IQWiG Reports – Commission No. A19-43

Concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a Social Code Book V (SGB V)[1]

Rapid report

[1] Translation of the rapid report A19-43 *Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach §35a SGB V* (Version 1.0; Status: 10 January 2020). Please note: This translation is provided as a service by IQWiG to English-language readers. However, solely the German original text is absolutely authoritative and legally binding.

Publishing details

Publisher:

Institute for Quality and Efficiency in Health Care

Topic:

Concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a Social Code Book V (SGB V)

Commissioning agency:

Federal Joint Committee

Commission awarded on:

2 May 2019

Internal Commission No.:

A19-43

Address of publisher:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8
50670 Köln
Germany

Phone: +49 221 35685-0
Fax: +49 221 35685-1
E-mail: berichte@iqwig.de
Internet: www.iqwig.de

IQWiG employees

- Sascha Abbas
- Ralf Bender
- Raphaela Gorris
- Elke Hausner
- Katharina Hirsch
- Thomas Kaiser
- Stefan Lange
- Jörg Lauterberg
- Guido Skipka
- Beate Wieseler
- Jürgen Windeler

Keywords: Product Surveillance – Postmarketing, Registries, Benefit Assessment, Concept

Executive summary

On 2 May 2019, the Federal Joint Committee (G-BA) commissioned the Institute for Quality and Efficiency in Health Care (IQWiG) according to §139a (3) Social Code Book V (SGB V), to develop scientific concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a SGB V.

Research question

The aims of this work are

- The creation of an overview of possible concepts for generating and analysing routine practice data. In particular, data collections that are not classified as randomized controlled trials (RCTs) should also be considered.
- The assessment of the identified concepts of data generation and their analysis with regard to their suitability to answer the research question of a benefit assessment according to §35a SGB V, especially with regard to the possibility of quantifying the added benefit of a new drug.
- The specification of criteria for data quality and the methodological requirements for the data collected within the framework of the respective generation of data. In this regard, the measures required to ensure data quality should also be addressed.
- The definition of requirements for reporting, as well as for the preparation and structure and the statistical analysis of the data collected within the framework of the respective generation of data.

Methods

Information retrieval and assessment

According to the project outline, the development of the concept for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a SGB V was supported by 3 modules:

- Empirical information from the benefit assessments of drugs according to the Act on the Reform of the Market for Medicinal Products (AMNOG[2], §35a SGB V).
- Exploratory literature search for scientific questions arising as part of the conceptual work (e.g. on the informative value of studies without randomization, depending on existing data constellations).
- Interviews with registry experts on criteria for the quality and methodological requirements of the data collected within the framework of the respective generation of data.

[2] *Arzneimittelmarktneuordnungsgesetz*

Determination of quality criteria for patient registries

During the course of the project, it became apparent that, for the generation of routine practice data for the benefit assessment of drugs, besides specifically conducting studies to generate data (study-specific data collection), data collection from registries is the second relevant data collection tool. The specification of quality criteria for the data collected was therefore limited to registries. The basis for the description of quality criteria was formed by the above-mentioned interviews with registry experts as well as a compilation of quality criteria for patient registries from national and international recommendations.

Results

Routine practice data for the benefit assessment of drugs are defined as follows:

- routine practice data are collected from the patient populations for which there is a therapeutic indication for the drug of interest within the scope of its marketing authorization
- in the collection of routine practice data, patients are treated without specific requirements

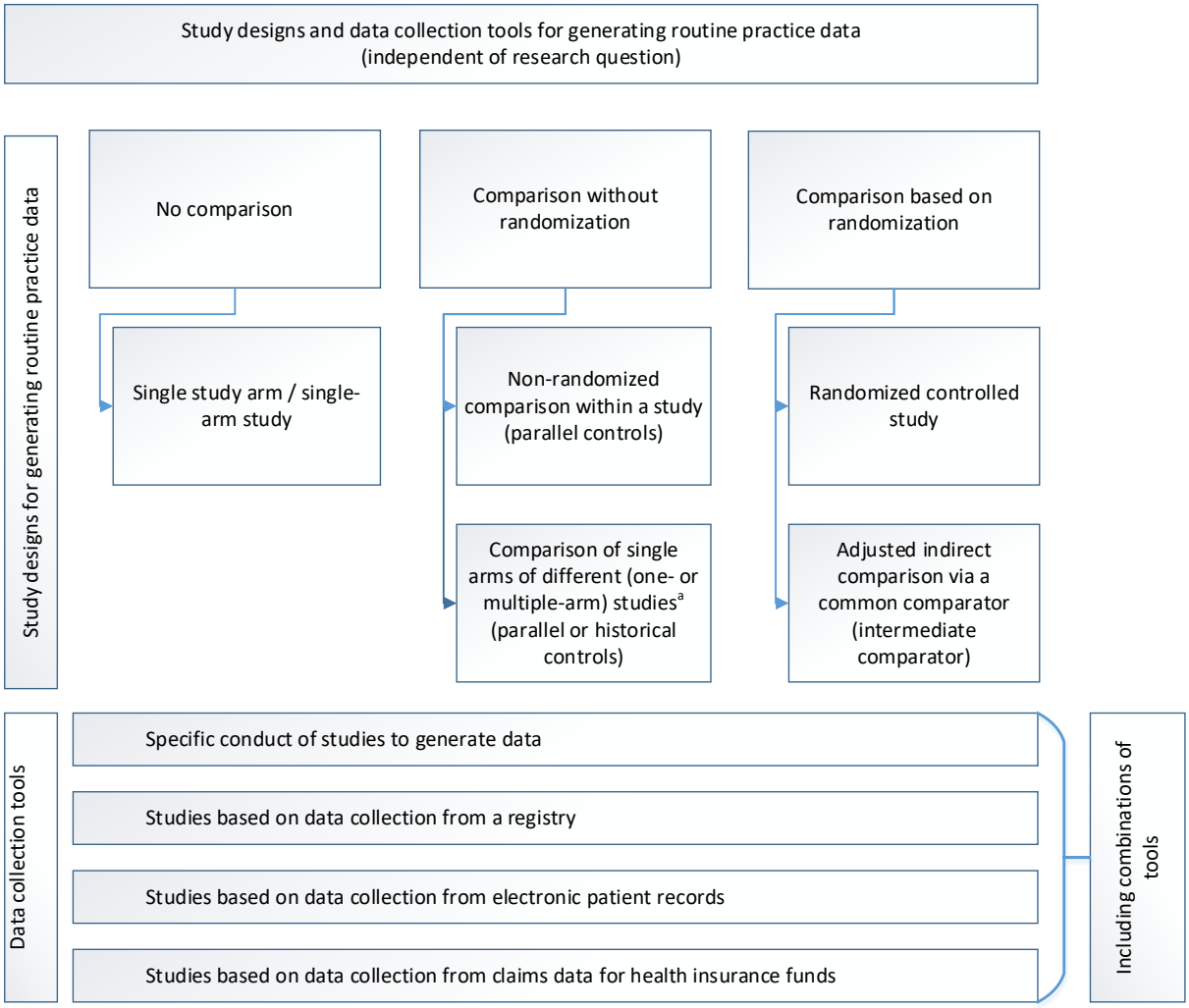
Since drug assessment according to SGB V is concerned with patient care in Germany, routine practice data must meet the two criteria mentioned above so that conclusions can be drawn for health care in Germany.

The definition of routine practice data implies neither a specific study design nor a specific data collection tool.

The goal of collecting routine practice data does not require that data collection be limited to data collected in routine practice per se. Rather, such a misconceived restriction of data collection would jeopardize the goal of the benefit assessment. The benefit assessment regularly requires data that are not collected in routine practice for all patients (e.g. data on health-related quality of life, symptoms or side effects). For use in a benefit assessment, routine practice data must also be sufficiently valid and structured.

Overview of study designs and data collection tools with the aim of generating routine practice data

The following figure provides an overview of study designs and data collection tools that can be used to generate routine practice data. The upper part of the figure describes the study designs that are basically conceivable, depending on the possible type of comparison of interventions. The lower part names the tools that can be used to collect routine practice data in studies with different designs. It becomes clear that the various data collection tools can generally be used for all study designs.



a: including studies on the spontaneous course of the disease

Figure 1: Study designs and data collection tools for generating routine practice data

Routine practice data in benefit assessments

If routine practice data are to be used for a benefit assessment, it must be taken into account that the basis of any conclusion on the effects of interventions is a comparison. This is because only on the basis of a comparison is it possible to distinguish between "**after** intervention A" and "**due** to intervention A"; this distinction is necessary for a causal conclusion.

It follows from these deliberations that the sole consideration of single-arm studies or individual study arms is not relevant for the benefit assessment. Thus, the left-hand strand of the overview of study designs in Figure 1, showing designs without a comparison, is not discussed further. Only comparative study designs are relevant to the research question of the benefit assessment.

Depending on the comparative study design chosen for the generation of routine practice data for a benefit assessment, different requirements for the conduct and analysis of the study arise. Table 1 shows the steps from the definition of the research question of the benefit assessment

to the result of the investigation of this question and summarizes the existing requirements in this process.

While general scientific principles, such as the formulation of the research question to be answered or the interpretation of the results (taking into account the achieved certainty of the results), are performed independently of the study design chosen, other steps of the benefit assessment differ depending on the study design. This is because, for certain study designs a fair, causally interpretable comparison can be assumed while for other designs, this fair comparison needs to be approximated by specific steps in study planning, data collection and analysis.

Table 1: Overview of general and specific requirements for the individual steps in the generation of routine practice data for benefit assessments, depending on study type

Process step	**General requirements (for all study types)**	**Specific requirements for comparative studies without randomization**	**Specific requirements for comparative studies with randomization**	**Specific requirements for adjusted indirect comparisons via a common comparator (intermediate comparator)**
Formulation of the research question and decision on a study design	▪ Identification of the evidence gap ▪ Formulation of the research question (PICO) according to the evidence gap ▪ Consideration of the requirements for the benefit assessment from §35a SGB V	▪ No factors that make it unlikely that sufficiently valid results can be achieved with this study design	▪ No very large (dramatic) effects to be expected for decision-guiding outcomes; outcomes of interest also achievable under comparator therapy	▪ Availability in principle of studies for such a comparison (preliminary search)
Study planning	▪ Detailed study protocol finalized before the start of data collection ▪ Prespecified analysis plan	▪ Emulation of the planning of comparative studies with randomization (the target trial) ▪ Prespecification of possible confounders and their adjustment in the analysis	▪ Adaptation of the study design to the daily treatment routine (pragmatic randomized study: inclusion and exclusion criteria, interventions, outcomes, visits to the doctor)	▪ Consideration of pragmatic randomized studies in the inclusion criteria ▪ Planning of a systematic review, including definition of the requirements for indirect comparisons
Data collection	▪ The data collection tool chosen must be able to provide data of the required quality	▪ Ensuring the availability of data for confounder control	▪ Use of existing data structures, e.g. registries	▪ If necessary, re-analysis of existing studies to meet requirements for indirect comparisons
Analysis und interpretation	▪ Consideration of the informative value of the different study designs and the specific data quality when interpreting the results	▪ Approximation to the similarity of the groups in terms of prognostic factors through adjustment	▪ Analysis and interpretation following existing standards	▪ Examination of the conditions for indirect comparisons (similarity, homogeneity, consistency of studies)
PICO: patient, intervention, comparison, outcome; SGB V: Sozialgesetzbuch V (Social Code Book V)				

Choice of design and study planning

The decision on a study design should take into account whether sufficiently valid results for a benefit assessment can be achieved with the design chosen.

An adequate study design that is also comprehensible in its timing is of decisive importance for the validity of the results of a study. A study protocol and a statistical analysis plan should be prepared before the study starts, and the study should be registered in a study registry.

For the planning of comparative studies without randomization, in order to compare treatment effects, the explicit replication of the planning of comparative studies with randomization is recommended (emulation of target trials).

Adjustments to compensate for the influence of structural inequality of treatment groups are essential for the data analysis of studies without randomization. In order to avoid a results-driven analysis, the relevant confounders and the procedure of the adjustment in the analysis must be prespecified comprehensibly and in the necessary depth of detail in the study protocol. The relevant confounders must be systematically identified (e.g. on the basis of scientific literature with the involvement of subject experts) and prespecified in the study protocol. The availability of corresponding data in the selected data source must be ensured before deciding on a comparative study without randomization. Adjustment only for the confounders available in the data set is insufficient if the relevant confounders are not covered.

When conducting a comparative study without randomization, it is possible to collect the data retrospectively or prospectively or in combination (partly retrospectively and partly prospectively). Retrospective data collection only makes sense if the data set on the basis of which the retrospective data collection is to be conducted contains the necessary data in the quality required. The availability of the relevant data must be ensured before deciding on a retrospective design. Historical controls are possible if the patient populations studied in the past are sufficiently similar to the patient population currently being treated and if data of sufficient quality for a meaningful comparison have been collected in the past. In addition, specific data relevant to the current study (e.g. individual patient data on confounders) must be available from the historical data set, and the data must be sufficiently similar (e.g. outcomes and confounders defined and corresponding data collected in a sufficiently similar manner).

If the necessary data are not available in sufficient quantity or quality, prospective data collection is required. If possible, existing data sources can be used (e.g. indication-specific clinical registry) in which any missing data (e.g. individual outcomes) can be added to the data set in the prospective data collection.

Data collection tools

The various data collection tools (study-specific data collection, registries, electronic patient records and claims data of health insurance funds) can in principle be used for comparative studies without randomization as well as for studies with randomization.

In practice, the collection of routine practice data from electronic patient records and claims data of health insurance funds for use in a benefit assessment does not appear realistically feasible at present and in the near future. The main reasons are the limited representation of relevant data for the benefit assessment (patient characteristics and outcomes) and the limited data quality of these sources.

In addition to study-specific data collection, patient registries covering a given disease (disease registries) are particularly suitable for data collection for benefit assessments. This is because, of the data collection tools that are not primarily geared towards comparative studies, such registries are most likely to offer the option of adapting the data collection to the requirements of these studies. This concerns both the specification of the necessary data and the data quality. In recent years, the aims and scope of the documentation implemented in registries have been expanded. In particular, the increasing documentation of clinical information in registries that can be used to describe the PICO (population, intervention, comparison, and outcome) for benefit assessments is relevant in this context. If a registry is expandable in principle, the combination with a supplementary, study-specific data collection for the respective registry study is also conceivable.

Studies based on data collection in a registry

Conceptually, it is important to distinguish between registries (active, prospective, standardized documentation of observation units on predefined questions, but expandable over time) and studies in these registries (registry studies). In principle, non-interventional and interventional comparative studies are possible in registries. In registries, comparative studies without as well as with randomization can be conducted.

Data quality requirements

Several national and international guidelines, overviews and position papers are available to describe data quality requirements in registries. These are broadly consistent in their main features. Ultimately, however, it is neither decisive nor necessary that all the measures mentioned there have been fully implemented, but rather that the data relevant to the specific research question are available in such a quality that an analysis within the framework of a registry study can be reliably interpreted. To ensure this, various categories of quality criteria for the data of a registry can be distinguished (see Table 2).

Table 2: Criteria for data quality and for ensuring the quality of routine practice data collection for the benefit assessment of drugs

Category	Quality criteria
Mandatory criteria to ensure data quality	▪ Detailed registry description (aim, registry protocol) ▪ Exact definition / operationalization of exposures, clinical events, outcomes and confounders ▪ Current data plan / coding manual ▪ Training on data collection and recording ▪ Clearly defined inclusion and exclusion criteria for registry patients ▪ SOP system for data collection ▪ Package of measures to ensure the accuracy of data and to provide information on error rates (e.g. source data verification, internal and external audits, IT-supported checks [e.g. cross-reference checks]) ▪ Documentation trail – documentation of process and definition changes in the registry ▪ Scientific independence of the registry ▪ Sustainable financing
General criteria that are regularly relevant for registry studies for benefit assessments	▪ Use of exact dates for patients, disease and events ▪ Detailed information on the drug therapy (active substance, dose, dose change, including dates) ▪ Timeliness (currentness and rapid availability of the required results)
General criteria that may be relevant for registry studies for benefit assessments, depending on the research question	▪ Use of standard classifications (e.g. ICD-10) and terminology (e.g. MedDRA) ▪ Use of valid standard survey tools (questionnaires, scales, tests) ▪ Flexibility and adaptability (e.g. for embedding studies, for further data collection, in the event of changes in the health care situation) ▪ Linkability with other data sources
Criteria whose degree of fulfilment is to be assessed with regard to components of the research questions[a]	▪ Representativeness of the sample / selection of the sample ▪ Completeness of data per data collection time point (lost-to-follow-up, drop-outs) ▪ Completeness of data collection time points ▪ Correctness of data ▪ Collection of data on all confounders relevant for the research question ▪ Data consistency over time
a: The criteria mentioned are important criteria of data quality, but can only be assessed in relation to specific questions. On the one hand, for example, “accuracy of data” and “consistency of data over time” only refer to data that are relevant to the respective question. On the other hand, “representativeness of the sample” refers only to the population relevant to the research question, but not to the entire registry population. ICD: International Statistical Classification of Diseases and Related Health Problems; IT: information technology; MedDRA: Medical Dictionary for Regulatory Activities; SOP: standard operating procedure	

In the context of the suitability testing of a specific registry, this list of criteria should be used to assess for each specific research question

- whether and to what extent the individual criteria are fulfilled
- what influence a possible non-fulfilment is likely to have on the quality of the results, and

- whether possible deficits can be corrected in a registry-based study using a reasonable amount of resources.

