements concerning potential conflicts of interest apply to third parties as to the external experts themselves (q.v. Section 4.3).

The Institute and the external experts agree to respect the confidentiality of the terms of the contract, especially its scientific content or study results (or parts thereof), until final acceptance. The external experts additionally undertake to adhere to any data protection requirements that may apply.

4.3 Guarantee of scientific independence

4.3.1 Objectives

The scientific independence of the Institute and the products it is responsible for and publishes are legally founded in § 139a Social Code Book V, as well as in the statutes of the Institute's foundation. The term "independence" can thereby only be approximately applied, as the assessment of scientific independence can vary from one individual to the next. Insofar as the term "independence" can only be a relative objective, the term "transparency" must at the same time be introduced to give form, in a fashion comprehensible both internally and externally, to any decision-making processes and their findings, against the background of "relative independence".

4.3.2 Guarantee of external scientific independence

Before any contract is signed between the Institute and an external expert or institution to provide professional advice, to conduct studies or to produce a scientific report, the Steering Committee must decide whether any reservations exist as to potential conflicts of interest. For this purpose, all external experts and institutions must provide the Steering Committee with a list of all activities that may potentially influence their scientific independence (conflicts of interest; q.v. Section 4.2). In particular, the following criteria, based on the respective guidelines of scientific and medical journals, are viewed as conflicts of interest: All financial agreements, employment, advice, fees, reimbursed expert opinions, reimbursed travel expenses, patent applications and share ownership

within the previous three years that could have influenced the work commissioned, as well as all existing personal contacts with other persons or organisations that could influence the assignment in question [1]. This list of criteria is also included in the form supplied on the Institute's website (q.v. Section 4.2). This form will be updated whenever necessary. The downloadable version of the form on the website always applies.

4.3.3 Guarantee of internal scientific independence

Internal scientific independence is guaranteed as far as possible by the selection of staff. On being appointed, staff must credibly outline their previous activities and are obliged to cease all (external) assignments likely to call their scientific independence into question where their work for the Institute is concerned. The Institute's research associates are 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 Director or administration. External assignments 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 Steering Committee will decide on a case-by-case basis whether a member of staff must be excluded from a certain activity or project on grounds of suspected bias.

References

[1] A. James, R. Horton. The Lancet's policy on conflicts of interest. Lancet 2003; 361: 8-9.

4.4 Report plan

The report plan, comparable to the study protocol of a clinical trial, contains the exact research question, including the target criteria (e.g. patient-relevant endpoints), as well as a description of the methodology of the literature search and the evaluation of the information acquired (including the inclusion and exclusion criteria used in this procedure). In addition, details on subsequent steps in the production of the final scientific report, such as the conduct of the peer review and the date of publication are described, insofar as they have been or can be defined at the time of production of the report plan. Where external advisors are included in the projects, they will be given the opportunity to go into more detail on certain aspects of the report plan, in agreement with the Institute's project group. The report plan will then be published on the Institute's website. Justifiable

amendments to the report plan are possible, as to the protocol of a clinical study, and will be published on the website.

The information provided in the report plan, especially on the inclusion/exclusion criteria for relevant information (e.g. scientific literature), is of crucial importance for the criteria that are to be met by future statements.

