IQWiG Reports – Commission No. MB13-01

Assessment and analysis of studies on rare diseases[1]

Executive Summary

Publishing details

Publisher:

Institute for Quality and Efficiency in Health Care

Topic:

Assessment and analysis of studies on rare diseases

Commissioning agency:

Federal Ministry of Health (BMG)

Commission awarded on:

10 December 2013

Internal Commission No.:

MB13-01

Address of publisher:

Institute for Quality and Efficiency in Health Care
Im Mediapark 8 (KölnTurm)
50670 Cologne
Germany

Tel.: +49 (0)221 – 35685-0
Fax: +49 (0)221 – 35685-1
E-Mail: berichte@iqwig.de
Internet: www.iqwig.de

External experts:

▪ Norbert Benda (Federal Institute for Drugs and Medical Devices [BfArM])

This rapid report underwent an external review. IQWiG thanks the external reviewer for his collaboration in the project. However, the reviewer was not involved in the actual preparation of the rapid report. Individual sections and conclusions in the rapid report therefore do not necessarily reflect his opinion. The information on conflicts of interest provided by the external reviewer is presented in Appendix B of the full report. No conflicts of interest were detected that could endanger professional independence with regard to the work on the present commission.

The responsibility for the contents of the report lies solely with IQWiG.

IQWiG employees[2]:

▪ Ulrich Grouven

▪ Ralf Bender

▪ Ulrike Lampert

▪ Stefan Lange

▪ Stefan Sauerland

▪ Ulrich Siering

▪ Volker Vervölgyi

---

[2] Due to legal data protection regulations, employees have the right not to be named.

Executive summary

In its letter of 10 December 2013, the Federal Ministry of Health (BMG) commissioned the Institute for Quality and Efficiency in Health Care (IQWiG) to produce a rapid report on the topic "assessment and analysis of studies on rare diseases".

Research question

The aims of this rapid report were

- to provide an expert opinion on methodological aspects in the conduct and analysis and on the assessment of the certainty of results of studies on rare diseases

- to describe the underlying studies for the approval of orphan drugs in Europe

Methods

Methodological expert opinion

The particular problems in the conduct and assessment of studies on rare diseases are presented in the methodological expert opinion. Rare diseases ($\leq 5/10\,000$) are differentiated from very rare diseases ($< 2/100\,000$ people). Criteria for the certainty of results and the reliability of the conclusions of study designs commonly used or proposed for rare diseases are presented and discussed as the basis of benefit assessments.

Empirical investigation on the underlying studies for the approval of orphan drugs in Europe

The search for the empirical investigation of the underlying studies for the approval of orphan drugs in Europe was conducted in Orphanet. European Public Assessment Reports (EPARs) on the approval studies were used for the extraction of relevant data on the orphan drugs identified. Relevant characteristics of the drug and of the underlying approval studies were extracted and analysed using descriptive statistical methods.

Results

Methodological expert opinion

Basic methodological aspects in the assessment and analysis of studies

According to international agreement the scientific basis of evidence-based medicine (EBM) is to systematically identify the clinical studies suitable for answering a clearly defined research question (according to the PICOS [patient, intervention, comparator, outcome, setting/study design] method), to assess the certainty of the results of the studies identified in a comprehensible way and to make a summarizing assessment under consideration of the certainty of results. 4 components play a role with regard to the reliability of the conclusions for answering the underlying research question according to the PICOS method. The certainty of the results of the clinical studies identified is mainly based on 3 components:

1) a qualitative component characterized by the risk of bias of the studies to be assessed (internal validity)

2) a quantitative component, determined by sample size(s), but also by the variance of the observations (precision of the results)

3) the size of observed differences (effect size)

There is one additional component with regard to the reliability of the conclusions:

4) the external validity (or applicability), i.e. to what extent the study conditions represent the specific research question

For the assessment of the risk of bias, 6 categories of bias can be differentiated, of which the first 4 categories deal with differences between the intervention groups to be compared: selection bias, performance bias, detection bias, attrition bias, reporting bias and other bias mechanisms, e.g. the use of inadequate statistical methods.

There are 3 main strategies that can be used to avoid the first 4 bias mechanisms mentioned: randomized and concealed allocation of patients to the intervention groups to be compared, blinding of the interventions for patients and study staff (also called "double-blinding"), and the analysis of all patients included in the study according to their allocated intervention (intention to treat [ITT] principle).

Barriers against randomized studies

There are currently no known alternatives to randomization, particularly if the aim is to avoid selection bias. However, occasionally arguments against randomized controlled trials (RCTs) are put forward, which concern their feasibility, often due to unspecified ethical concerns or logistical problems.

A controlled study is in fact ethically doubtful if the benefit (or added benefit in comparison with standard treatment) of an intervention in a specific therapeutic indication is more or less proven. However, then studies with the aim of gaining knowledge on the benefit or added benefit are unnecessary anyway, and not only in case of rare diseases.