Interviews with registry operators

From the interviews with the registry operators, it emerged that the suitability of the respective registry for the benefit assessment of drugs cannot be answered in a generalized manner, but depends on the specific research question. However, from a technical and organisational point of view, the registries are usually prepared to implement any necessary extensions of the data set.

It was also possible to deduce various factors from the interviews that are beneficial or obstructive to the operation of the registry. From this and generally from the results of the interviews, recommendations for action can be derived for registry operators, those responsible for registry studies, as well as health care and health policy decision-makers.

Requirements for the analysis

In studies without randomization, the groups to be compared cannot be considered similar in terms of prognostic factors. This similarity that is required for a fair comparison is generally not given in these studies. Group differences concerning possible confounders, i.e. factors that are related to both treatment and outcomes and can consequently distort a treatment effect, must therefore be considered when estimating effects. A detailed study protocol and analysis plan should thus describe, among other things, the systematic identification of relevant confounders (e.g. by means of the scientific literature with the involvement of experts), since confounder adjustment must be based on which confounders are relevant to the research question and not on which ones are included in the data set.

Various approaches are available for confounder adjustment: for a benefit assessment of drugs, as a rule only those approaches using individual patient data are meaningful. The use of propensity scores is a frequently applied method for the consideration of confounders in comparative studies without randomization based on registries. When using the propensity score method, important criteria include positivity, overlap and balance. The relevant decision structure must be defined in the analysis plan; this structure should also contain specifications for decisions depending on the specific data situation (e.g. minimum level of overlap and balance).

In practice, even if the usual methodological guidelines are strictly followed, the accuracy of the assumptions regarding confounder adjustment cannot be fully verified and unmeasured or completely unknown confounders may play a role. Therefore, results from comparative studies without randomization as a rule at best provide only a low degree of qualitative certainty of results. Even if studies without randomization only show a low qualitative certainty of results, they can increase the certainty of results of the overall conclusion on added benefit if combined with other data (e.g. if reliable data on important outcomes are supplemented by the study

without randomization in other outcome categories, or if a [small] study with randomization is combined with a [larger] study without randomization).

Even with the most careful analysis and fulfilment of the quality requirements mentioned above, due to potentially unknown confounders, a conclusion on the benefit or harm of an intervention should only be derived from the effects observed in the study if these effects exceed a certain effect size. A (positive or negative) conclusion on the benefit or harm can be drawn if the confidence interval for the effect observed exceeds a threshold that must be defined. Since the fulfilment of the above-mentioned quality requirements is a prerequisite for the observation of effects, this threshold value should be significantly below the value for the "dramatic effect" (relative risk of 5–10), e.g. in a range of 2–5 for the relative risk. The specific threshold depends on the quality of the data in the individual case.

In benefit assessments of drugs according to §35a of SGB V, starting from this threshold for a conclusion on benefit or harm, if the threshold is exceeded, there is at least a minor added benefit for the respective outcome. Exceptions are outcomes in the category "non-serious/non-serious complications", because, according to the Regulation for Early Benefit Assessment of New Pharmaceuticals[3], a "not only marginal improvement" is additionally required for these outcomes. For all outcome categories, classification into the extent categories "considerable" or "major" requires higher (i.e. above the above-mentioned threshold) effect sizes that are graded according to magnitude.

Especially for rare diseases, it may be useful and necessary to conduct studies in international collaboration. On the one hand, such analyses require standardized data harmonization. On the other, analyses that use data generated outside of the German healthcare context of interest must justify that these data can be classified as routine practice data in terms of health care in Germany or that deviations are not relevant for the effect estimate. In the case of analyses from several registries, it can for efficiency reasons be useful not to form a common data pool and then analyse it, but to plan and conduct identically designed studies in the individual registries and then to summarize these studies meta-analytically.

Reporting requirements

Irrespective of the study type, the complete documentation of a study includes the study protocol (planning of the methods and conduct of the study), the analysis plan (planning of the data analysis) and the results report (description of the planned methods [including the analysis] and conduct of the study, deviations from this planning and reporting of complete results). The study protocol and the analysis plan serve not only to describe the methods and conduct of the study in the case of prospectively collected data, but also to prespecify the study planning. This prespecification is an essential quality feature of a study with prospective data collection. For prospective comparative studies without randomization, this prespecification should cover

[3] *Arzneimittel-Nutzenbewertungsverordnung*

confounder control (documentation of confounders and definition of adjustment methods) as comprehensively as possible.

The possibility of results-driven analyses and reporting is an unsolved problem for retrospective study designs. Irrespective of this, a study protocol including an analysis plan should also be prepared for studies with retrospective data collection from existing data sets. It is recommended to make these documents publicly available.

Optimized studies for decision-making in health care

The conduct of a high-quality comparative study without randomization is resource-intensive. In this context, the current discussion about adjustments in the conduct of comparative studies with randomization is relevant. On the one hand, this should ensure that the results are meaningful for broader populations (pragmatic studies with randomization) and, on the other hand, reduce the necessary effort ("large simple trials" and registry-based studies with randomization). In summary, it may be easier and more purposeful to conduct a comparative study with randomization considering these adjustments than to try to generate high-quality results from a comparative study without randomization.

Suggestions for an approach to routine practice data collection according to GSAV and §35a (3b) SGB V

In the present rapid report, the results of the project were also analysed in connection with the possibility of routine practice data collection introduced by the "Gesetz für mehr Sicherheit in der Arzneimittelversorgung" (GSAV[4], Law for More Safety in the Supply of Medicines [own translation]).

Definition of the research question

The basis for routine practice data collection according to GSAV is the definition of the research question to be answered by this data collection. The question at least contains the components of the PICO format and the required duration of data collection. The exact specifics of the research question are derived from the evidence gap shown in the benefit assessment and that is to be closed by the data collection. A research question defined in this way is also the starting point for the description of the necessary scope of data collection (including duration of observation and sample size calculation).

Evidence gaps in benefit assessments of drugs (orphan drugs)

In order to be able to better assess the evidence gaps in benefit assessments and their importance in determining the extent of added benefit, G-BA decisions on benefit assessments of orphan drugs were examined in more detail. Decisions on orphan drugs with market access in the years

[4] *Gesetz für mehr Sicherheit in der Arzneimittelversorgung*

2014 to 2018 were examined, including decisions on new therapeutic indications for orphan drugs in this period.

On the one hand, the analysis shows that, for orphan drug assessments in the years 2014 to 2018, relevant data were submitted for a large part of the research questions in the corresponding dossiers for the benefit assessments at the time of market access (80 of 85 questions, 94%). In about two-thirds of the cases, these were studies with randomization, and in one third of the cases, without randomization. Nevertheless, an added benefit of the intervention was quantified for only about a quarter of all research questions. It could not be inferred from the analysis that the added benefit was particularly non-quantifiable in very small target populations.

The analysis of the research questions with the conclusion of a non-quantifiable added benefit showed that in 61% (52 of 85) of the research questions assessed by the G-BA for the years 2014 to 2018, evidence gaps were identified that were decisive for the lack of quantifiability of added benefit. In consequence, almost two-thirds of the orphan drug assessments are potential candidates for routine practice data collection according to GSAV. In all 52 cases, data on the control group were also missing, which is why targeted routine practice data collection according to GSAV must as a rule be planned and conducted in a comparative manner involving a control group (comparator therapy).

Evidence gaps are often present in several outcome categories (mortality, morbidity, health-related quality of life, and adverse events). With regard to morbidity and health-related quality of life, information on patient-reported outcomes (PROs) will often be required for targeted routine practice data collection according to GSAV, as these are essential for addressing the evidence gaps in these outcome categories.

It is therefore overall foreseeable that a data collection required by a regulatory authority, which is intended in particular to identify rare or late-onset side effects of the respective orphan drug, will in unchanged form often not represent a suitable data collection for a benefit assessment (i.e. a targeted routine practice data collection according to GSAV). Which change or extension to a regulatory data collection is necessary to achieve suitability for a benefit assessment has to be examined in each individual case based on the existing evidence gap for the quantification of the added benefit.

Possible process steps of routine practice data collection according to GSAV

Based on the analysis of the orphan drug assessments from 2014 to 2018 and on the requirements of SGB V, Table 3 shows possible process steps of the routine practice data collection according to GSAV in the benefit assessment procedure according to §35a SGB V.

Table 3: Process steps for routine practice data collection according to GSAV for benefit assessments according to §35a SGB V

Process step	Comment
Identification of an evidence gap in the G-BA decision on a benefit assessment according to §35a SGB V	▪ Evidence gap: relevant data gap for the comparison of the new drug with the (appropriate) comparator therapy with regard to patient-relevant outcomes (especially if the evidence gap does not allow quantification of the added benefit)
Description of the G-BA specifications for routine practice data collection according to GSAV and transmission to the pharmaceutical company	▪ Definition of the research question ▪ Duration, type and scope of data collection (duration of data collection per patient, sample size based on a sample size estimation) ▪ Type and scope of the analysis (depending on the study type used) ▪ Specification of the time points for the evaluation of the data obtained (at least every 18 months) ▪ Specification of the requirements, taking into account ongoing and planned data collection, especially those resulting from requirements of the regulatory authorities (e.g. EMA)
Evaluation of the data collected and the obligation to collect data	▪ At the time of the first evaluation, the G-BA will check whether a (publicly available) study protocol including an analysis plan is available that reflects the routine practice data collection according to GSAV as requested ▪ At the first and each subsequent evaluation time point, the G-BA will evaluate the available data and decide whether the data collection can be stopped or should be continued
EMA: European Medicines Agency; G-BA: Gemeinsamer Bundesausschuss (Federal Joint Committee); GSAV: Gesetz für mehr Sicherheit in der Arzneimittelversorgung (Law for More Safety in the Supply of Medicines [own translation])	

Conclusion

Study design and data collection

- The use of routine practice data for benefit assessments of drugs according to §35a SGB V requires a comparison between the new drug and the appropriate comparator therapy specified by the G-BA; this requires the conduct of comparative studies.
- The collection of routine practice data from electronic patient records and from claims data of health insurance funds for benefit assessments according to §35a SGB V is currently and foreseeably not considered realistic; rather, a study-specific data collection or data collection from patient registries is necessary

Routine practice comparative studies without randomization

- If comparative studies without randomization are to be used for the benefit assessment, it must be ensured at the stage of study planning that the study conduct and the data collected are of the quality required to generate interpretable results.
- Essential components of such a study planning are a study protocol including an analysis plan, the emulation of a target trial that deals with the relevant research question, and ensuring that sufficient data are collected for confounder control.

- A key aspect of the analysis of a comparative study without randomization is adequate confounder adjustment; this adjustment must be pre-specified as far as possible and the assumptions made (e.g. the definition of the relevant confounders) must be substantiated.
- No effects can be derived from comparative studies without randomization if the data quality in the data sources used and the quality of analysis and reporting is not high.
- Even under high quality requirements (for data, analysis and reporting), no more than a hint of an effect can normally be derived from comparative studies without randomization.
- Due to the inherent uncertainty of the results from comparative studies without randomization, because of potentially unknown confounders, a conclusion on the benefit or harm of an intervention should only be derived from the effects observed in the study if these effects exceed a certain effect size. Quantification of an added benefit according to the legally prescribed extent categories requires corresponding effect sizes graded according to magnitude.
- The possibility to consider retrospective study designs depends on whether the available data sources already contain the necessary data in the required quality; comparisons of a new drug with historical controls only appear realistic if the same data source (e.g. a disease-specific clinical registry) is used for the new drug and the historical control.

Routine practice comparative studies with randomization

- Routine practice comparative studies can also be randomized (pragmatic clinical trials).
- The effort required for a routine practice comparative study with randomization will generally – with comparable data quality – be less than the effort required for a study without randomization, as confounder data collection and confounder adjustment can be omitted.
- Routine practice comparative studies with randomization are of higher informative value than those without randomization, and the quantification of added benefit is more reliable.
- Especially after market authorization, depending on the existing research question, routine practice comparative studies with randomization can be conducted with limited data collection ([large] simple trials); conducting them in registries has an additional potential to accelerate the conduct of the studies and make them less complex and resource-intensive (registry-based comparative studies with randomization).

Routine practice data collection according to GSAV

- Whether the various patient registries are currently already suitable for data collection according to §35a SGB V cannot be answered in general, as this depends on the respective registry and the specific research questions posed.
- On the basis of the analyses and the discussions with the registry operators, fields of action can be described that serve to support the individual registries in particular and the

registry landscape in Germany in general in routine practice data collection according to GSAV; these fields of action are described in the report.

- The findings of the present report can be used for routine practice data collections according to GSAV to close evidence gaps after a benefit assessment; a proposal for the corresponding approach is provided with the report.

Table of contents

List of tables

Page

List of figures

Page

List of abbreviations

Abbreviation	Meaning
AMNOG	Arzneimittelmarktneuordnungsgesetz (Act on the Reform of the Market for Medicinal Products)
ATC	Anatomical Therapeutic Chemical
BSC	Best supportive care
CI	Confidence interval
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information (German Institute for Medical Documentation and Information)
DNVF	Deutsches Netzwerk Versorgungsforschung (German Network for Health Services Research)
DRG	Diagnosis-related groups
EHR	Electronic health record
EMA	European Medicines Agency
EUnetHTA	European Network for Health Technology Assessment
FDA	Food and Drug Administration
G-BA	Gemeinsamer Bundesausschuss (Federal Joint Committee)
GP	General practitioner
GSAV	Gesetz für mehr Sicherheit in der Arzneimittelversorgung (Law for More Safety in the Supply of Medicines)
hdPS	High-dimensional Propensity Score
HES	Hospital episode statistics
HIS	Hospital information systems
ICD	International Statistical Classification of Diseases and Related Health Problems
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care)
IT	Information technology
MDK	Medizinischer Dienst der Krankenkassen (Medical Service of the Statutory Health Insurance Funds)
MedDRA	Medical Dictionary for Regulatory Activities
OPS	Operationen- und Prozedurenschlüssel (Operation and Procedure Classification System)
PICO	Population, intervention, comparator, outcome
PRO	Patient-reported outcome
RCT	Randomized controlled trial
SGB	Sozialgesetzbuch (Social Code Book)

Abbreviation	Meaning
SHI	Statutory health insurance
SVR	Sustained virological response

1 Background

On 2 May 2019, the Federal Joint Committee (G-BA) commissioned the Institute for Quality and Efficiency in Health Care (IQWiG), according to §139a (3) Social Code Book V (SGB V), to develop scientific concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a SGB V.

Sections 5.1 to 5.6 show the results of this work. Based on these results, Section 5.7 presents proposals for the collection of routine practice data according to the Law for More Safety in the Supply of Medicines (GSAV[5] [1], own translation), as this represents one of the potential uses of the collection of routine practice data.

[5] *Gesetz für mehr Sicherheit in der Arzneimittelversorgung*

2 Research question

The aims of this work are

- The creation of an overview of possible concepts for generating and analysing routine practice data. In particular, data collections that are not classified as randomized controlled trials (RCTs) should also be considered (see Sections 5.1 to 5.4).
- The assessment of the identified concepts of data generation and their analysis with regard to their suitability to answer the research question of a benefit assessment according to §35a SGB V, especially with regard to the possibility of quantifying the added benefit of a new drug (see Section 5.6).
- The specification of criteria for data quality and the methodological requirements for the data collected within the framework of the respective generation of data. In this regard, the measures required to ensure data quality should also be addressed (see Sections 5.5.1 and 5.5.2)
- The definition of requirements for reporting, as well as for the preparation and structure and the statistical analysis of the data collected within the framework of the respective generation of data (see Sections 5.5.2 and 5.5.3).

3 Course of the project

On 2 May 2019, the G-BA commissioned IQWiG to develop scientific concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a SGB V.

The work on the research question was based on a project outline and documented in a rapid report. This report was submitted to the G-BA and published on the IQWiG website 2 weeks later.