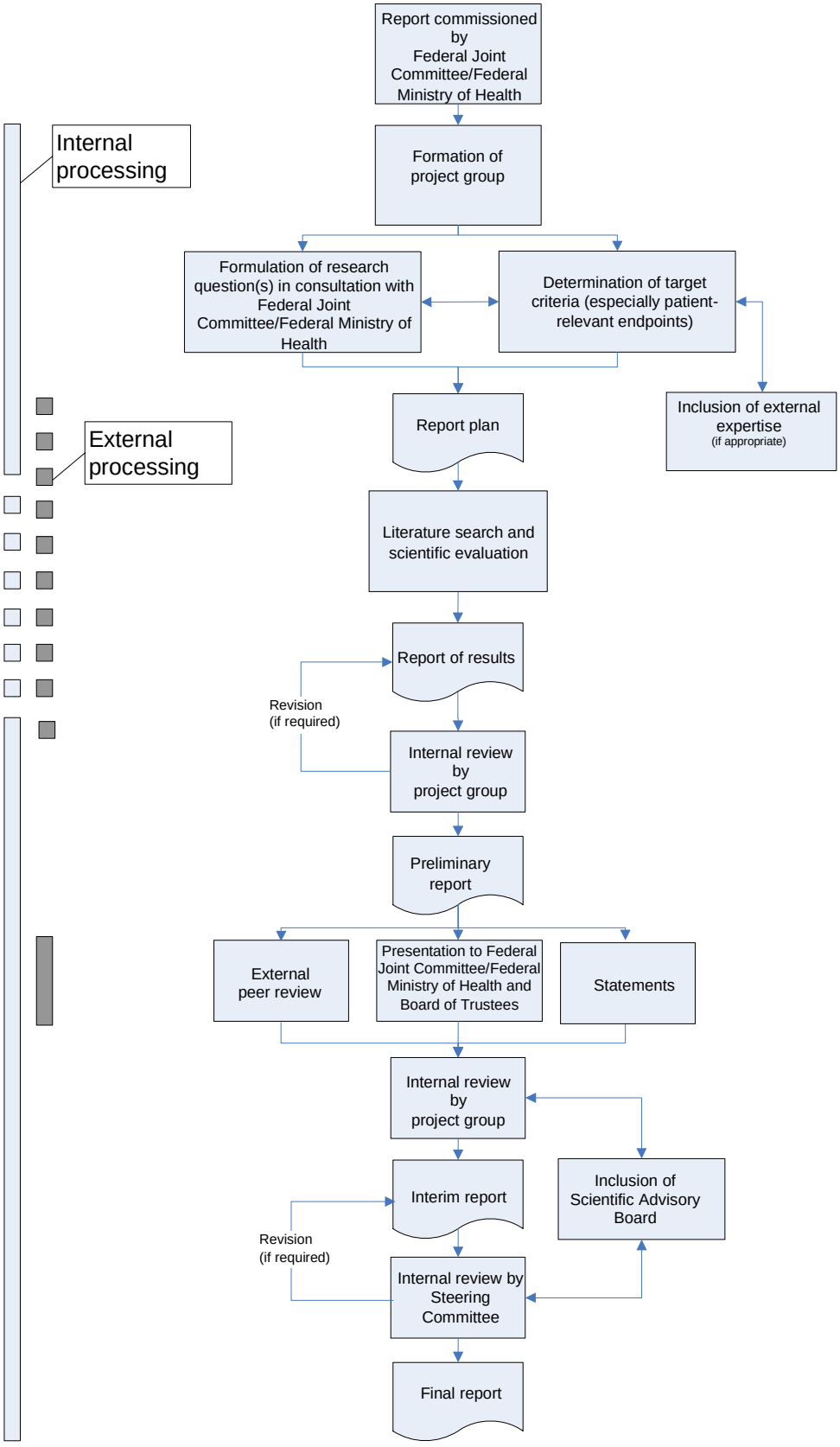
In principle, statements will be possible relating to the completeness of the literature search and to the correctness of the scientific evaluation of the information acquired.

4.5 Structure of report production

The procedure for report production is outlined in Figure 4.

After a report has been commissioned by the Federal Joint Committee or the Federal Ministry of Health, a project group is formed under management of the department concerned and a project manager appointed to take responsibility. The composition of the group is not fixed at this point, as the need for changes may arise during subsequent stages of work. If necessary, specification of the assignment then takes place in agreement with the relevant committees of the Federal Joint Committee or the Federal Ministry of Health. This also includes the (rough) definition of target criteria, especially patient-relevant endpoints. If required, this definition is refined by the project group, considering external advice where appropriate. Finally, the project group produces the report plan (q.v. Section 4.4), which provides the basis for the subsequent literature search and scientific evaluation.

Figure 4: Flow diagram of a report production



The literature search and the scientific evaluation of the information acquired can, at least in part, be conducted by one or more external experts who have demonstrated their suitability according to the selection criteria described in Section 4.2. In individual cases, consultation will be necessary with the external expert(s) on the steps described in the report plan, including a more precise formulation of the research question posed. The report plan will then be published on the Institute's website. One or more reports of results (depending on the number of external experts) will be produced from the literature search and subsequent scientific evaluation and, after internal review by the project groups, summarised as a preliminary report.