There are 2 scenarios that might be arguments against randomization: Conducting an adequate central randomization maintaining concealment is too involved or a certain intervention is not sufficiently available for various reasons (e.g. in terms of quantity or because expertise in its use is lacking) and cannot be made available in the short term.

Situation of rare diseases

It obviously becomes more and more difficult to achieve a sufficient sample size for answering a specific research question the rarer this disease is. Hence the certainty of results in rare diseases is particularly impaired with regard to precision. Among other things, this affects the type 2 error (not finding an existing difference). The probability of this error

depends on the actual difference (on the "truth") and can therefore only be controlled to a limited extent by the sample size based on assumptions on this "truth". In contrast, the type 1 error (erroneously assuming a difference although none exists) can at least theoretically be controlled also in case of small sample sizes by specifying the significance level.

In contrast, the 2 other components of the certainty of results and reliability of the conclusions that can be influenced (internal and external validity) are non-quantifiable and may result in systematic errors (deviations from the "truth"). Particularly with regard to the risk of bias, such weaknesses in the study design can usually not be compensated by statistical methods, however sophisticated they may be. Hence in the context of the assessment of medical interventions in rare diseases it seems reasonable to first search for possibilities to use optimized statistical methods to compensate for the limitations caused by possibly decreased precision.

Proposals for the design of randomized treatment studies in rare diseases

The starting point for proposed modifications is the classic randomized parallel group design, in which patients are randomly allocated to one of 2 or more different interventions and are observed for a predefined period of time. The methods and designs can be roughly divided into 5 groups, which cannot be clearly differentiated however:

- designs to reduce the variance between the observations, e.g. stratification, regression analysis

- methods with adaptations of the design in the course of the study, e.g. sequential designs

- methods in which prior information (outside the study) are included in the statistical analysis (Bayesian methods)

- designs in which all patients receive the treatment under investigation

- other designs, e.g. adaptive randomization

Certainty of results of non-randomized studies

Non-randomized studies inherently entail major limitations regarding internal validity (first component of the certainty of results). Currently no method is known that is sufficient to overcome these limitations. Depending on the reason for the possible dispensing with randomization, there are immediate consequences for non-randomized studies used instead. If ethical concerns are raised, any parallel comparative studies are out of the question, and, in principle, only historical controls remain.

Historical control studies rank very low in the commonly used hierarchies of evidence. Besides the usual bias mechanisms (e.g. selection bias), the factor "time" or "chronology" is an additional confounder. Results from historical control studies only allow to draw conclusions in the sense of an intervention effect only if there is (nearly) a reversal of a more or less deterministic course of the disease ("dramatic effect"). In the literature, a risk that is

increased by the factor 10 in the intervention group in comparison with the control intervention in combination with an (adequate) statistical test with a significance level of 1% is the proposed criterion for the presence of a dramatic effect.

Observational studies, which are either aimed to a lesser degree at answering the question of the effects of interventions or the database of which does not primarily serve to gain medical-clinical questions, are differentiated from prospective intervention studies. In the commonly used hierarchies of evidence, they are therefore ranked below randomized and non-randomized intervention studies in their informative value.

Recommendations

No scientific rationale can be derived for using different approaches in the assessment of medical interventions for rare and non-rare diseases. Studies with low precision or with insufficient protective mechanisms against potentially biasing factors or with deviations of the study conditions from the research question of interest (e.g. use of surrogate outcomes instead of patient-relevant outcomes) have the same consequences on the reliability of the conclusions in rare and non-rare diseases. Conversely, no specific designs and statistical methods exist for rare diseases that could not also be relevant for (more) common diseases.

In the context of the assessment of medical interventions in rare diseases it will therefore be (particularly) necessary to choose a methodological approach that is as efficient as possible, creating efficient structures at the same time. Moreover, it may be necessary, particularly in very rare diseases, to make compromises with regard to the reliability of the conclusions, which may also be due to external (political) requirements.

In the framework of the Act on the Reform of the Market for Medicinal Products (AMNOG), German legislation defined political guidelines by stipulating that the added benefit of drugs for the treatment of rare diseases (orphan drugs) is regarded as proven by their approval. However, this only applies for as long as the turnover of such a drug in the statutory health insurance does not exceed 50 million euros within 12 calendar months. On the one hand, this is justified by the content-related argument that it can be regularly assumed that there is no therapeutically equivalent treatment alternative. On the other hand, the justification also shows the political guideline by linking the added benefit to the turnover of the drug. From a scientific point of view it cannot be justified why content criteria, which – if they do apply – make the assessment (of the [added] benefit) futile, are (supposed to be) no longer valid above a certain turnover volume.

In rare diseases it becomes increasingly difficult with decreasing frequency of the disease to sufficiently account for the quantitative component of the certainty of results. Hence it can be stated, although trivial in principle, that working in supraregional and supranational networks is of particular importance for the clinical, patient-oriented research of rare diseases. A strong appeal can be derived from this for disease registries with clear quality criteria regarding

completeness and wholeness as the basis for high-quality clinical, particularly non-randomized studies.