4 Methods

The present commission comprises the conceptual development of proposals on the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a SGB V. The conceptual work was supported by information retrieval and assessment.

4.1 Information retrieval and assessment

According to the project outline, the development of the concept for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a SGB V was supported by 3 modules:

Empirical information from benefit assessments of drugs according to the Act on the Reform of the Market for Medicinal Products (AMNOG[6], §35a SGB V)

The evidence base and the results of the completed benefit assessments according to § 35a SGB V were to be presented systematically. In this context, constellations that could contain information for the present research question (especially assessments based on non-randomized studies) were to be described and analysed. Case reports and aggregated data for case constellations represented potential formats (if applicable, including effect sizes with confidence intervals [CIs] in the constellation selected).

The analysis was performed for assessments of orphan drugs, as, due to the GSAV, these will probably be the focus of future data collections commissioned by the G-BA.

Exploratory literature search

An exploratory literature search was conducted for research questions arising as part of the conceptual work (e.g. informative value of non-randomized studies, depending on the available data constellations) in order to support the concept to be developed by the current state of scientific knowledge.

Interviews with registry experts

The commission comprised the specification of criteria for the quality and methodological requirements with regard to the data collected within the respective data generation frameworks. In this context, measures to ensure data quality should also be described.

Registries will play an important role in data generation. Questionnaire-based interviews with experts on the structure and use of registries were conducted in order to consider practical experience in the requirements for the registries (see Section 4.2 for the conduct of the interviews).

[6] *Arzneimittelmarktneuordnungsgesetz*

4.2 Determination of quality criteria for patient registries

Interviews with registry experts

Between the end of June and August 2019, 6 expert interviews were conducted with representatives of selected patient registries in Germany. The aims were, among other things, to get to know the perspective of the practical work in the registries with regard to key quality requirements for patient registries and the studies based on these registries, as well as to gain insight into important beneficial and hindering factors in the establishment and maintenance of registries. The selection of registries was restricted to those in which drug therapy plays an important role in the treatment of patients and is also reflected in the documentation in the registry. Table 4 lists the registries included. The selection comprised regional, national and international registries including patient populations of varying sizes and documenting both rare and common as well as both oncological and non-oncological diseases. In addition, an expert workshop was conducted involving 3 statisticians ("statistical workshop") with well-founded expertise in analyses of patient registry data (see also Table 4).

Table 4: Interviews with registry operators and statistical experts

Registry interview	
Registry name	**Interview partners (affiliation)**
Cancer Registry Bavaria of the Bavarian Health and Food Safety Authority and the Tumour Centre Regensburg	▪ Monika Klinkhammer-Schalke (DNVF; Tumour Centre Regensburg) ▪ Brunhilde Steinger (Tumour Centre Regensburg) ▪ Vinzenz Völkel (Tumour Centre Regensburg)
CRISP – Clinical Research platform Into molecular testing, treatment and outcome of (non-)small cell lung carcinoma Patients of AIO and iOMEDICO[a]	▪ Frank Griesinger (Pius-Hospital Oldenburg; AIO e. V.) ▪ Martina Jänicke (iOMEDICO) ▪ Martin Sebastian (University Hospital Frankfurt; AIO e. V.)
German Cystic Fibrosis Registry of the German Cystic Fibrosis Association	▪ Manuel Burkhart (German Cystic Fibrosis Registry) ▪ Lutz Nährlich (Justus-Liebig University Gießen, German Cystic Fibrosis Registry) ▪ Miriam Schlangen (German Cystic Fibrosis Association)
German MS Registry of the German MS Association	▪ David Ellenberger (MS Research and Project Development, MS-Registry of the DMSG, Federal Association) ▪ Alexander Stahmann (MS Research and Project Development, MS-Registry of the DMSG, Federal Association)
Patient Registry of the European Society for Blood and Marrow Transplantation (**EBMT**)	▪ Nicolaus Kröger (EBMT; University Hospital Eppendorf, Hamburg)
RABBIT – Rheumatoid Arthritis: Observation of Biologics Therapy; Registry of the German Rheumatism Research Centre (DRFZ)	▪ Anne Regierer (DRFZ) ▪ Anja Strangenfeld (DRFZ) ▪ Angela Zink (DRFZ)
Epidemiological-statistical interview partners	
Name	**Affiliation**
Oliver Kuß	Institute for Biometry and Epidemiology, German Diabetes Centre of the Heinrich-Heine University, Düsseldorf; Scientific Advisory Board of IQWiG
Rolf Lefering	Institute for Research in Operative Medicine (IFOM), University of Witten/Herdecke
Claudia Spix	German Paediatric Cancer Registry, University Hospital Mainz

a: Definition of abbreviation corrected (the term "AIO" was missing in the German original version 1.0).

AIO: *Arbeitsgemeinschaft Internistische Onkologie in der Deutschen Krebsgesellschaft e. V.* (Working Group Medical Oncology in the German Cancer Society); DMSG: *Deutsche Multiple Sklerose Gesellschaft Bundesverband e. V.* (German Multiple Sclerosis Society – National Association); DNVF: *Deutsches Netzwerk für Versorgungsforschung* (German Network for Health Services Research); DRFZ: *Deutsches Rheuma-Forschungszentrum* (German Rheumatism Research Centre); EBMT: European Society for Blood and Marrow Transplantation; IFOM: *Institut für Forschung in der operativen Medizin* (Institute for Research in Operative Medicine); iOMEDICO: International Organisation of Medical Oncology; MS: multiple sclerosis; RABBIT: *Rheumatoide Arthritis: Beobachtung der Biologika-Therapie* (Rheumatoid Arthritis – Observation of Biologics Therapy)

Interviews with registry operators

The interview partners from the selected patient registries received a 3-part questionnaire (see Appendix A) in order to prepare for the interview. The first 2 parts of the questionnaire were to be filled in and sent to IQWiG before the interview. The answers were to help IQWiG to prepare the interview and some answers were discussed in more detail during the interview. In the first part, the questionnaire records important features of the patient registry of interest. In the second part, the respondent was to evaluate a list of 29 quality criteria for patient registries according to the aspects of relevance, simple or complex feasibility as well as the effort required. Finally, for all criteria, the degree of fulfilment in the respective registry was to be estimated.

The third part, only included as advance information and interview preparation, contains 2 typical assessment scenarios for drugs in the work of the G-BA, which were to be discussed in the interview with regard to the topic "potential collection of additional routine practice data".

The first and third parts of the questionnaire were designed on the basis of expertise and according to plausibility aspects. The necessity to quickly set interview dates and conduct interviews within the tight timelines of the rapid report only allowed for an exploratory, not a systematic, literature search for the development of the second part (quality criteria) of the expert questionnaire. The memorandum of the German Network for Health Services Research (DNVF[7]) [2] was initially chosen to organize the quality criteria list. The structure of the publication seemed well suited for this, as, in contrast to other documents on quality standards for patient registries, it particularly considers the aspects of data quality and data quality assurance, which are the focus of the G-BA's commission to IQWiG.

Four sources were used for preselection in order on the one hand to identify commonly named quality criteria and on the other, to obtain information on relevant criteria that are particularly relevant for drug assessments (e.g. dates on the start and end of treatment). Besides the above-mentioned DNVF publication [2,3], these were

- the manual by the Agency for Healthcare Research and Quality (AHRQ) as an established standard source [4] (Chapter 25)
- the requirements for patient registries proposed by the European Medicine Agency (EMA), which are currently undergoing public discussion [5] (Chapters 5 to 6)
- the aspects formulated by the European Network for Health Technology Assessment (EUnetHTA) in the Registry Evaluation and Quality Standards Tool (REQueST) tool [6] (Criteria 1 to 19)

This selection of publications reflects national and European expertise and is also both up-to-date and internationally representative of the discussions on the quality of patient registries held at regulatory and health technology assessment (HTA) agencies. As a result, for the interviews,

[7] *Deutsches Netzwerk Versorgungsforschung e. V.*

an initially limited list was prepared for the second part of the questionnaire, which was then expanded and quality-assured as described in the next section.

Statistical workshop

Before the statistical workshop was conducted, issues for discussion were compiled and sent to the participants (see Appendix B). An open discussion of these issues was then held. These issues are considered in this rapid report, particularly in Section 5.5 ("Requirements for data and analyses").

Compilation of quality criteria for patient registries

In order to expand the list of quality criteria used in the questionnaire with registry operators, firstly, further relevant publications with corresponding recommendations on the topic were identified and secondly, information provided by the experts on quality criteria relevant for the practical work with registries, but not yet considered, was integrated. In the section "superordinate criteria", these include aspects such as a secure funding basis and the timeliness of the provision of the required results of the analyses and the corresponding reports.

As Mandeville 2018 [7] determined after a systematic search for publications on registry quality conducted within a current EUnetHTA project, the documents with recommendations showed some variations (e.g. purpose, content, structure, level of detail, method of development, format). However, if compared, they can be used to identify common or consistently mentioned quality criteria.

For the list of quality criteria for patient registries, the above-mentioned sources [2,4-6] were supplemented by a fifth publication [8] for the current report. This publication contained the criteria formulated from a regulatory perspective with regard to the evaluation procedure of the US Clinical Trials Transformation Initiative (CTTI) for the suitability of patient registries in the conduct of embedded clinical studies. The quality criteria for registries and registry studies were extracted from these 5 sources and compared with each other. Table 14 in Appendix C shows the results of the comparison. The comparison of sources with regard to the naming of the individual quality criteria was conducted by 2 researchers independently of each other. Inconsistent evaluations were discussed and solved by consensus.

5 Results

5.1 Study designs and data collection tools for generating routine practice data

This chapter outlines study designs and data collection tools for generating routine practice data. Such data can be used to answer different questions, such as determining the prevalence of different diseases in Germany, describing the health care provided to certain patient groups or assessing the benefit of drugs. The respective research question to be answered must be considered in the evaluation of the suitability of routine practice data for supporting decisions in the health care system.

What are routine practice data for benefit assessments of drugs?

For the present report, routine practice data are defined as follows:

- routine practice data are collected in patient populations eligible to receive the drug of interest in routine practice (in the approved therapeutic indication)
- in the collection of routine practice data, patients are treated without specific requirements

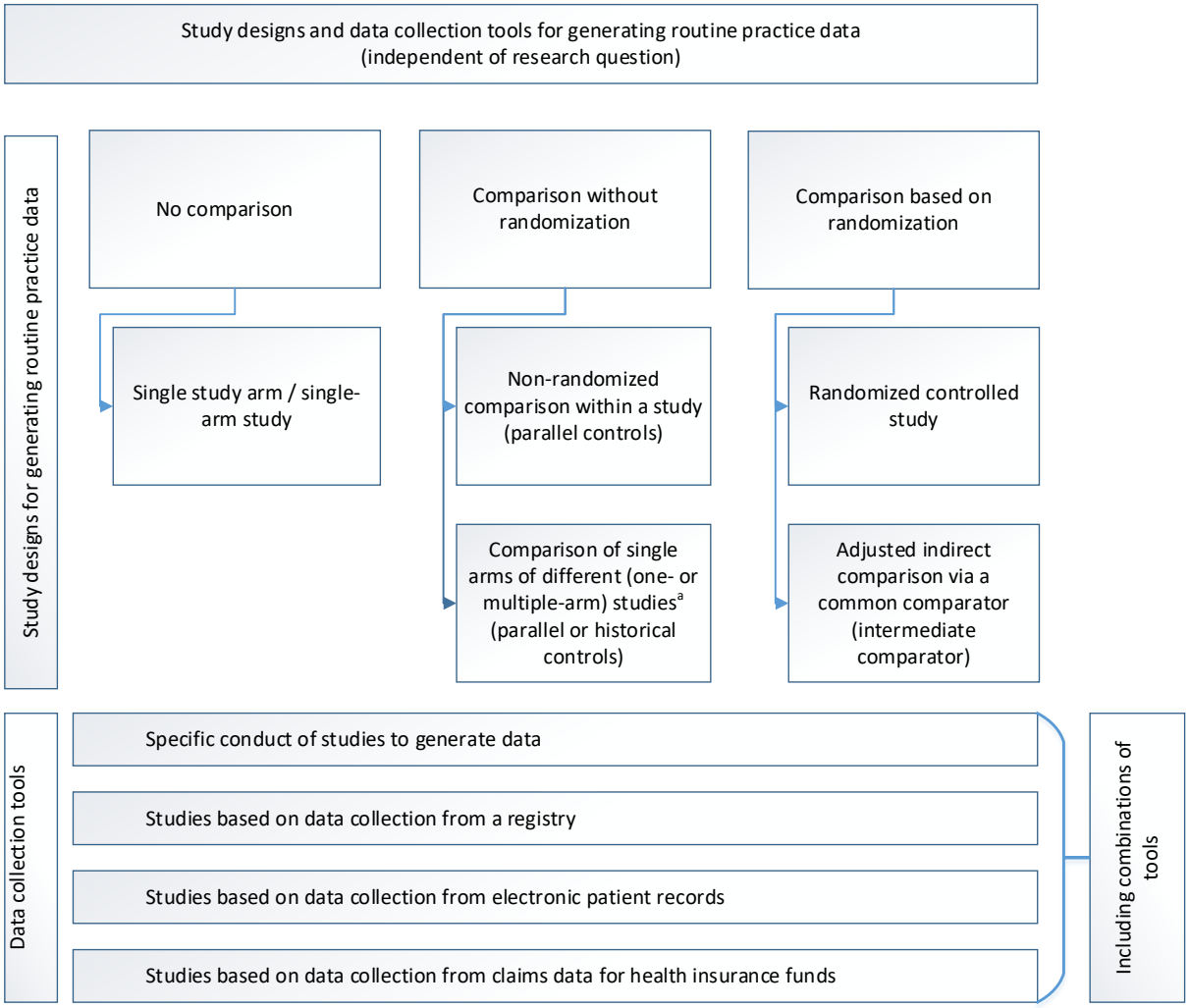
Since drug assessment according to SGB V is concerned with patient care in Germany, routine practice data must meet the 2 criteria mentioned above so that conclusions can be drawn for health care in Germany.

The definition of routine practice data implies neither a specific study design nor a specific data collection tool (see below) [9] .

The goal of collecting routine practice data does not require that data collection be limited to data collected in routine practice per se. Rather, such a misconceived restriction of data collection would jeopardize the goal of the benefit assessment. The benefit assessment regularly requires data that are not collected in routine practice for all patients (e.g. data on health-related quality of life, symptoms or side effects). For use in a benefit assessment, routine practice data must also be sufficiently valid and structured.

Overview of study designs and data collection tools

The following figure provides an overview of study designs and data collection tools that can be used to generate routine practice data. The upper part of the figure describes the study designs that are basically conceivable, depending on the possible type of comparison of interventions. The lower part names the tools that can be used to collect routine practice data in studies with different designs. It becomes clear that the various data collection tools can generally be used for all study designs.



a: including studies on the spontaneous course of the disease

Figure 2: Study designs and data collection tools for generating routine practice data

To assess the components of this overview with regard to their suitability for answering the research question of the benefit assessment according to §35a SGB V, the following section first describes the prerequisites for a benefit assessment. Then, starting from the research question of the benefit assessment, the requirements for the generation and analysis of routine practice data are discussed within the various strands of the overview.

5.2 Prerequisites for a benefit assessment

The starting point for considerations on a meaningful procedure for benefit assessments can be succinctly described with the following quotation:

“The effect of any treatment for a given patient is the difference between what happened to the patient as a result of giving him the treatment and what would have happened had treatment been denied.” [10].

This core statement makes the principle and the resulting dilemma clear; the basis of any conclusion on the effects of interventions is a comparison. It does not matter whether the effects are positive or negative and whether these effects are called “effect”, “effectiveness”, “benefit”, “added benefit” or “harm”. Only on the basis of a comparison is it possible to distinguish between “**after** intervention A” and “**due** to intervention A”; this distinction is necessary for a causal conclusion. The identical initial situation as described in the above quotation would provide the necessary “fair” comparison.

The dilemma arises from the fact that this comparison is in principle not possible in the individual patient, since (see the subjunctive in the quote) 2 alternative situations cannot be observed simultaneously.

One way out of the dilemma would be to observe the same patient consecutively, once with and once without the treatment of interest. Apart from the fact that this approach would only be feasible with reversible outcomes, it has 2 further disadvantages: On the one hand, the auxiliary condition implicit in the opening sentence, namely the identical initial situation, cannot be guaranteed (the comparison would not be “fair”), and on the other, the knowledge gained in this way is hardly general, i.e. also useful for other patients.

N-of-1 trials or cross-over studies offer further options. However, the critical prerequisites and limited options of interpretation mean that the areas of application are very limited.