As a rule, this preliminary report will be simultaneously:

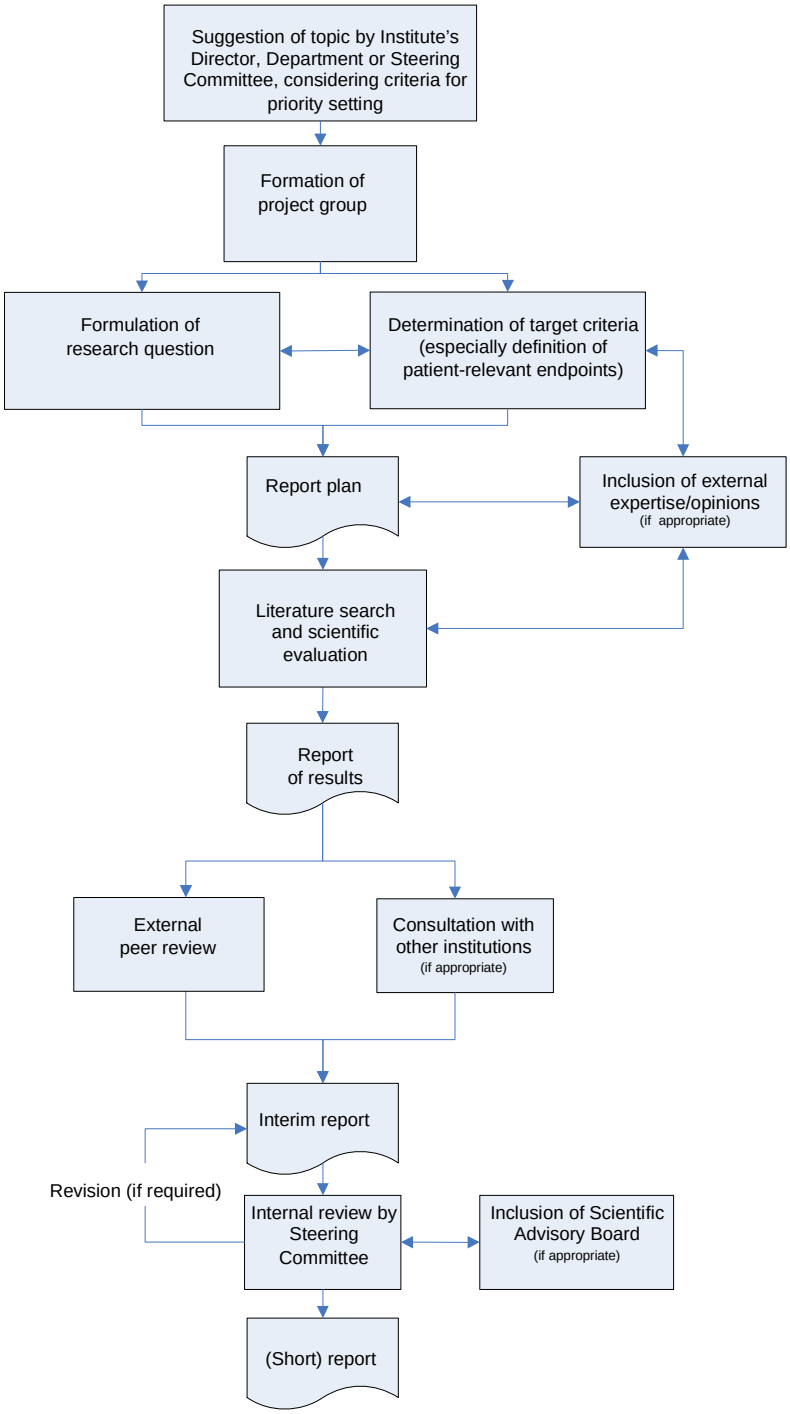
- Forwarded to one or more external peer reviewers with proven methodological and/or professional competence;
- Forwarded to the Federal Joint Committee or Federal Ministry of Health, and Board of Trustees, also to verify the completeness of the work originally commissioned;
- Published on the Institute's website to solicit statements fulfilling the criteria outlined there.