When there are inherent limitations regarding one component of the certainty of results, it seems not to be productive to compromise other components as well. Hence those strategies of the ones described above would have to be preferred that allow reduced sample size while largely maintaining internal and external validity, i.e. sequential designs.

In very rare diseases, it might be considered as an alternative option or in addition to the commonly used methodological approach to allow a larger statistical significance level for regulatory decisions, e.g. to raise the usual two-sided significance level from the usual 5% to 10%. The advantage of this approach would be that the probability of errors could at least be quantified. Such an approach can also be understood as an approximation to Bayesian methods: Whether data from a randomized study with a small sample size with increased significance level is used for making a decision, or whether the usual significance level is maintained, e.g. in combination with an optimistic prior, probably leads to comparable results regarding the resulting decisions.

In descending priority, limitations of external validity could also be accepted, e.g. by including data from similar therapeutic indications or by using established surrogate outcomes within composite outcomes. Using adaptive designs will also be generally associated with limitations of external validity. At least internal validity would be maintained, however.

For logical reasons, abandoning internal validity by not using randomization would rank lowest in such a hierarchical approach. One of the key prerequisites to be able to use them for (regulatory) decisions would be that the underlying data are from a disease registry with the quality criteria mentioned above or that an observed effect is so large that it can no longer be explained by bias alone.

Results of the empirical investigation on the underlying studies for the approval of orphan drugs in Europe

85 drugs with European orphan drug designation and European market approval from 2001 to 2013 were identified. The approval of the 85 drugs was based on 125 main studies (without 6 approvals based on literature reviews), including 82 RCTs.

Analyses at drug level

All approvals

58 of the 85 drugs identified (68%) are used for the treatment of rare diseases, and 27 drugs (32%) are used for the treatment of very rare diseases. The approval of 59 drugs (69%) was based on RCTs (55 exclusively on RCTs, and 4 on RCTs in combination with non-RCTs). Non-RCTs were the basis of approval for 20 drugs (24%).

Approvals without literature reviews

The approval of the 79 orphan drugs without literature reviews was based on 1 to 5 main studies each. In very rare diseases, the approval was based on a maximum of 3 studies. The proportion of approvals based on data from RCTs was approximately 75%. The number of patients in the approval studies per drug was between 27 and 2961 patients (median: 165). Approximately 70% of all patients were treated in RCTs. Patient-relevant outcomes were considered in the studies for 66 drugs (84%) (for 31 orphan drugs as primary outcome, for 57 as secondary outcome). The majority of the studies were multicentre, multinational, and multicontinental. In the majority of the cases, the study characteristics presented showed no noticeable differences between rare and very rare diseases.

Analyses at study level

The descriptive analyses on characteristics and methods of the approval studies were based on all 82 RCTs (66%) among the 125 main studies (without approvals based on literature reviews) of the approved orphan drugs (1 to 3 RCTs per drug). Between 8 and 769 patients (median: 160.5) were included in the RCTs.

The proportion of double-blind RCTs was 74%. An active comparator was used in 28% of the studies. With 6% of the studies, non-inferiority and equivalence studies were only used in exceptional cases. There were no noticeable differences between the studies on rare and very rare diseases.

Methods to control confounders were used in 52% of the RCTs. Crossover designs and sequential methods were used in 5% and 12% of the studies, respectively. Significance level was increased in 1 study. Adaptive randomization, Bayesian methods or other specific randomization designs were not used in any of the approval studies. Overall, specific methods of analysis were used less frequently in the studies on very rare diseases; sequential methods were not used at all.

Conclusions

No scientific rationale can be derived for using different approaches in the assessment of medical interventions for rare versus non-rare diseases. Conversely, there are also no specific designs and statistical methods for rare diseases that could not also be relevant for (more) common diseases. This applies in the same way to drug and non-drug interventions.

Approvals and approval studies for orphan drugs, also in very rare diseases, are largely based on conventional (randomized) designs so that the general feasibility is not in question.

However, in the context of the assessment of medical interventions in rare and particularly in very rare diseases, it may be necessary or politically required to make compromises with regard to the reliability of the conclusions. Such compromises are principally conceivable at 3 levels:

The (statistical) significance level could be raised above the usual value of (two-sided) 5% (compromise regarding the required precision).

In descending priority, limitations of external validity could also be accepted, e.g. by including data from similar therapeutic indications or by using established surrogate outcomes within composite outcomes.

For logical reasons, abandoning internal validity by not using randomization would rank lowest in a hierarchical approach. One of the key prerequisites to be able to use them for (regulatory) decisions would be that the underlying data are from a disease registry with excellent quality regarding completeness and wholeness.

Keywords: rare diseases, benefit assessment

The full report (German version) is published under https://www.iqwig.de/de/projekte-ergebnisse/projekte/medizinische-biometrie/mb13-01-bewertung-und-auswertung-von-studien-bei-seltenen-erkrankungen-rapid-report.3685.html#overview.