The existing dilemma can be solved in such a way that the necessary comparison is not made within a patient but between patients. In order to achieve a certain degree of generalizability – the results are to apply to future patients – larger groups of patients are considered. One part of the patient population receives the treatment of interest, the other part does not, so that a comparison of the results is possible. In order to come as close as possible to the ideal of an identical initial situation and to enable a fair comparison, it must be ensured in the best possible way that, between the groups compared, neither the initial situations nor the patients differ in their characteristics (similarity of the treatment groups in terms of prognostic factors).

Just as in other scientific fields, but also in everyday life (e.g. in sports competitions), the lower the expected differences or effects, the higher the demands on ensuring fair comparisons. In modern medicine with its rather modest advances, a fair comparison is therefore the key prerequisite for conclusions on the benefit of interventions.

It follows from these deliberations that the sole consideration of single-arm studies or individual study arms is not relevant for the benefit assessment. Thus, the left-hand strand of the overview of study designs in Figure 2, showing designs without a comparison, is not discussed further. Only comparative study designs are relevant to the research question of the benefit assessment.

5.3 Routine practice data in benefit assessments

Depending on the study design chosen for the generation of routine practice data for a benefit assessment, different requirements for the conduct and analysis of the study arise. Figure 3 shows the steps from the definition of the research question of the benefit assessment to the result of the investigation of this question.

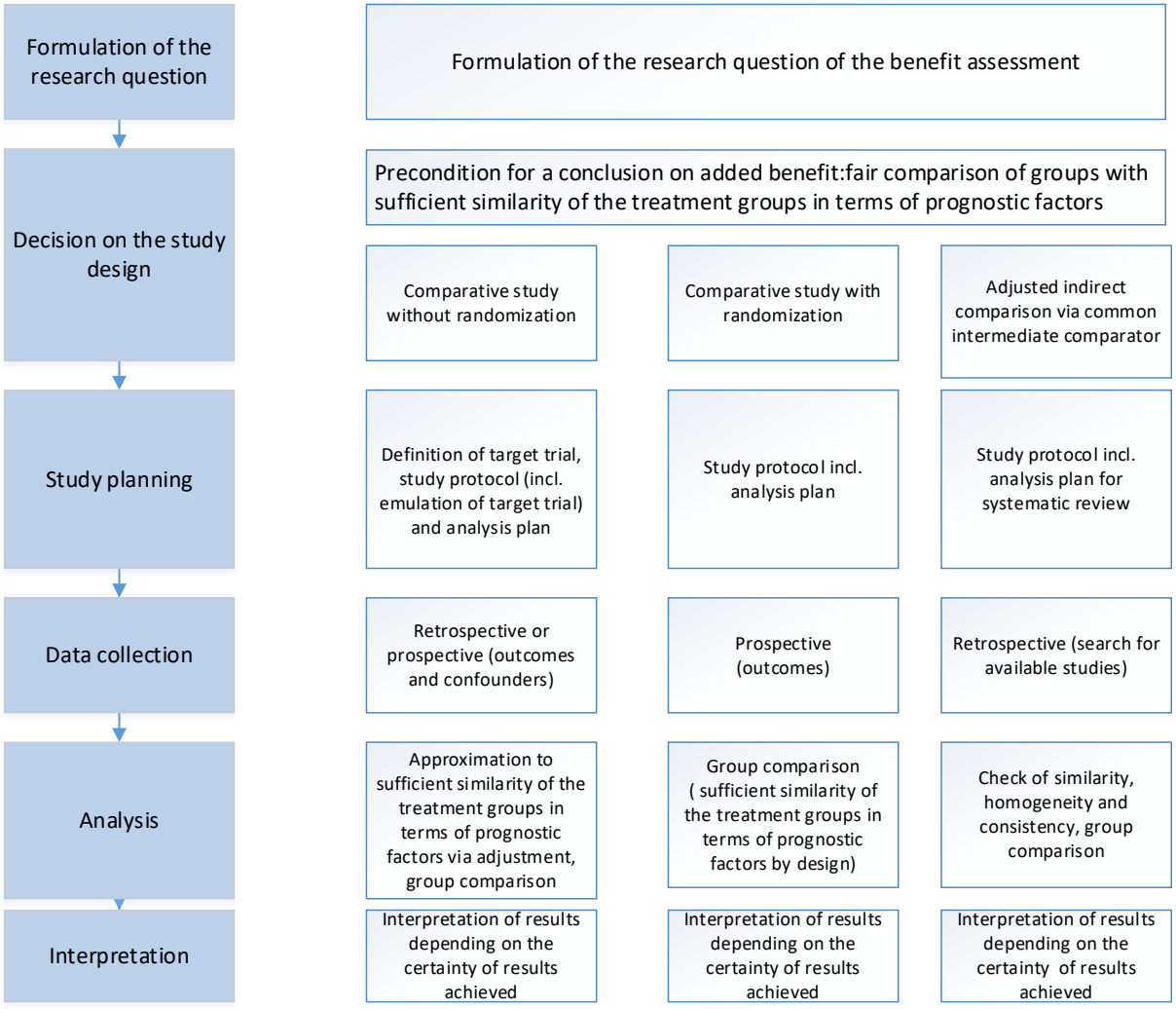


Figure 3: Process from the definition of the research question of the benefit assessment to the result of the investigation

While general scientific principles, such as the formulation of the research question to be answered or the interpretation of the results (taking into account the achieved certainty of the results), are performed independently of the study design chosen, other steps of the benefit assessment differ depending on the study design.

The following chapters describe general and study design-specific methodological and procedural requirements for study planning, data collection and analysis, and interpretation of the results. The following recommendations, among others, will be considered:

- Guidelines of the International Conference on Harmonization [11]
- Standards for the reporting of different study designs of the EQUATOR[8] Network [12]
- Framework for the Real-World Evidence Programme of the Food and Drug Administration (FDA) [13]
- PCORI[9] Methodology Standards [14]
- Cochrane Handbook for Systematic Reviews of Interventions [15]

5.3.1 Formulation of the research question and decision on a study design

General requirements

The prerequisite for the generation of routine practice data that can make a relevant contribution to the decision-making process of the G-BA is the formulation of the research question to be answered by the data collection. The starting point for the research question is the evidence gap that is to be closed. The research question must cover the elements of the PICO (population, intervention, comparison, and outcome) scheme. This means that the relevant patient population, the intervention to be tested, the comparator, and the outcomes must be defined.

The specific research question must be formulated pursuant to the requirements for a benefit assessment according to §35a SGB V. The following points must be considered [16,17]:

- the patient population corresponds to the approved therapeutic indication of the drug to be assessed
- the drug to be assessed is to be used in accordance with its approval
- the control arm of the study represents the (appropriate) comparator therapy
- the outcomes are patient-relevant outcomes

If the routine practice data are to be collected following an initial benefit assessment in order to close evidence gaps, the research question should explicitly address these gaps. In this context, the patient population, the control arm or the outcomes can be restricted to aspects not yet covered by the evidence, if necessary.

The decision on a study design should take into account whether sufficiently valid results for a benefit assessment can be achieved with the design chosen.

[8] Enhancing the QUAlity and Transparency Of health Research

[9] Patient-Centered Outcomes Research Institute

Special requirements for comparative studies without randomization

Under certain conditions, comparative studies without randomization are not feasible. For example, the literature describes that no valid effects can be expected in cases where the disease status or the expected treatment result strongly influences the choice of treatment [18]. A prospective comparative data collection may therefore not be possible at all, if the preference for a treatment option (e.g. a new drug) is so strong that a sufficient sample size for treatment with a suitable comparator therapy cannot be achieved within a reasonable period. Another factor may be the expected effect size. The smaller the expected effect size, the more meaningful it is to conduct a comparative study with randomization. Before deciding on a comparative study without randomization, it should therefore be checked whether sufficiently valid results can be expected with this study design.

If a decision is made in favour of a comparative study without randomization, the question as to whether to work with a historical or a parallel control arm should be clarified. Historical controls are possible if the patient populations observed in the past are sufficiently similar to the currently treated patient population and if data of sufficient quality for a meaningful comparison have been collected in the past. In addition, the specific data relevant to the current study must be available from the historical data set (e.g. individual patient data on confounders) and the data must have been collected in a sufficiently similar way (e.g. similar outcome definition). This requirement becomes a regular problem, for example, when populations are to be studied that are characterized by certain biomarkers that were not recorded in the past, or when individual patient data are not made available from the historical data set for the necessary confounder adjustment.

If it is to be expected that the new drug will not be available in parallel in all regions or centres in Germany immediately after market launch (e.g. due to the complexity of the intervention and the associated requirements for quality-assured usage), a prospective design with arms that are not parallel in time and/or place can be considered: The allocation to the new drug would then be possible at a sooner or later date in the respective regions or centres, depending on availability. Thus, the allocation would be considerably influenced by the availability of the new drug and not primarily by patient criteria.

Special requirements for comparative studies with randomization

A randomized study may not be necessary if very large (dramatic) effects are expected or if it is sufficiently certain that the outcome of interest will not be met at all under the appropriate comparator therapy (e.g. sustained virological response [SVR] in hepatitis C if no active treatment options are available, as e.g. in the assessment of sofosbuvir in pretreated adolescents with genotype 2 or 3 [19]).

As in the prospective study with non-parallel study arms described above, such a design is also theoretically conceivable as a randomized design (stepped-wedge design [20]). Randomized is the point in time from which a new drug can be prescribed in a certain location (region, centre).

For benefit assessments of drugs in Germany, however, such studies are less realistic due to the general prescribability of new drugs at market launch.

Special requirements for adjusted indirect comparisons via a common comparator

Before deciding on an adjusted indirect comparison via a common comparator, it should be checked whether in principle studies are available that allow such a comparison (comparative studies with randomization). Since the aim is to assess the studies on the basis of routine practice data, the studies to be included should fulfil the criteria for routine practice data mentioned in Section 5.1. If the search for relevant studies for an adjusted indirect comparison shows that the evidence base is inadequate and that the gap cannot be closed in the near future (see Section 5.3.3), a different design must be selected to answer the research question.

5.3.2 Study planning

General requirements

Regardless of the study design chosen, a study protocol must be developed and finalized. Ensuring adequate study planning without knowledge of the data is easier to ensure with prospective study designs than with retrospective ones. In prospective studies, the requirement to finalize a study protocol before the first patient is included in the data collection is generally established. Even in retrospective data collections, the study protocol should be finalized before the data are accessed. Likewise, regardless of the study design, changes in the study conduct should be documented in protocol amendments. These changes should be made either without knowledge of the data or within the framework of an adaptive design with appropriate statistical validation.

Prior to the start of data collection, the study should be registered in a study registry; such registration is possible regardless of the study design, e.g. at ClinicalTrials.gov [21].

The planning of the study also includes the planning of statistical analyses of the data. The sample size and duration of observation must be defined in such a way that the study is able to answer the research questions of the benefit assessment. The completion of an analysis plan before the data are known is essential and must therefore be documented in a comprehensible manner.

An adequate study protocol that is also comprehensible in its timing is of decisive importance for the informative value of the results of a study. The protocol, including the amendments and the analysis plan, must therefore be part of the study results report (see Section 5.5.3). It is recommended that the planning of the study be made transparent by publishing the complete study protocol and analysis plan (e.g. in a study registry).

The study protocol (and the final study report) should describe how it is ensured that the patients selected for the study meet the requirements for “routine practice” within the research question under investigation and how the study participants are identified, selected, included and retained in the study in order to minimize selection bias. The methods by which an unbiased

and systematic data collection from all study participants are to be achieved should also be described.

Special requirements for comparative studies without randomization

For the planning of comparative studies without randomization, to determine treatment effects the explicit replication of the planning of comparative studies with randomization is recommended ("emulation of target trials" [22,23]). In this context, initially a concept for a randomized study is defined that would be suitable to investigate the given research question. In the next step, the aim is to replicate the characteristics of this randomized study from the data set of the observational study.

The following table shows the components of the protocol for the emulation of the target trial.

Table 5: Components for the emulation of the target trial from a non-randomized data set

Component of protocol	Requirement for the data set selected
Inclusion criteria for patients	It must be possible to check the various inclusion/exclusion criteria on the basis of the data available in the selected data source (if, for example, patients should not have received certain treatments in the 2 years prior to inclusion in the study, appropriate observation periods and information on the treatments used must be available).
Treatments under investigation	Patients who meet the inclusion/exclusion criteria are allocated to the intervention they received at the beginning of their treatment for the disease or therapeutic indication under investigation (e.g. second-line therapy), (new user design).
Treatment allocation	In order to approximate the similarity of treatment groups in terms of prognostic factors required for the comparison (which is not achieved by randomization in this design), adjustment for the relevant confounders is necessary. If the data set of the non-randomized study contains insufficient information on the relevant confounders, it is not possible to replicate a randomized treatment allocation and thus to approximate a fair comparison. In this case, the results of a study are usually not meaningful for a benefit assessment.
Duration of observation (including clear definition of start of observation)	At the start of the observation (in the target trial as well as in the non-randomized study) 3 conditions should be fulfilled: the inclusion and exclusion criteria are met, treatment begins and the recording of the outcomes begins. These conditions must be derivable from the data set used. For example, it is necessary that details are available on the dates for the start of treatment and the recording of outcomes. The data set should contain the outcomes over the planned observation period.
Outcomes	The relevant outcomes should be contained in the data set.
Comparison of interest	The comparison of interest should specifically be prespecified (e.g. the intention-to-treat analysis of the predefined outcomes after 2 years).
Analysis plan	An analysis plan should be prespecified and the depiction of the planned analyses from the data set should be possible (e.g. recording of outcomes independent of changes in treatment during the observation period to enable an intention-to-treat analysis).

According to this concept, the data set envisaged for the comparative study without randomization must contain the necessary information to emulate the target study.

Adjustments aiming to compensate for the influence of the lack of similarity of the treatment groups in terms of prognostic factors are essential for the data analysis of studies without randomization. In order to avoid a results-driven analysis, the relevant confounders and the process of adjustment during the analysis must be comprehensible in the study planning and be pre-specified in the required level of detail (see also Section 5.5.3).

Special requirements for comparative studies with randomization

The requirements for planning the conduct and analysis of randomized trials are established (e.g. [11]). These requirements must also be implemented for the collection of routine practice data in comparative studies with randomization. In order to ensure the collection of routine practice data, the inclusion/exclusion criteria for patients, the application of the interventions, and the conduct of visits to the doctor's office must be adapted to routine practice. Randomized studies that meet these requirements are called pragmatic randomized trials [24-26]. Tools are available to support the planning of such trials [27-29].

Special requirements for adjusted indirect comparisons via a common comparator

An adjusted indirect comparison via a common comparator draws on existing randomized studies. In order to enable the assessment of the drug on the basis of routine practice data, the underlying studies of the indirect comparison should be designed to collect such data.

The study pool for such a comparison must be generated through a systematic literature search. The study design must correspond to that of a systematic review, while the analysis plan must meet the special methodological requirements for indirect comparisons. Corresponding standards are available [15,30-33].

5.3.3 Data collection

General requirements

When selecting the data collection tool, it must be taken into account whether the chosen tool can provide the data (components of PICO) necessary to answer the research question in sufficient quality.

For studies that use data from different data sources (data linkage), a detailed description of the data sources used, the information required for successful data linkage, and the algorithms used and their verification are necessary [34]. Before using data from different data sources, it must be checked and documented in a comprehensible manner whether linkage is possible at all or whether, for example, differences in the definition of data points stand in the way of joint use.

For each data source, suitable tools must be used to ensure that the data to be collected are sufficiently valid (e.g. plausibility analyses, automated checks, source data verification, use of

standard classifications and terminology, option of an audit) and have sufficient integrity (e.g. availability of log files).

Special requirements for comparative studies without randomization

For the analysis of comparative studies without randomization, data for confounder control are required in addition to the components of PICO. The relevant confounders must be systematically identified (e.g. on the basis of scientific literature with the involvement of subject experts) and pre-specified in the study protocol (see Sections 5.3.2 and 5.5.2). The availability of corresponding data in the selected data source must be ensured before deciding on a non-randomized study. Adjustment only for the confounders available in the data set is insufficient if they fail to cover the relevant confounders.

When conducting a comparative study without randomization, it is possible to collect the data retrospectively or prospectively or in combination (partly retrospectively and partly prospectively). Retrospective data collection only makes sense if the data set on the basis of which the retrospective data collection is to be conducted contains the necessary data in the quality required. The availability of the relevant data must be ensured before deciding on a retrospective design.

If the necessary data are not available in sufficient quantity or quality, prospective data collection is required. If possible, existing data sources can be used (e.g. indication-specific clinical registry) in which any missing data (e.g. individual outcomes) can be added to the data set in the prospective data collection.