According to the nature and extent of the statements received, including the results of peer reviews and statements by the Federal Joint Committee or Federal Ministry of Health, and Board of Trustees. A discussion including the external experts may then be held, if it is felt this could improve the scientific quality of the final scientific report. The project group will subsequently prepare an interim report to be presented to the Steering Committee for internal review and the final scientific report will then be forwarded to the Federal Joint Committee or Federal Ministry of Health. After their agreement, at the latest after a latency period previously determined in the report plan, the final scientific report will be published by the Institute.

4.6 Accelerated production of reports

In exceptional situations it may be necessary to react without delay to topical issues. This requires a shortened advisory phase and rapid decision-making within the Steering Committee. Furthermore, consultation with external institutions may be required. Due to pressure of time and the corresponding danger of imprecise or incorrect statements, orientation towards previously defined procedures is essential to minimize errors. The procedure of accelerated report production is shown in Figure 5 in analogy to the normal procedure (q.v. Section 4.5).

Figure 5: Flow diagram of an accelerated report production



4.7 Publication of scientific reports

The Institute's essential task is to perform a careful evaluation of available evidence and utilise its findings with the aim of improving individual and general health care.

The Institute sees itself as having the responsibility to ensure the transparency of its findings and to publish all results obtained from reports produced by the Institute or in the Institute's name.

To guarantee the Institute's independence, it must be ruled out that the Federal Joint Committee, Federal Ministry of Health or any other interested party has any influence on the content of the reports, as this could lead to conflation of scientific results with political or other vested interests. At the same time it must be avoided that the Institute for its part withholds certain findings. To prevent this, all the results produced by the Institute according to its legal obligations, including the report plan and a summary of the underlying materials used in the process of the scientific report, are to be published without delay.

If not otherwise agreed, all copyright lies with the Institute.

The time frame for publication is included in the report plan.

4.8 Literature search

The information, which together with its scientific evaluation forms the basis of the Institute's scientific reports, can take a variety of forms, e.g. scientific studies or data collections. The handling of raw data is outlined in Section 1.13. In the following section, the process of a topic-related literature search is described, referring to research conducted both by the Institute's staff and external experts.

4.8.1 General principles of topic-related literature search

The aim of a topic-related literature search is to identify all relevant publications that contribute to a greater understanding of the respective research question. The methodology of systematic literature searches therefore follows the general principle that the topic-related literature search concerned must at least fulfil all the criteria which, according to the results research on this issue, have or could have an important influence on the result (i.e. on the answer to the research question posed).

These criteria include in particular:

- The selection of data sources (public databases, private databases, manual searches in selected journals, contacts with experts/industry/patient organisations, etc.);
- Search techniques relating to the selection of publication or study types (RCTs, case reports, etc.);

- Search techniques relating to the medical criteria determined by the research question posed (target population, type of intervention, endpoints, etc.);
- Search techniques relating to formal characteristics of the publication (publication of abstracts, language, etc.);
- The conduct of the search itself (e.g. by one person or two persons working independently of each other).

Examples are provided by publications [1-8].

As the relevance of these criteria can vary with different research questions, the selection of the criteria to be applied will be on a case-by-case basis. If no valid research results for the individual criteria with regard to the research question are available, or if these have not yet been reviewed and evaluated by the Institute itself, it will be primarily assessed whether established strategies of international or national working groups (e.g. the Cochrane Collaboration), where wide experience already exists, can be applied.

4.8.2 Search techniques for primary publications in bibliographical databases

The literature search consists of three parts:

1. Formulation of the search strategy;
2. Application of the search strategy (primary search);
3. Selection of relevant publications from the results of the primary search.

The formulation of the search strategy is made on the basis of the research question agreed upon with the Federal Joint Committee or Federal Ministry of Health and on the target criteria defined. Where appropriate, this includes the results obtained from the procedure of determining patient-related effects and measures (q.v. Section 1.5).

In particular it involves:

a) Selection of data sources;

b) Selection of search key words and search criteria for database searches;

c) Specification as to which studies are to be regarded as relevant, giving details of inclusion/exclusion criteria.