Special requirements for comparative studies with randomization

In a randomized study, the data are generally collected prospectively. If the randomized study is to be conducted within an existing data structure (e.g. in a registry), it must be checked whether this structure provides for the collection of the information required. If this is not the case, the collection in the data structure can be extended or the missing information can be collected in parallel for the specific study (i.e. the specific conduct of a study to generate data, study-specific data collection).

Special requirements for adjusted indirect comparisons via a common comparator

An adjusted indirect comparison via a common comparator draws on existing studies. If the search for studies for such a comparison shows that the evidence base is inadequate, it may be possible to close the data gaps within the indirect comparison by means of additional studies still to be performed. In such a case, however, it is generally preferable to conduct a direct comparative study, as this can achieve greater certainty of conclusions.

Under certain circumstances, the evidence for an adjusted indirect comparison can be expanded by re-analysis of existing studies, for example, if not the entire study population, but a subpopulation of existing studies would be sufficiently similar for the indirect comparison (see e.g. [35]). This should not only be based on one’s own individual patient data. If necessary,

corresponding analyses should also be requested from other study sponsors. This option should be examined, even if it may mean that only a sub-question can be addressed.

5.3.4 Analysis und interpretation of results

General requirements

The prespecification of the planned analyses with the aim of obtaining reliable answers to existing research questions and preventing results-driven analyses is one of the standards of high-quality studies [11,14,15]. This also includes the description of possible changes to the analysis plan.

When interpreting the results, the informative value of the different study designs and the specific data quality must be taken into account. Within each study design, the certainty of results of the specific study results is also considered in the final decision on the certainty of conclusions for answering the research question [36].

The routine practice data must be suitable to answer the research question of the benefit assessment. If the specific data deviate from individual aspects of the research question (e.g. different line of treatment, dosage or mutation status), justification must be given as to why, despite these deviations, valid conclusions for answering the research question are possible. Routine practice data are suitable for a benefit assessment if deviations with regard to the key aspects of the research question (PICO) are not relevant for the effect estimate.

Special requirements for comparative studies without randomization

Since in studies without randomization, the similarity of the treatment groups in terms of prognostic factors is not guaranteed per se due to the lack of random allocation, an attempt is made to approach this similarity by means of adjustment.

The methods used to approximate the similarity of the treatment groups in terms of prognostic factors are based on certain assumptions. These assumptions should be described and, where possible, checked. The results of this check, as well as assumptions that cannot be verified, should be documented and their impact on the interpretation of study results evaluated (see Section 5.5.2).

Since analyses based on retrospective data collections are particularly susceptible to results-driven analyses and reporting (see Section 5.5.3), there are special requirements for the documentation of the planning of the analyses of these data. In this context, it should as far as possible be ensured that the planning of the analyses was conducted before the data were viewed.

Special requirements for comparative studies with randomization

In studies with randomization, the similarity of the treatment groups in terms of prognostic factors as a prerequisite for a fair comparison is ensured by the study design. It should also be

examined in studies with randomization whether other specific aspects of the design call a fair comparison into question (e.g. different treatment goals in the study arms [37]).

Special requirements for adjusted indirect comparisons via a common comparator

The prerequisite for meaningful adjusted indirect comparisons via a common comparator is the similarity of the studies included, the homogeneity of the results of the individual studies considered in the analysis, and the consistency of the estimated effects from direct and indirect evidence [38,39]. These assumptions must be checked and the correct or incorrect assumptions must be documented in a comprehensible manner. If there are strong doubts that one or more of the basic assumptions are sufficiently fulfilled, indirect comparisons should not be used [40]. Uncertainties regarding the assumptions should be investigated by means of sensitivity analyses. The final certainty of conclusions of adjusted indirect comparisons using a common comparator depends on the risk of bias of the studies included and the fulfilment of the assumptions mentioned above [36].

5.4 Data collection tools

The data collection tools described below can be used for different research questions. However, the focus of the present report is on the benefit assessment of drugs. The discussion of the data collection tools refers to their suitability for the research question of the benefit assessment and therefore addresses their use in comparative studies.

5.4.1 Study-specific data collection

Study-specific data collection can be used for routine practice comparative studies with and without randomization.

The advantage of study-specific data collection is that the data to be collected can be tailored to the needs of the study. Standards for study-specific data collection have been established. This type of data collection will therefore not be discussed further in this report.

In the literature, the combination of study-specific data collection with the use of other data collection tools such as registries or electronic patient records is under discussion [41]; according to this discussion, the gaps that exist in registries or electronic patient records are to be closed by additional, study-specific data collection.

5.4.2 Studies based on data collection in a registry

Data collection in a registry can be used for routine practice comparative studies with and without randomization.

In recent years, the goal and documentation scope of registries have been expanded [3]. For the present report, the increasing documentation of clinical information in registries covering patients with specific diseases is particularly important because this can be used to describe the PICO for benefit assessments. Registries are thus likely to provide more relevant information

for benefit assessments than electronic patient records or statutory health insurance (SHI) routine data (see Sections 5.4.3, 5.4.4 and 5.5.1.2).

Conceptually, it is important to distinguish between registries (active, prospective, standardized documentation of observation units on predefined but expandable questions over time for which a precise reference to the target population can be presented [3]) and studies in these registries (registry studies). While the registry (in the best case) exists as a permanently available and continuously maintained infrastructure and documentation, registry studies are conducted on a specific research question (e.g. for the benefit assessment of a new drug) with the support of a registry.

In principle, non-interventional and interventional comparative studies are possible in registries. In registries, comparative studies either with or without randomization can be conducted. While the majority of comparative registry studies are currently conducted without randomization, registry studies with randomization are described as an important tool for the efficient collection of reliable evidence [42-44]. A number of examples show the potential of this use of registries [44-48].

Since registry studies are particularly important for the collection of routine practice data with the aim of generating evidence for the benefit assessment of drugs, the specifications contained in G-BA commissions with regard to quality criteria for the routine practice data to be collected focus on registries (see Section 5.5.1).

5.4.3 Studies based on data collection in electronic patient records

Data collection in electronic patient records is conceivable for routine practice comparative studies with and without randomization. The systematic use of these diagnostic and treatment data (which are digitally documented in routine practice) for research purposes is seen by some as a future model for a “learning health care system”, but especially also for drug development, approval, benefit assessments and pharmacovigilance [49]. Compared to studies with primary data collection, analyses based on electronic patient records are expected to lead to more cost-effective and faster longitudinal analyses of populations that are not highly selective [50]. However, data from such records are not documented for research purposes, and therefore documentation is less stringent, standardized and complete. Analyses based on these data are also susceptible to various forms of bias and confounding to varying degrees [51-53].

The possibilities of using data from electronic patient records are also discussed in the context of decisions on the approval of drugs and medical devices. Various regulatory recommendations have therefore been made [54-56]. They address various questions of compliance with the standards of good clinical practice (GCP), which through the use of electronic patient records arise in a partly novel way with regard to the transparency and quality of research and the preservation of data protection and safety.

Definition

Internationally, the term “electronic patient record” is still understood differently. Haas 2017 [57] speaks of a “historical confusion of terms”, which is accompanied by a variety of terms and different approaches to the definition. For example, it is understood to mean either an institutional digital file of a health care facility (e.g. hospital, medical practice) or a cross-institutional file on a patient [58]. Opinions also differ as to the structure, content, and functions of such a file, the access options, who fills in or manages the file, e.g. the patients themselves, their medical caregivers and health care facilities or the health insurance fund. Increasingly, however, national concepts for overarching so-called electronic health records (EHR) are being established; these are fed from routine clinical documentation within digital files of individual health care facilities (electronic medical records, EMR).

As database applications, across cases and sectors, cross-institutional EHR aim to continuously collect or link patient-related data on medical history, treatments, medication, diagnostic findings, special risks and other health characteristics. They can be linked or aggregated for large patient populations in electronic EHR platforms for analysis and research purposes.

Electronic patient records vary across a wide range of characteristics. In the simplest case, it may be a digital archive that is limited to one health care facility and contains a few PDF documents of a patient (supplemented by X-rays) that are difficult to search and analyse. Or, ideally, it could be a long-term, easily linked and analysable file with a comprehensive, highly formalized and structured collection of validated data that has been granularly coded according to standard terminology, and that, as an integral element in a patient record system, also provides various functions for patients, medical providers and researchers.

International and German situation

An expertise [59] commissioned by the Organization for Economic Cooperation and Development (OECD) recently showed that probably only 10 of 28 nations analysed worldwide (without Germany) will be able to provide a solid basis for health reporting and medical research in the near future by means of their solutions for electronic patient records. This is due, among other things, to a lack of organizational-political and technical prerequisites, data protection regulations, insufficient coverage, incomplete and insufficiently standardized data collection or insufficient interoperability of the information technology (IT) solutions used. Investigations by EMA researchers focusing on the purposes of drug approval [60] show that data sets based on electronic patient records also vary greatly in Europe, that they currently only rarely meet regulatory requirements, that there is a lack of transparent information on them, and that they are unequally distributed (establishment mainly in Central, Western and Northern Europe). This clearly limits the number of national EHR platforms theoretically available in the future for benefit assessments of drugs.

For some years now, various international associations of industry, research, patient organizations and health authorities have been organizing themselves. Especially for the purpose of research and drug development, they want to create new possibilities through platforms based

on electronic patient records [61-63]. The goals include improved research planning, patient recruitment, study conduct, and generalizability of results by bringing research closer to routine practice.

In Germany, the SHI funds are required to offer their insured members an electronic patient record by 2021, for which preparations and coordination processes have been underway for some time. However, experts [57,64] have identified numerous obstacles to the digitalization of the German health care system and relevant hurdles for the realization of the national electronic patient record project. In this context, the below-average digitalization in the hospital sector compared to the rest of Europe is also of great importance [65].

Since the planned national solution for the electronic patient record envisages that patients themselves will determine which data or which treatment episodes will be stored in their records and who, in addition to themselves, is to have access rights, the following question arises: In future, how many patient records will a) be complete enough for use in research and b) generally be made available for this purpose by patients? It seems certain that the electronic patient record planned for Germany does not currently, and probably will not in the medium term either, offer any options for conducting collections of routine practice data to determine the added benefit of drugs.

In Germany, too, current research initiatives exist for the patient-related linking of institutional electronic patient records with other digital data sets, e.g. from biobanks, SHI funds or digital health applications [66]. It is not foreseeable at present whether or when these initiatives will result in reliable possibilities for conducting collections of routine practice data with the databases created.

Validity of data from electronic patient records

It is known that, given the diversity of electronic patient records, their data quality also varies greatly and is influenced by many factors [67,68]. For secondary use of these data, such as in clinical research, it is therefore recommended that the data quality and suitability for the respective analysis purpose is thoroughly examined in each individual case (“fit for purpose”) [41,68]. There are also widespread calls for the validity of data from electronic patient records and other so-called “real world data” to be systematically checked in specially established study programmes and to develop standards for this purpose [60,69-72]. However, the current transparency regarding the validity and quality of data from electronic patient records (used in RCTs, for example) is considered to be completely inadequate [73].

In routine practice, entries in electronic patient records are often made under a heavy workload by a large number of physicians and nurses who are not primarily committed to research goals and the associated data quality requirements. This therefore explains results shown in an analysis by Brennan 2012 [74] of health services usage coded according to the English Hospital Episode Statistics (HES), which is based on electronic hospital documentation. The authors reported that according to the coding in the HES data, 1600 adults apparently attended

outpatient child and adolescent psychiatry services in England each year, while 3000 children and adolescents attended outpatient geriatric services, and 17,000 men were admitted to obstetric wards.

In the USA, Green 2013 [75] used a national database on hospital treatments that is based on information from electronic patient records to investigate emergency department visits by 875 patients with a coded intubation procedure during their emergency department visit. In 27% of cases, he identified implausibilities, which after a comparison with text notes were presumably caused by coding errors, because, according to the data, 81 intubated patients were apparently immediately discharged and went home and 153 were immediately transferred to a normal ward.

In a joint attempt by 4 healthcare providers in the USA to replicate a case study of comparative effectiveness in different anti-hypertensive drug treatments with clinical data from their electronic patient records, their qualitative analysis identified 5 main problems with the extracted data: "missing data, erroneous data, uninterpretable data, inconsistencies among providers and over time, and data stored in noncoded text notes" [76].

According to a systematic search for methods and dimensions of data quality assessment in the context of EHR data reuse for research [68], there is currently no agreed standard on how the 5 data quality dimensions (completeness, correctness, concordance, plausibility and currency) derived from the literature review should be determined systematically and by means of statistical methods. Possible methods include "comparison with gold standards, data element agreement, data source agreement, distribution comparison, validity checks, log review, and element presence" [68]. According to the authors, "if the reuse of EHR data for clinical research is to become accepted, researchers should adopt validated, systematic methods of EHR data quality assessment" and that "… the majority of the studies we identified relied upon an 'intuitive' understanding of data quality and used ad hoc methods to assess data quality" [68]. The guideline of the International Society of Pharmacoepidemiology is a good starting point for systematic quality checks of electronic databases [77].

A major problem of electronic patient records is that a lot of important information on studies (e.g. inclusion criteria, outcomes) is only documented in text fields in a non-standardized way [78]. Manual analyses are considered too time-consuming, not least because a large number of different terms are used for the same facts in the medical notes. Wassermann 2011 [79] reported, for example, that on one day in a children's hospital in Philadelphia, in 465 notes in patient records of children with otitis media, the doctors used 278 different ways of expressing fever (e.g. fever, pyrexia, 39.2 degrees, elevated temperature) and 123 different ways of expressing ear pain in 213 patients. Extracting such data in sufficient quality for coding purposes automatically from free-text notes with text-mining applications and natural language processing (NLP) tools will certainly remain a major challenge in the near future. This is also shown by a recent German publication on the automated analysis of computed-tomography (CT) findings within an epidemiological study. In the conclusion, under the aspect of data

quality for future studies, the authors stated that “the combination of human intellect and intelligent, adaptive software appears most suitable for mining unstructured but important textual information for research” [80] .

PICO

The research questions in the benefit assessments of the G-BA are regularly structured according to the PICO scheme, which requires the most precise description possible of the patient population of interest, the intervention to be assessed and its control intervention (comparator), as well as the relevant outcomes. Important aspects for the precise depiction of PICO by data from electronic patient records are discussed below.

Population

For the determination of the patient population in the context of the collection of routine practice data, an electronic patient record should contain the necessary information on inclusion and exclusion criteria (emulation of target trial, see also Section 5.3.2).

Köpcke 2013 [81] examined electronic patient records of 5 German university hospitals to determine whether they contained sufficient information on the protocol-based inclusion and exclusion criteria of 15 randomly selected clinical studies. On average, the records, which varied from clinic to clinic, contained data fields for documenting 55% of the patient characteristics that would have been required to determine inclusion and exclusion criteria. If one also considers the data completeness of 64% in these fields, the proportion of patient characteristics that were depicted in the digital files was only 35% of the patient characteristics required for patient selection. Additional extensive data collection, for example, from parallel documentation (electronic or paper), would have been unavoidable for the definition and recruitment of the patient population within the collection of routine practice data based on electronic patient records stored in the inpatient sector.

Lau 2011 [82] compared data from electronic patient records of 52 community oncology clinics with registry and claims data (Surveillance Epidemiology and End Results [SEER] database of the National Cancer Institute; Medicare and commercial claims data). They found that 70% of the information on tumour stage and 40% on ethnic classification was missing in the electronic patient records. Thus, the incompleteness of important data was again shown to be problem for the determination of inclusion and exclusion criteria based on electronic patient records.

Intervention / comparator

The aforementioned study in the 5 German university hospitals [81] also showed that structured data on treatments and especially on medication hardly existed in their electronic patient records. Standardized and coded claims-relevant inpatient operations and procedures that are

standardized via the Operation and Procedure Classification System (OPS[10]) were an exception. Even according to international experience, data on medication prescriptions in electronic patient records mostly originate from the outpatient sector, which was confirmed by a recent EMA study of 34 electronic databases in Europe, of which only 2 (5.9%) contained prescriptions from hospitals [60]. However, since new and innovative drug therapies are often started in the inpatient sector, the lack of precise prescription data in hospitals means that the possibility of collecting routine practice data on the basis of electronic patient records is considerably limited.

Outcomes

Data from electronic patient records depict routine practice. This is why important data for the benefit assessment of drugs, such as data on symptoms and adverse effects, are incomplete in electronic patient records. In addition, they are usually not coded in a standardized way (e.g. coded according to the Medical Dictionary for Regulatory Activities, MedDRA). Rather, such information is usually contained in the unstructured text notes on the course of treatment and would have to be extracted and secondarily coded from there for research purposes with great effort [76]. Serious adverse events may be depicted as International Statistical Classification of Diseases and Related Health Problems (ICD) codes. However, for three major pharmacovigilance initiatives in the USA and Europe, Moore 2015 [83] established that both in health insurance data and data from electronic patient records, the specificity of the data on adverse effects and adverse events depicted via ICD-9 codes was shown to be highly variable and overall insufficient in previous validation studies, depending on the data basis. For instance, in a medical record validity study of the data basis for the FDA-launched Sentinel Project, it was shown that of the cases coded as severe acute liver injury, only 24.7% could be confirmed by medical records; of the 56 cases coded as acute liver failure (a subset of the severe liver injury cases) just one was confirmed. Reliable information on sensitivity, i.e., how many of the actual adverse events are at all depicted in the EHR-based data sets, is scarce in the absence of appropriate validation studies.

Certain data, such as those on health-related quality of life and PROs, are generally not part of electronic patient records, or related information is only found as unstructured information in the text notes of the treating staff. Thus, important outcomes are not included in these records.

Confounding and effect modification

Since the content and data fields in electronic patient records are extremely variable, for each research question of a comparative benefit assessment it must be checked individually whether the corresponding data required to generate confounder variables are available and, if yes, in what quality. This applies in the same way to possible effect modifiers.

[10] *Operationen- und Prozedurenschlüssel* (German modification of the International Classification of Procedures in Medicine, ICPM)

Representativeness / applicability

Various aspects are important for the question of the applicability of international study results based on electronic patient records to SHI-insured persons in Germany, including the comparability of populations. According to the results of the aforementioned study by Lau 2011 [82], selection effects already play an important role at the national level for data from EHR sources. For 6 tumour entities studied, the comparison of data from more than 160,000 electronic patient records with registry and claims data showed systematic differences in the distribution of tumour types, ethnic classification, and outpatient oncological therapies. As an explanation, the authors cite geographical differences in the data collections, as well as the fact that certain groups of cancer patients are less often treated in special oncological facilities because they are only treated by their office-based specialists. For example, this is the case for many older prostate cancer patients who receive hormonal treatment from their urologists and who are therefore, for example, more strongly represented in Medicare data. The study makes it clear that, as long as analyses are not based on a nationally representative database of cross-institutional electronic patient records, the question of the generalizability of study results based on data from EHR sources already arises at the national level. In addition, the applicability of international EHR studies to the respective German patient group from the SHI population must be evaluated.

Conclusion

Against the background of

- the lack of cross-institutional national electronic patient records that are suitable for research purposes
- the very often questionable quality of data extracted from electronic patient records
- the limitations of data from electronic patient records mentioned under the issue of the PICO scheme, and
- the questionable applicability of international studies based on electronic patient records

collection of routine practice data from electronic patient records for benefit assessments according to §35a SGB V is currently not considered realistically feasible.

5.4.4 Studies based on data collection from claims data of health insurance funds

In principle, studies based on data collection from claims data of health insurance funds are conceivable both for routine practice comparative studies with or without randomization.

The present report focuses on the data collected for claims purposes in the German SHI system, so-called SHI routine data. A good overview of the existing data in the claims data of the social insurance providers is given in Chapter 2 of an expert report prepared for the German Institute

for Medical Documentation and Information (DIMDI[11]) [84]. This report presents the advantages and disadvantages of SHI routine data. Chapter 4 of the report also describes why the claims data of the approximately 40 private health insurance companies are probably incomplete (especially for outpatient services and drug prescriptions) due to a different claims system and a large variety of optional tariffs (e.g. with deductibles and exclusions). Therefore, they are to be regarded as unsuitable for research purposes in the context of benefit assessments. In the following text, key aspects are discussed that are particularly relevant for the benefit assessment of drugs using SHI routine data.

Validity of SHI routine data

In the context of research with SHI routine data, the validity of the diagnoses given according to the ICD classification is of particular importance, since otherwise no clinical details are documented for the services billed. This diagnosis data can be validated externally in comparison with other sources (e.g. hospital records, medical practice documentation) or internally. In an internal validation in a data set it is checked, for example,

- whether diagnoses are congruent with drug prescriptions (e.g. type 1 diabetics with insulin prescription)
- whether inpatient and outpatient diagnoses match
- whether diagnoses of irreversible chronic diseases persist over time, or
- whether claims for newly diagnosed diseases included the billing of the diagnostic procedures to be expected in direct chronological proximity to the time of diagnosis [85].

Although external validation in the sense of source data verification is of greater informative value than internal validation, systematic comparisons of SHI routine data have rarely been performed or published.

For the primary care sector, they point to a relevant degree of underreporting, over-reporting and miscoding. For example, in a random sample of 250 patients from 10 general practitioner (GP) practices, Erler 2009 [86] examined whether, on the basis of the patients' medical records, health problems treated within a one-year period matched claims-based diagnoses. They found a 40.1% agreement with ICD-10-coded diagnoses from claims records for 2158 treatment events. In 29.7% of the cases, reasons for treatment, in particular screening tests, psychosocial counselling services and non-specific illnesses and symptoms did not result in claims-based diagnoses ("under-reporting"), while in 19%, diagnoses of illnesses were found in the claims-based data set that, according to the patients' records, had not provided a reason for treatment in the relevant period ("over-reporting"). In 11.2% of cases, the ICD-10 codes of claims-based diagnoses and the diagnoses in the records did not match ("erroneous codes"). However, for 6 of the diagnoses most common in GP practices, the correctness of the routine data (specificity)

[11] *Deutsches Institut für medizinische Dokumentation und Information*

was between 71% and 93%, at least on a rough coding level (3-digit code). The quality of GP documentation in 6 practices regarding diagnosis data in patients with thyroid disease was the subject of a study by Münch 2016 [87]. In 548 patients who were permanently prescribed thyroid medication, the diagnosis was incomplete in 147 (26.8%) and not documented at all in 100 (18%) patients.

Internal validation studies from the outpatient sector also point to a considerable degree of inconsistency and associated uncertainty regarding the validity of diagnoses, even for well-defined diseases [88-92].

Against the background of morbidity-orientated risk structure compensation [93] between SHI funds, in the past some of these funds have provided financial incentives to "correct diagnosis coding" for groups of office-based physicians in special contracts. This has long been the subject of controversial discussion in health policy and has recently led to various legal initiatives. In order to reduce concerns about the susceptibility of outpatient SHI routine data to manipulation in this regard, the Appointment Service and Health Care Act (TSVG[12]) of 2018 stipulates the binding introduction of coding guidelines for outpatient diagnoses and procedures with effect from the beginning of 2022. Electronic data processing (EDP) integration into the certified practice management systems of medical practices is planned accordingly, and uniform nationwide test standards for coding are to be created. The effects of these legal measures on the validity of outpatient diagnosis data cannot yet be predicted with certainty.

Since the introduction of diagnosis-related groups (DRGs) in Germany in 2003, the quality of coding in the inpatient sector has improved steadily as a result of a continuous review process by the SHI funds and their medical services (MDK[13]), also with regard to claims-based diagnoses [94]. Parts of the official hospital statistics are therefore now based on routine data such as discharge diagnoses. Binding DRG coding directives are constantly being refined and adapted. Approximately 10% of hospital bills are checked by the MDK, although doubts about the accuracy of a primary or secondary diagnosis are relatively rarely a reason for this; instead, other issues such as possible misallocation are in the foreground [95]. Diagnosis data from the hospital is not completely represented in the routine data per se, because according to coding regulations, only those secondary diagnoses that increase the need for treatment may be reported in the claims data.

In the context of a patient safety study, Maas 2015 [96] reviewed 3000 patient files of German hospitals and compared them with the corresponding claims data. There, only 72.7% of cases of high-grade pressure ulcers, 25% of postoperative deep vein thromboses, 44.8% of wound infections, 20.8% of heart attacks, and 23.3% of pneumonia had been documented during the

[12] *Terminservice- und Versorgungsgesetz*

[13] *Medizinischer Dienst der Krankenkassen*

inpatient stay. Despite low sensitivity, however, at 99% to 100% the accuracy of the information on these adverse events in the routine data was high in terms of specificity.

The validity of SHI prescription data (quarterly data on prescribed drugs according to the Anatomical Therapeutic Chemical [ATC] Classification System, prescription and dispensing date, package size) is generally considered to be relatively high and stable. For the period from 2000 to 2006, Hoffmann 2008 [97] described a further improvement in the already high data quality, e.g., in the documentation of the dates when the prescriptions were issued and the corresponding medications dispensed.

Representativeness of SHI routine data

Although the SHI routine data is available for all 73 million SHI patients in Germany and thus for the vast majority of the population (completeness), it is not combined in a single body of data, but is only available separately for the currently existing 109 SHI funds [98]. Several studies [99-101] prove beyond doubt that the members of SHI funds in Germany differ, sometimes considerably, even after adjustment for age, gender and co-morbidity. For example, Schäfer 2017 [101] showed in an adjusted comparison that melanoma patients insured with the largest SHI fund (AOK[14]) or the agricultural SHI fund more often suffer from a considerably more advanced stage of disease at the time of initial diagnosis than those insured with other SHI funds. Only analyses of routine data sets containing insured persons of all SHI fund types from all regions would allow generalizable conclusions. Occasional attempts to extrapolate the results determined by specific SHI funds to the entire SHI population must be viewed critically because of the unclear effects of bias [100]. DIMDI's data body, which is currently the only one that is organized across all types of SHI funds according to §303 SGB V, does not have up-to-date data or the necessary completeness and data depth for benefit assessments [102]. In summary, with regard to their use in benefit assessments, it can be said that analyses based on SHI-fund-specific routine data lack the necessary representativeness for the entire group of SHI-insured persons.

Confounder control in SHI routine data

For an analysis of SHI routine data in comparative studies without randomization, an adjustment of the comparison between intervention and control would be necessary. As mentioned above, the SHI routine data lack important information for this adjustment with the aim of controlling confounding. This includes clinical and genetic data, diagnostic findings such as laboratory values and other test results, and information on disease severity. Anthropometric data such as weight and height are not documented, so that, for example, the body mass index cannot be calculated as a possible confounder. Data on socioeconomic status are only partially available, and information on lifestyle factors (e.g. smoking) is missing [84].

[14] *Allgemeine Ortskrankenkasse*

This means that in many cases it is not possible to adjust comparisons, or only to an insufficient extent.

Depiction of PICO from SHI routine data

The research questions in the benefit assessments of the G-BA are regularly structured according to the PICO scheme, which requires the most precise description possible of the patient population of interest, the intervention to be assessed and its control intervention (comparator), as well as the relevant outcomes. The following text describes the extent to which SHI routine data can be used to precisely depict PICO.

Population

The identification of patients/insured persons belonging to a specific population is conducted in SHI routine data via a "case definition". This is primarily based on sociodemographic (age, gender) and ICD diagnosis data, but often needs to be supported by additional criteria such as the documentation of disease-specific medication or other prescriptions over several quarters to increase validity. This is because the SHI routine data do not contain any clinical information that, for example, beyond a possible ICD coding, depict symptoms or the severity and stage of the disease. The lack of laboratory data or genetic data also makes it difficult to narrow down the population more precisely by means of SHI routine data. However, genetic data play an increasingly important role in the benefit assessment of drugs through "personalized" treatment approaches. In practice, diseases are often only coded with "9" at the level of the 4th digit of the ICD, which may form a larger residual class (other, not specified) and in many cases, this makes a specific definition of a population impossible [84].

If the temporal reference of the disease plays a role in the specification of a population, it is important to know that in the outpatient sector, diagnoses are documented without an exact date and only in a quarterly reference. In contrast, in the inpatient sector they are depicted by the admission and discharge date, at least for newly occurring diseases and acute events. The clear definition of newly ill (incident) patients is difficult, because the distinction from prevalent cases can only be made by a longer retrospective consideration of the previous years (without a corresponding diagnosis).

Since diagnoses are of outstanding importance in the case definition in SHI routine data, the problems described in the section on validity presumably markedly limit the possibility of forming sufficiently delimited and at the same time complete patient populations from the data set for the benefit assessment. According to the current state of knowledge, it is not possible to ensure a sufficient certainty of conclusions for benefit assessments on the basis of these data.

According to DIMDI, the ICD-10 system only contains specific codes for rare diseases for about 355 of more than 8000 diseases (approx. 4.4%) [103], which in most cases makes it impossible to identify these patients in SHI routine data via diagnosis data. Currently, a double coding system is being discussed for the ICD-11 version, which aims to alleviate this problem [103]. It is currently unclear to what extent double coding can contribute to the solution.

In summary, against the background of the usually very specific research questions posed by the G-BA in drug assessments with SHI routine data, it will rarely be possible to identify the study populations with the precision required for benefit assessments, e.g. by genetic characteristics, disease severity or stage of treatment.

Intervention / comparator

Whereas inpatient DRG claims data contain codes that precisely describe operations and procedures performed according to a standard code (OPS), information on inpatient drug therapy is missing in SHI routine data (apart from very rare exceptions of high-priced special drugs). Drug therapies started, continued or modified in hospital are therefore not included in the data.

In the outpatient sector, quarterly billings with the SHI system contain information on prescribed drugs (according to the ATC code), the date of the prescription, the date of dispensing in the pharmacy and the package size. As mentioned above, these data are valid in themselves, but do not allow sufficiently precise conclusions to be drawn on exposure, e.g. whether dose or combination changes or treatment switches or discontinuations occurred. To this end, exposure-estimation calculations based on the prescription pattern and certain assumptions are typically performed in routine data analyses [104]. Over-the-counter (OTC) drugs and drugs obtained through private prescriptions are not included in the SHI routine data.

Non-drug active comparator therapies (e.g. surgical procedures, remedies, psychotherapy, etc.) are documented in varying degrees of detail in the SHI routine data according to claims regulations for the respective health service area.

In summary, it can be said that, for a data collection to precisely describe intervention and comparator treatments within a comparative benefit assessment of drugs, SHI routine data have general and, in particular, sector-specific gaps that make such analyses appear problematical, depending on the research question and the demands on the certainty of conclusions.

Outcomes

Mortality as a clear outcome is reliably recorded in the SHI routine data in the master data of the insured persons as a reason for termination of membership.

Due to the lack of clinical data, laboratory data, data from diagnostic imaging, and other test results, there is only limited direct information on morbidity within SHI routine data. However, since the benefit assessment of drugs is typically based essentially on the precise recording of symptoms and complications of the disease or the treatment, the lack of clinical data on symptoms (nausea, headaches, diarrhoea, etc.) is a major shortcoming in SHI routine data for this purpose. Adverse events are not fully recorded like in clinical trials and are not coded using standard terminology such as MedDRA. Instead, recording is only very incomplete and moreover not accurately represented via the ICD coding in the outpatient (only quarterly data without dates) and inpatient area. As described in the section on validity, even diagnoses of

serious adverse events are not reflected in the SHI routine data in a sufficiently sensitive manner, for example in hospitals [96]. In a study by Kuklik 2017 [105], a comparison of medical records with claims data was conducted in 4 German hospitals, focusing on the sensitivity and specificity of ICD-10 German Modification (GM) codes for the designation of drug-therapy-related adverse events acquired in hospitals (e.g. L27.0 – "Generalized skin eruption due to drugs and medicaments"). Out of 807 reviewed cases with specific codes in the claims data, after studying the patient file, 65.1% were finally confirmed as a drug-therapy-related adverse event acquired in the hospital (specificity). For the determination of sensitivity, 1510 randomly selected patient files were analysed and in 358 cases, a total of 495 adverse events acquired in hospitals were identified. Of these, 246 events were depicted in the claims data, which corresponds to a sensitivity of 49.7%. These results would be insufficient for a benefit assessment.

Data on the health-related quality of life of patients or on other PROs, which are also of great importance for a benefit assessment, are not available in SHI routine data.

With SHI routine data, rougher indicators beyond individual treatment episodes could be considered over longer periods of time, which could indirectly say something about the condition and fate of the patient [106]. These include mortality, re-hospitalization and the reasons for it, duration and intensity of treatment, occurrence of or changes in the degree of need for long-term care, use of medical services, etc. With the exception of mortality, such indicators, which are mainly used in comparative quality assurance of inpatient care, do not usually correspond to the outcomes currently used for benefit assessments of drugs, also due to their operationalization.