In each case, it will be reviewed whether consultation with external experts is useful. This can be the case particularly if no professional expertise with regard to the research question posed is avail-

able in the Institute. The provisionally formulated search strategy is to be presented to the Institute's project manager and subsequently agreed on.

The application of the search strategy (database query) can be made by one person.

In this context, it is often useful to search first for preliminary work by other working groups, e.g. by searching in specialised databases (Cochrane Database of Systematic Reviews, HTA databases, DARE database, etc.). Recourse can, if required, be taken to the search for primary publications conducted for these reviews, insofar as relevant reviews can thereby be identified. The prerequisite for this is that the respective search is in line with the Institute's methodology, and the transferability of the results to the research question posed is assured, particularly with regard to the inclusion/exclusion criteria in the report plan. In these cases, it may also be necessary to conduct a search for primary publications. This search then refers to the publication period not considered by the reviews. If no relevant reviews are identified, a search is conducted for primary publications for the whole publication period relevant to the research question posed.

4.8.3 Other data sources for primary searches

Besides bibliographical database searches, it can be useful to conduct a manual search in selected professional journals. This is to be decided upon from case to case. Moreover, the materials for consideration provided by the Federal Joint Committee or Federal Ministry of Health should also be regarded as a component in a primary search. These materials will then be treated according to the same criteria applying to the other information searches and evaluation processes.

In addition, depending on the research question posed, further data sources are of considerable importance, e.g. study registers, abstract volumes of scientific congresses, or in the case of drug evaluations, drug approval databases or correspondence, insofar as these are available to the Institute.

If indications as to unpublished or not fully published studies can be derived from these data sources, the Institute will contact the respective principal investigator and/or sponsors and request access to all information required for evaluation of the studies. The Institute will not accept any obligations regarding confidentiality in this regard. Scientists, institutions, commercial enterprises and other persons or groups who are not prepared in this respect to provide the information needed by the Institute will in consequence not be allowed an opportunity to make statements in the subsequent report production process.

4.8.4 Selection of relevant publications

The selection of relevant publications from the results of the primary search is usually made in two steps:

- Perusal of the abstracts with the aim of excluding evidently irrelevant publications;
- Perusal of the full papers of the remaining potentially relevant publications.

Both steps are, as a matter of principle, to be performed by two persons working independently of each other. After completion of these steps the search findings are to be presented to the project leader before evaluation of the literature; in individual cases, an expert can again be consulted to assess the completeness of the literature. If modification or extension of the literature search is then advised, this is to be initiated according to the principles mentioned and the results presented once again before evaluation of the literature.

4.8.5 Documentation

All the steps taken in literature search are to be fully documented. This especially includes:

- The search strategy, including the grounds for the selection of data sources and the search techniques applied.
- In addition, if bibliographical database queries are conducted:
    - Evidence of the search, preferably in the shape of a screen shot of the research history;
    - A list of the primary results (quotations, abstracts) obtained by applying the search strategy;
    - A list of the publications judged relevant to the research question posed, after perusal of the primary results;
    - A list of the literature sources noted in the primary result but not judged relevant, including the grounds for their exclusion.
- Copies of correspondence (if personal contact with manufacturers, experts or professional societies was established).
- The date or period of the search.

The type of documentation (paper copy, database format, etc.) is to be clarified beforehand with the project manager and should as a matter of principle follow the general guidelines laid down by the Institute.

References

[1] 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: 109-125.

[2] 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: 1228-1231.

[3] Moher D, Pham B, Klassen TP et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol 2000; 53: 964-972

[4] 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: 159-163

[5] 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: 115-123.

[6] 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: 1109-1119.

[7] Sampson M, Barrowman NJ, Moher D et al. Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol 2003; 56: 943-955.

[8] 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: 44-51.

4.8.6 Literature screening

Besides topic-related retrospective searches, early detection and evaluation of current and relevant publications are necessary, based on the systematic monitoring of important data sources. The term "data sources" includes not only professional journals, but also the lay press, daily, weekly and monthly press, electronic media, etc.

A publication is to be 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 great public interest is discussed in the publication;
- The publication touches on a topic high up on the Institute's internal priority list.