In summary, Table 6 shows possibilities and limitations of SHI routine data for answering research questions on drug therapies according to the PICO scheme:

Table 6: Possibilities and limitations of SHI routine data for answering research questions of benefit assessments according to the PICO scheme

Population	**I**ntervention - **C**omparator	**O**utcomes
▪ Limited representativeness for data sets specific to single or selected SHI funds ▪ Differentiation by age, gender and other "auxiliary characteristics", as well as by questionably valid, often unspecified ICD diagnosis data, is only possible to a limited extent, depending on the disease ▪ No clinical and genetic characteristics, laboratory data, degree of severity data ▪ Rare diseases are only specifically depicted in the ICD 10 in about 4.4% of cases ▪ Incident cases only identifiable through longer retrospective analyses	▪ Medication, prescription and dispensing date, as well as package size in the outpatient sector are clearly documented ▪ Inpatient drug therapies not documented ▪ Exact administration (e.g. duration, dose) unclear ▪ Treatment discontinuations, interruptions, and switches are poorly depicted ▪ Non-drug add-on or comparator therapies are depicted in SHI routine data	▪ Mortality depicted ▪ Due to the lack of clinical data, morbidity only very limitedly and indirectly depicted via ICD-10 codes ▪ No depiction of symptoms ▪ Side effects (adverse events) incompletely recorded and not specifically coded (e.g. MedDRA), but only depicted via ICD codes, with poor sensitivity and specificity ▪ Exact temporal reference for outcomes such as side effects is difficult (outpatient > inpatient) ▪ No data on health-related quality of life ▪ No data on PROs ▪ Generation of mostly rather unspecific indicators for long-term treatment results
ICD: International Statistical Classification of Diseases and Related Health Problems; MedDRA: Medical Dictionary for Regulatory Activities; PICO: population, intervention, comparator, outcomes; PRO: patient-reported outcome; SHI: statutory health insurance		

International claims data

The analyses of Moore 2015 [83] and Pacurariu 2018 [60], which also refer to claims data, indirectly show that the challenges identified there in the use of claims data for research purposes are internationally congruent in their nature with those arising from SHI routine data for benefit assessments. This refers, for example, to various aspects of data quality as well as to the extensive lack of important clinical data and possible confounders. External validation studies with comparison of medical records [107] have produced sobering results, particularly with regard to the sensitivity of ICD-coded diagnoses in the claims data, and the methodological problems associated with the use of SHI data are not new [108], but were already discussed in detail in the international literature more than 2 decades ago [109]. In countries with different health insurance systems, such as the USA, an additional problem is that for employees and their relatives, a change of employer is often associated with a change of health insurance fund, and thus the continuity of information required for research can be lacking.

Conclusion

The limitations of SHI routine data summarized in Table 6 and the similar limitations and quality problems with claims data from other countries show that they are not suitable for the

robust benefit assessments of drugs as conducted by the G-BA according to the specifications laid down for this purpose.

5.5 Requirements for data and analyses

5.5.1 Criteria for data quality

This chapter describes criteria for data quality of registries. The reason for focusing on this data collection tool is that registries are of particular importance for the research question of the present report. Routine practice data from electronic patient records and from claims data of SHI funds currently play a subordinate role for the research questions of benefit assessments and will continue to do so in the near future (see Sections 5.4.3 and 5.4.4). The criteria for the quality of data from study-specific data collections are largely established as standards.

5.5.1.1 Criteria for the quality of registries according to national and international standards

Quality criteria for patient registries and registry studies

As defined in Gliklich 2014, a patient registry should be understood as "an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes" [4].

Such patient-related data collections are of various types and purposes. In addition to routine basic analyses, they can become the basis for additionally planned observational or interventional studies. In this case, we speak of a registry study, in which additional criteria beyond the basic quality requirements for a patient registry must often be fulfilled, for example, with regard to a study protocol, extended patient consent, the collection of additional data, more intensive assurance of data quality or the analysis methods (see Figure 4).

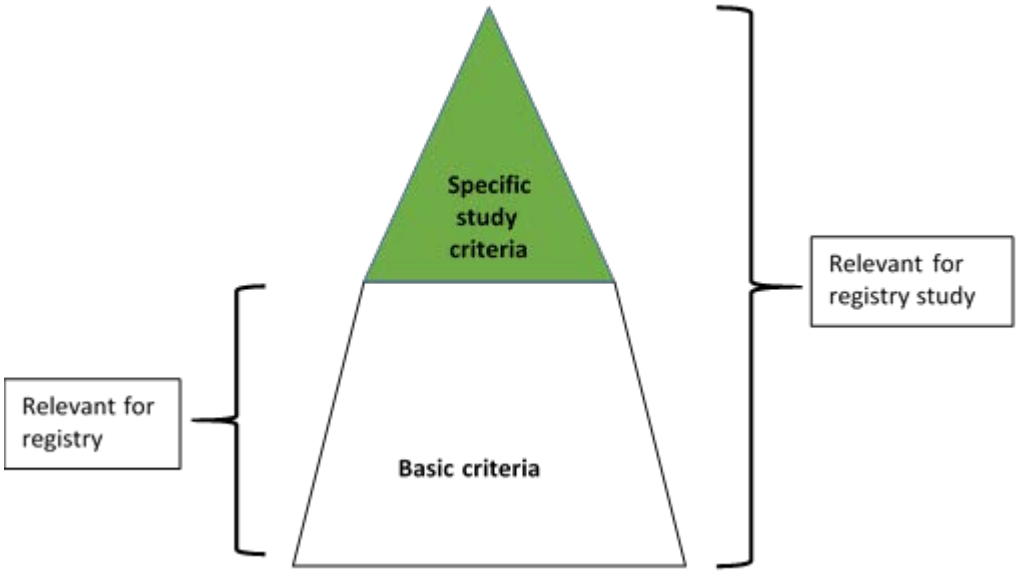


Figure 4: Quality criteria for patient registries and registry studies

Publications on the quality of patient registries generally recommend a mix of structural, process and outcome quality criteria that characterize good registry practice. As in Donabedian's quality model, the present rapid report assumes that structures and processes aim at the highest possible quality of outcomes, which thus "remain the ultimate validators of the effectiveness and quality of medical care" [110, S. 694].

Transferred to patient registries, a structural quality characteristic would be, for example, the existence of a steering committee or a comprehensive, up-to-date registry protocol. Process quality characteristics include regular audits in local registration centres or automatic checks during data entry. Ultimately, structures and processes in a patient registry aim at a high outcome quality, which is expressed above all in key characteristics of data quality (completeness, integrity, correctness). A further aspect of the outcome quality can be seen as effective data protection or the fast provision of analysis results.

Section 4.2 described how the list of quality criteria for patient registries contained in Table 7 was developed for this report. This list also represents a mixture of characteristics of the structural, procedural and outcome quality and, like other published lists of criteria [4,6,8], also contains basic criteria for a registry and those that are of particular or even specific relevance to registry-based studies.

Table 7: Compilation of nationally and internationally used quality criteria for patient registries and registry-based studies

No.	Quality criterion
Systematics	
1	Detailed registry description (protocol)
Standardization	
2	Precise definition / operationalization of exposures, clinical events, outcomes and confounders
3	Current data plan / coding manual
4	Use of standard classifications (e.g. ICD-10) and terminology (e.g. MedDRA)
5	Use of validated standard data collection tools (questionnaire, scales, tests)
6	Training courses on data collection and recording
7	Implementation of a consensual disease-specific core data set
8	Use of exact dates for the patient (e.g. birth, death, pregnancy)
9	Use of exact dates of disease (e.g. definitive diagnosis, clinically relevant events)
10	Use of exact dates for important examinations
11	Use of exact dates for treatments / interventions (e.g. for drugs: start / stop date, dose, dose changes)
Achievement of the recruitment goal / sample composition	
12	Clearly defined inclusion and exclusion criteria for registry patients
13	Completeness of registry patients (full data collection or representative sample)
14	Strategies to avoid selection bias in patient inclusion to achieve representativeness
Validity of data collection	
15	Completeness of data per time point of data collection (loss-to-follow-up, drop-outs)
16	Completeness of data collection time points
17	Accuracy of data
18	Data consistency over time
19	Source data verification (e.g. for 10% randomly selected patients per study centre)
20	Registry monitoring by internal audits
21	Registry monitoring by external audits
22	Quality management system (if necessary, with regular collection of quality indicators)
23	Standard operating procedures for data collection

(continued)

Table 7: Compilation of nationally and internationally used quality criteria for patient registries and registry-based studies (continued)

No.	Quality criterion
Superordinate quality criteria	
24	Registry transparency (e.g. funding, decision pathways conflicts of interest)
25	Scientific independence
26	Secure funding (for planned data collection period)
27	Steering committee, executive committee
28	Currency of the registry documents (e.g. protocol, data plan, statistical analysis plan, declaration of consent etc.)
29	Respect of patient rights and data protection, consideration of ethical aspects
30	Timeliness (currentness and rapid availability of the required results)
31	Flexibility and adaptability (e.g. for embedding studies, for further data collection, in the event of changes in the health care situation)
32	Documentation trail - documentation of all process and definition changes in the registry
33	Audit trail - documentation and attributability of all data transactions
34	Linkability with other data sources
Validity of statistical analyses and reports on registry studies	
35	Public registration of the planned registry study
36	Preparation of a study protocol and a statistical analysis plan for the planned registry study
37	Prespecification of the analysis methods in the statistical analysis plan
38	Explanation of the handling of missing values
39	Adjudication committee for key outcomes
40	Adjustment of results of comparisons with regard to potentially confounding variables and consideration of effect modifying variables
41	Sensitivity analyses (e.g. for different case definitions or consideration of confounders)
42	Analysis / control of site effects
43	Report on measures to avoid other types of bias (e.g. selection bias)
44	Full report of the results
45	Publication of the results report including study protocol and analysis plan
Other possible criteria from a regulatory perspective	
46	Recording and handling of adverse events according to regulatory requirements
ICD: International Statistical Classification of Diseases and Related Health Problems; MedDRA: Medical Dictionary for Regulatory Activities	

This list of criteria can be the starting point for an overview and evaluation of the structural, procedural and outcome quality of a patient registry.

In order to assess the quality of a patient registry, the credibly presented outcome quality with regard to the representativeness, completeness and accuracy of the data is particularly relevant. In the evaluation of the suitability of a registry for conducting studies, it is therefore inappropriate to consider all individual structural or process quality characteristics of the list as

mandatory. The relationship between the individual structural and process quality characteristics and outcome quality achieved is also not linear for patient registries. Rather, it can be assumed that the more these criteria are fulfilled to a high degree, the better the outcomes to be expected. At the same time, however, even a low degree of fulfilment of individual criteria does not necessarily mean that a registry is fundamentally unsuitable for the collection of routine practice data for benefit assessments.

In a suitability test for a specific registry, the following items should therefore be assessed with the help of a list of criteria targeted towards the present case of use (benefit assessments of drugs) for the respective specific research question:

- whether and to what extent the individual criteria are fulfilled
- what influence a possible non-fulfilment is likely to have on the outcome quality, and
- whether possible deficits can be corrected within a reasonable timeframe and with reasonable effort in a registry-based study.

Such a list of criteria geared to the present case of use, the benefit assessment of drugs, can be found in Section 5.5.1.3. The list was formed by means of the above complete list on the basis of the literature analysis of nationally and internationally used quality criteria for registries, the discussions with registry operators (see Section 5.5.1.2), as well as the experience from more than 300 benefit assessments of drugs and the requirements for study design and analysis relevant for the assessments. Criteria that refer to the preparation of a study protocol and an analysis plan for a registry study, as well as to specific analyses in a registry study (e.g. sensitivity analyses), are not listed here, since these are addressed separately in Section 5.5.2 (“Methodological requirements for the analysis”).

5.5.1.2 Results of discussions with registry operators

As described in Section 4.2, a 3-part questionnaire was sent out in preparation for the discussions with the registry operators. Not all registry operators were able to complete the first 2 parts of the questionnaire (Part 1: Information on the registry; Part 2: General quality criteria for registries) before the interview. In these cases too, the questions (Part 1) and criteria (Part 2) listed in the questionnaire served as an interview guideline.

In addition to the aim of obtaining a general understanding of the structure, organization and quantitative and qualitative data stock, as well as an evaluation of the importance and degree of fulfilment of the various quality criteria, the practical experience of the registry operators was to be used to identify factors particularly beneficial or obstructive to the operation of an existing registry or to the establishment of a new one. This experience was also to be used to estimate the time required for different starting situations (new registry vs. existing registry) and different approaches (prospective data collection in the registry vs. retrospective analysis based on an already existing data stock).

In the following, the feedback of the registry operators on the quality criteria listed in Part 2 of the questionnaire is first summarized, especially with regard to the respective degree of fulfilment (self-evaluation of the registry operators) and the importance of the criteria mentioned. Subsequently, the factors are listed that were highlighted in the interviews as being particularly beneficial or obstructive for the operation of the registry.

This is followed by a general evaluation of the suitability of the registries examined for the collection of routine practice data for benefit assessments. Finally, it is described which time period is likely to be usually required for the collection of routine practice data for benefit assessments.

Fulfilment of the general quality criteria from the perspective of the registry operators

The quality criteria mentioned in the questionnaire were all known to the registry operators and were all considered useful in principle, albeit partly in a modified form (see e.g. below for information on source data verification).

According to the self-evaluation of the registry operators, the degree of fulfilment of the criteria serving to standardize the data in the registry was generally high. An exception existed in some registries for classification systems that are not regularly used in clinical practice (e.g., the MedDRA system for coding adverse events / adverse drug reactions) and for exact dates, whereby, depending on the registry, this referred to dates on the patients themselves, on the disease (dates of diagnosis and examinations during the course of the disease) or on drug therapy. Especially in the case of the latter, a lower degree of fulfilment was reported in some cases with regard to information on the initial dose or on dose changes during the course of treatment.

With regard to the criteria for the validity of the generation of the sample, a lower degree of fulfilment was described in some cases, particularly with regard to ensuring completeness and to strategies for avoiding patient selection (exceptions particularly include registries with a legal basis for recording data for all eligible patients). In this context, some registry operators stated that the representativeness of the registry population could only be insufficiently evaluated due to the lack of valid external data sources.

The degree of fulfilment was more often classified as being incomplete with regard to the criteria on the validity of data collection (data completeness and correctness). This was mostly explained by the extensive effort involved. With regard to source data verification, most of the registry operators stated that they would perform it, but that the target value of a 10% proportion of data to be verified, as stated by the EMA for example, could not be achieved at present due to the high financial and organizational effort involved. However, some registry operators noted that such a fixed target value would not make sense. This is because, on the one hand, other methods of ensuring data quality are also used (e.g. coding aids and training, plausibility checks, cross-reference checks [when data on a patient are entered by several people, e.g. when patients have contact with different levels of care]). On the other hand, from the perspective of some

registry operators an adapted source data verification would be more sensible (e.g. by checking a small proportion of data in each centre with an extension of the check in case of inconsistencies). According to the registry operators, external audits (external monitoring) were not routinely conducted in the registries examined.

The evaluation of the degree of fulfilment of the criteria for the validity of the statistical analysis and reports varied between the registry operators. In some cases, no limitations were noted, in others an incomplete degree of fulfilment was found for the reporting of dealing with missing values and for confounder adjustment. In the latter case, particularly the limited recording of potentially relevant confounders was decisive for this evaluation.

Finally, the self-evaluation of registry transparency and scientific independence was unreservedly positive. A review of the publicly accessible information on the registries (websites, results reports and scientific publications) did not produce any contrary evaluation. In contrast, no registry considered was fully linked to the spontaneous reporting systems for suspected adverse drug reactions. Here, too, the extensive effort required was mentioned as a reason.

Factors particularly beneficial or obstructive to the operation of the registries

In the case of the factors emerging from the registry discussions, i.e. factors that are particularly beneficial and obstructive to the operation of registries, a striking consistency between the various registries and registry operators was shown. Deviating evaluations arose in particular if the operation of the registry is supported by a special framework, such as the legal requirement for a registry and the data to be collected.

The particularly beneficial factors included:

- a grown community of clinically active physicians and other medical personnel with intrinsic motivation to support the operation of the registry
- local feedback for the participating centres within the registry software (access to data on centre patients, including the possibility of displaying temporal progressions, possibly also benchmarking by means of comparison with other participating centres)
- on the motivation of the centres: usability of the data for research tasks / participation of the centres in scientific publications, scientific independence from sponsors
- easiness of contact between centres and registry operators, with timely support in the event of problems
- with regard to data quality: extensive training for the participating centres, professional support for the centres in data entry, e.g. by trained documentation staff
- for the establishment of a new registry: an existing technical infrastructure (hardware and software), e.g. due to the established operation of a registry for another (related) disease

The particularly obstructive factors included:

- lack of long-term funding possibilities
- variations in data protection requirements (at different levels, e.g. between federal states, between different universities, between different hospitals even within a federal state)
- lack of standardization of the IT infrastructure in the centres (especially variations in hospital information systems [HIS], partially outdated hardware)
- lack of or difficult access to other data sources (e.g. death registries)
- in some cases, lack of willingness for source data verification in the centres (perceived as control, disruption of operations)

Some of the points mentioned above are discussed in more detail below, as they were often and specifically emphasized in the discussions with the registry operators.