The Steering Committee decides which data sources (professional journals, websites, etc.) are important and should therefore be regularly monitored. The departments evaluate the data sources allocated to them and forward relevant publications to the other departments. A publication may evoke an ad hoc evaluation and an official statement from the Institute. The respective procedure follows defined standards (q.v. Section 4.6).

4.9 Evidence to support claims of effectiveness

Evidence of the alleged favourable effect of a medical procedure is to be supplied by those institutions or persons promoting this position to the Institute. A "favourable effect" in this regard means the predominance of patient-relevant advantages compared with any risks involved in the application of the medical procedure in question. Those who propagate the procedure must supply evidence of the presence of the desired patient-relevant effect and evidence of the exclusion of relevant risks.[v] It is therefore in the interests of these institutions and persons to supply the Institute with all the available data for evaluation by the Institute. It is not the Institute's responsibility to prove the absence of a patient-relevant effect[w] or the presence of relevant risks. However, the Institute can – as far as the available data allow this – provide such evidence in its evaluations (according to the level of scientific knowledge prevailing at the time).

4.10 Priority-setting

A major part of the Institute's resources will be used in response to external assignments awarded by the Federal Joint Committee or Federal Ministry of Health; it is their responsibility to set priorities in consultation with the Institute for the start of work on assignments. Moreover, the Institute needs to set internal priorities for processing assignments.

The Institute also needs to set its own priorities with regard to:

- Internal projects of the individual departments;
- Production of health information by the Department of Patient Information and Research;
- Assignments awarded to external experts and institutions;

[v] As a rule, this will only be possible by means of inferential statistics, that is, as a rejection of the hypothesis (subject to error [significance level]) that an undesired effect exceeds a more than irrelevant order of magnitude.

[w] Q.v. previous footnote.

- Methodological and scientific work conducted by the Institute.

Priority-setting may involve particular issues that need to be considered in individual areas. Each department has the option of developing specific priority-setting procedures based on the specialised activities of the department. However, priority-setting processes for the Steering Committee and individual departments will aim to be straightforward and consistent with the general rationale, values and methods set out below.

Priority-setting is not carried out following a rigid process. Nevertheless, it is essential to aim for fairness and transparency in the distribution of intellectual and financial resources in a public organisation such as the Institute. Priority-setting in research activities and in the systematic evaluation of medical techniques and technologies (e.g. HTA) needs transparent mechanisms, and must be in agreement with the rationale, values, methods and criteria of the Institute. By this means, it is ensured that the activities and priorities of the Institute can be reviewed and understood externally.

4.10.1 Background to the Institute's priority-setting

Through its activities, the Institute aims to make a beneficial impact on the health of the German population and contribute to the development of public and scientific understanding of health.

In decision-making procedures for the work undertaken, the Institute will consider the:

- Proportion of the population likely to benefit from the work the Institute chooses to undertake;
- Burden of disease (including cost), currently and in the future, for individuals and society, particularly for disadvantaged groups;
- National health priorities;
- Ability of the Institute to influence clinical practice, clinical decisions and the health status of the population;
- Resources required by the Institute to address the activity effectively;
- Potential of a beneficial impact on society or the health care system, considering the principles of equity;
- Potential risks;
- Unique contributions the Institute could make, including assessing what others are doing or who else could potentially undertake certain activities;

- Contribution of the activity to the quality assurance of the Institute's work;
- Potential contribution of the activity to the evolution of scientific knowledge.

4.10.2 Processes for priority-setting

A wide variety of models and processes in use for priority-setting in similar organisations to the Institute generally have some features in common. These form the basis of the Institute's model of priority-setting and include:

- Collection of data and opinions to provide a basis for decision-making, and their ongoing documentation;
- Application of the relevant criteria developed by the Institute or its departments;
- Provision of a report of the data and prevailing opinions, together with a recommendation, to the Steering Committee, which will make the final decision and document the reason for making this decision.

4.11 Production times of reports

The evaluation of a method or intervention is possible at any time. There will be no general directive that the production of a scientific report by the Institute will take place, at the earliest, following passage of a certain period of time after approval or establishment of a method or intervention. If, in the case of a report ahead of schedule, there is great uncertainty about results due to a lack of long-term studies, this will be described in accordance with general working procedures.