Funding and independence

Most of the registry operators consulted pointed to the need for sustainable funding of the registry. For many registries, this is a permanent issue, as funding often has to be ensured from the operating organization's own resources, with or without financial support from the (mostly pharmaceutical) industry, possibly supplemented by temporary third-party funding of research projects. Industry funding was viewed critically by some registry operators, as in their view this endangers the independence of the registries. Regardless of the specific funding of their own registry, the registry operators agreed that all data collected in the registry must be available without restriction for analysis by the registry operators (or institutions or other research groups commissioned by them) and that it must be possible to publish these analyses independently of the results.

According to the registry operators, the sometimes severely limited resources prevented desirable extensions of the data collection in some registries, and in some cases, a reduction in the current data collection is also being considered. This does not only concern data fields and data time points, but also the professional support of data collection on site. All in all, most of the registry operators believe that sustainable (i.e. secured for several years) basic funding by a body independent of industry and SHI / private health insurance is necessary, commensurate with the importance of registries and the associated effort involved. In this context, the levying of a system surcharge for the funding of registries was mentioned as an option.

Data protection

For all registry operators, data protection was an essential quality feature of the registries to ensure trustworthiness. Accordingly, it was not data protection as such that was described as a major obstructive factor, but rather the local or regional procedures associated with it. In this context, 3 points in particular were mentioned:

- Different requirements and evaluations of data protection by the competent authorities in the federal states ("Bundesländer"), including the appointed data protection officers: identical topics (procedures, declarations of consent etc.) are assessed differently in the different states in some cases
- In part, specific requirements at centre level: even if the competent authorities (e.g. the "Bundesländer") approve the registry in principle, negotiations at centre level can be lengthy because of specific additional requirements that cannot be easily implemented in the operation of the registry
- Lack of practical assistance (documents, process descriptions) in the event that a new registry is set up; the information and documents provided by the platform "Technology, Methods, and Infrastructure for Networked Medical Research" (TMF[15]) were described by most registry operators as too remote from practice.

Lack of standardization of the IT infrastructure in the centres

In the discussions, the lack of standardization of the IT infrastructure, as well as the partially outdated hardware in the participating centres, was repeatedly mentioned as an obstructive factor. It was clearly described why the lack of standardization of the HIS depending on the centre can make data entry on site unnecessarily complex and error-prone. This is because a lot of the information also required for the registry (e.g. demographic data of patients, laboratory values, examination results, etc.) must also be entered into the HIS or is contained in it in a patient-related manner. Two points were especially emphasized in this context:

- On the one hand, depending on the HIS, interfaces are missing that would enable the transfer to the registry of this already digitally available information. This not only causes unnecessary additional work due to double entry (which also hampers the recruitment of new centres), but also represents a potential source of error when manually transferring this information into the registry software. For example, it was described that on-site documentation staff responsible for data input into the registry software work in parallel with 2 IT systems, because the necessary information is read off the screen of one system (HIS) and then transferred to the other system (registry software).
- Furthermore, the lack of semantic interoperability (depending on the data field) between the various HIS does not allow data to be transferred to the registry using a uniform interface, even from HIS with an existing interface. According to some registry operators, this problem is exacerbated by individual adaptations of the HIS in the individual centres.

Linkage to other data sources to use data collected elsewhere

From the perspective of the registry operators, it would be welcome if high-quality data already collected elsewhere could be made available to the registries to complete the patient-related information. Access to the death registry was mentioned as an example: for reasons of data

[15] *Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.*

protection or the lack of a uniform patient-related identification number, according to some registry operators, access to these data is hardly possible. Instead, for registry studies that, for example, aim to draw conclusions on the survival of patients, a great deal of effort is put into obtaining the relevant information through targeted contact with centres. It was also unclear to the registry operators what potential data sources existed at all, and what possibilities and limitations the respective data sources offered. In this respect, a central information platform, which ideally should also provide a low-threshold opportunity for exchange between the users of the data sources, was lacking.

Suitability of the data for the purpose of benefit assessments of drugs

It became clear from the interviews that, on the one hand, the existing registries are generally not primarily designed to generate data for benefit assessments of drugs. On the other hand, due to the existing technical and organizational infrastructure, and also in part due to the already existing data structure, it can be assumed that such data can in principle be generated at least prospectively in the registries considered. Whether or what additional effort this would involve depends strongly on the respective research question, which will be explained in the following text using examples.

- Most of the registries also record PROs, although some only since recently, resulting in a small up-to-date data stock. It depends on the specific research question whether the tools used (symptom questionnaire, questionnaire on generic or disease-specific quality of life) and the recording intervals are suitable for a benefit assessment.
- In comparison to clinical studies, adverse effects are only recorded to a limited extent in the registries considered. In most cases, known adverse effects of selected drugs are the criterion for recording them. In some registries, at least the recorded adverse effects are coded according to the MedDRA system as used in clinical studies, so that in principle, using the same type of recording for registries as for studies seems possible. In some registries, this is supported by the distinction between centres with limited (core data set) and extended data collection. In some registries, centres with extended data collection are in principle geared towards data collection as in studies. Some registry operators are also aiming for this in the context of registry-based RCTs. In the view of some registry operators, the associated legal requirements (especially according to the German Medicines Act [AMG[16]] and SGB V) represent important hurdles for the implementation of interventional registry studies.
- In some of the registries considered, not all levels of care (e.g. primary care, care in centres, rehabilitation) are directly involved in data collection. In these cases, the information flow between the levels of care determines whether information arising from contact with an uninvolved level of care is included in the registry data set. This potentially influences not only the completeness of the data (is this information recorded

[16] *Arzneimittelgesetz*

in the registry?), but also the quality of the data (if necessary, a third party may transfer to the registry information that has already been processed and is available in unstructured form [e.g. diagnosis in the physician's letter] without the possibility of source data verification). In turn, it depends on the specific research question whether this information is relevant for the generation of data for benefit assessments and should therefore be recorded in sufficient quality with corresponding additional effort. On the one hand, depending on the disease, there may be hardly any contact with physicians outside the centres participating in the registry, or the information generated in other levels of care may not be decisive for the benefit assessment (e.g. diagnosis of minor diseases). On the other hand, necessary treatments and complications in a (not directly recorded) emergency could be of essential importance for the benefit assessment.

The above points describe the possibilities and limitations of conducting a (retrospective or prospective) comparative study within the same data source, i.e. the same registry. Furthermore, it became clear from the overall view of the interviews that a comparative study (conducted using different data sources for the drug to be assessed or the comparator therapy) will probably only be suitable for benefit assessments in exceptional cases. In such analyses, potentially a confounder adjustment alone is insufficient to avoid systematic bias, since supposedly identical outcomes are potentially defined differently in the different data sources, and even if the definition is similar, the type of data collection leads to relevant differences. Examples include:

- Adverse events: Are the events systematically recorded during patient contact? Are events recorded when the patient refers to them? Are the events classified according to severity? If so, does the patient or physician classify severity according to a standardized system?
- PROs: Where are the data collected, at home or at the centre? Is data collection conducted in a temporal context with the contact to a doctor? If so, before or after the visit? Is the completion of the questionnaires supported, e.g. by staff or by an IT application?
- Recording of diagnoses, secondary complications, events in general: According to which specifications are diagnoses coded in the registry, and at what level of detail? How detailed are events documented (e.g. silent heart attack)? How is unstructured information (e.g. doctors' letters) transferred into structured data fields?

With regard to the examples mentioned, through various measures efforts are being made in the interviewed registries to ensure consistency between the centres and thus within the registry. Consistency with other registries or with clinical studies is not the focus and cannot be the goal, since the various external data sources and clinical studies are themselves not consistent in this respect. Therefore, when comparing a new drug with a comparator therapy using different data sources (e.g. single-arm study of the new drug vs. control for the comparator therapy from a registry), depending on the outcome of interest, confounder adjustment alone is insufficient to address potential bias. A comparison using different data sources for a targeted data collection seems reasonable only for outcomes where recording does not potentially differ markedly between sources. These include mortality and, where appropriate, standardized laboratory

values or other measurements validated as surrogates (e.g. SVR for hepatitis C). For other outcomes, this would have to be examined and confirmed on a case-by-case basis. Independent of this, comparable availability and operationalization of the relevant confounders in the different data sources would be necessary.

From the above-mentioned explanations it also follows that, for the collection of routine practice data for benefit assessments, those registries in which patient characteristics are the inclusion criterion (e.g. disease registries) are primarily considered. Registries in which the specific therapy (e.g. treatment with a new drug) represents the inclusion criterion are probably only suitable for benefit assessments in exceptional cases. This is particularly true in the case of product or intervention registries in which a large part of existing treatment options (e.g. established standard therapies) are not considered in the registry, as probably in this case comparisons for benefit assessments within this data source will normally not be possible.

Estimation of the necessary timeframe for registry studies

The following estimate of the necessary timeframe for registry studies is based primarily on the specific practical experience of registry operators as outlined in the discussions. The estimation is limited to registry studies that are conducted solely within the respective registry, i.e. without connection to another data source. Possible delays due to lack of funding of the studies are not considered, since it is assumed for the present situation of use that funding will at least be secured by the contracting party of the planned study.

This estimation is made subsequently for the prospective data collection in a new registry to be established and the prospective data collection in an existing (and basically suitable) registry. In addition, the timeframe for purely retrospective data collection in an existing (and basically suitable) registry is presented, even though this will probably not be a frequent case due to the requirements and limitations described in Section 5.3. For the more probable case of use of a combined prospective data collection (for the drug to be assessed) and retrospective analysis (for the comparator therapy) within the same data source, the estimation of the timeframe for retrospective analyses can be helpful independently of this.

1) Prospective data collection in a new registry

- Necessary timeframe to set up a new registry: approx. 1 to 4 years, depending in particular on:
  - existing software (possibly already used by the registry operator for another registry) and other technical and organizational infrastructure for registry operation
  - widespread willingness to participate (e.g. through an already existing registry for another disease in the same specialty, an existing network of specialists [e.g. competence networks, certified centres with established joint specialist meetings or health care goals], groups of physicians otherwise organized [e.g. long-term cooperation with patient representatives])

  - number of levels of care required for data collection, including the associated need for decentralized coordination in the area of data protection
  - existence of an established data set vs. the need to define a large part of the data to be collected
- Additional timeframe for the registry study, see 2), with the necessary preparations being made partly as the registry is being set up

2) Prospective data collection in an existing and basically suitable registry

- Timeframe necessary, e.g. for a registry study with a one-year observation period: approx. 2 to 4 years. This period consists of 6 to 24 months of preparation time, the recruitment and observation period (depending, among other things, on the sample size and recruitment possibilities) as well as a period of 3 to 6 months for the analysis and preparation of the results report.
  - Explanation of the time required for preparation: The time required depends strongly on the specific research question and the associated time required for the preparation of the protocol and analysis plan, plus internal and external approval processes for the registry study. In addition, a time factor that is sometimes large and difficult to estimate in general terms is the potentially necessary extension of the data set, less from a technical point of view, but primarily from an organizational one. It may be necessary to separately recruit centres that are willing to participate in the extended data collection, plus training of the participating centres, plus lead-time to ensure high-quality data collection for collection in the context of the specific registry study; it may also be necessary to extend the declaration of consent with decentralized coordination of data protection.

3) Retrospective analysis in an existing and basically suitable registry

- Timeframe necessary: 6 to 18 months, depending strongly on the specific research question and the associated time required for the protocol and analysis plan as well as registry-internal and external approval processes for the registry study.

5.5.1.3 Data quality criteria for the generation of routine practice data for benefit assessments and measures to ensure data quality

The prerequisite for the suitability of routine practice data for benefit assessments of drugs is, among other things, adequate data quality, which must be ensured by appropriate measures. What measures are suitable according to international standards and are named as quality criteria for registries are described in Section 5.5.1.1.

Ultimately, however, it is neither decisive nor necessary that all the measures mentioned have been fully implemented, but rather that the data relevant to the specific research question are available in such quality that analyses within the framework of a registry study can be

interpreted with sufficient certainty. To ensure this, the following criteria for the data quality of a registry can be distinguished:

1) Mandatory criteria for ensuring data quality
2) Data quality criteria, including
    a) general criteria that are always relevant for registry studies considered in benefit assessments of drugs
    b) general criteria which, depending on the research question, may be relevant for registry studies considered in benefit assessments of drugs
    c) criteria whose degree of fulfilment is to be assessed on a question-related basis

Based on the internationally and nationally defined quality criteria, the consideration of the discussions with the registry operators (see Section 5.5.1.2), as well as knowledge of the information on patient characteristics, outcomes, study protocol, and inclusion criteria relevant for benefit assessments of drugs; these relevant criteria are listed in Table 8 below. Criteria that refer to the preparation of a study protocol and an analysis plan for a registry study, as well as to specific analyses of a registry study (e.g. sensitivity analyses), are not listed here, since these are addressed separately in Section 5.5.2 (“Methodological requirements for the analysis”).

Table 8: Criteria for data quality and for ensuring the quality of routine practice data collection for the benefit assessment of drugs

Category	Quality criteria
Mandatory criteria to ensure data quality	▪ Detailed registry description (aim, registry protocol) ▪ Exact definition / operationalization of exposures, clinical events, outcomes and confounders ▪ Current data plan / coding manual ▪ Training on data collection and recording ▪ Clearly defined inclusion and exclusion criteria for registry patients ▪ SOP system for data collection ▪ Package of measures to ensure the accuracy of data and to provide information on error rates (e.g. source data verification, internal and external audits, IT-supported checks [e.g. cross-reference checks]) ▪ Documentation trail - documentation of process and definition changes in the registry ▪ Scientific independence of the registry ▪ Sustainable financing
General criteria that are regularly relevant for registry studies for benefit assessments	▪ Use of exact dates for patients, disease and events ▪ Detailed information on the drug therapy (active substance, dose, dose change, including dates) ▪ Timeliness (currentness and rapid availability of the required results)
General criteria that may be relevant for registry studies for benefit assessments, depending on the research question	▪ Use of standard classifications (e.g. ICD-10) and terminology (e.g. MedDRA) ▪ Use of valid standard survey tools (questionnaires, scales, tests) ▪ Flexibility and adaptability (e.g. for embedding studies, for further data collection, in the event of changes in the health care situation) ▪ Linkability with other data sources
Criteria whose degree of fulfilment is to be assessed with regard to components of the research questions[a]	▪ Representativeness of the sample / selection of the sample ▪ Completeness of data per data collection time point (lost-to-follow-up, drop-outs) ▪ Completeness of data collection time points ▪ Correctness of data ▪ Collection of data on all confounders relevant for the research question ▪ Data consistency over time
a: The criteria mentioned are important criteria of data quality, but can only be assessed in relation to specific questions. On the one hand, for example, “accuracy of data” and “consistency of data over time” only refer to data that are relevant to the respective question. On the other hand, “representativeness of the sample” refers only to the population relevant to the research question, but not to the entire registry population. ICD: International Statistical Classification of Diseases and Related Health Problems; IT: information technology; MedDRA: Medical Dictionary for Drug Regulatory Affairs Activities; SOP: standard operating procedure	

In the context of the suitability testing of a specific registry, this list of criteria should be used to assess for each specific research question

- whether and to what extent the individual criteria are fulfilled
- what influence a possible non-fulfilment is likely to have on the quality of the results, and

- whether possible deficits can be corrected in a registry-based study using a reasonable amount of resources.

5.5.2 Methodological requirements for the analysis

The methodological requirements for the analysis of comparative studies with randomization and of adjusted indirect comparisons using a common comparator are established standards [15,36]. The present chapter therefore essentially describes the methodological requirements for the analysis of comparative studies without randomization. This is based on the published scientific literature as well as on discussions with statisticians with in-depth expertise of analyses of patient registry data (see also Section 4.2).

In addition, at the end of the section, special requirements for the analysis of comparative studies with randomization as well as the merging of data from different data sources are briefly discussed.

Special requirements for the analysis of comparative studies without randomization

Statistical analysis plan

The planning of comparative studies without randomization for the purpose of comparing medical interventions should meet the requirements of comparative studies with randomization [111] (see also Section 5.3.2). This also includes a detailed analysis plan, which is defined in advance, and and which should include

- which statistical methods and models are used
- which methods and criteria are used for model selection and adaptation
- to what extent and for what reasons missing data can be expected
- which measures are taken to avoid missing data
- which analysis strategies are chosen to handle missing data
- how implausible data and outliers are dealt with, and
- which sensitivity analyses are used to check the robustness of the results.

Confounder adjustment

In studies without randomization, the sufficient similarity in terms of prognostic factors of the groups to be compared, which is necessary for a fair comparison, is usually not given. Group differences in possible confounders, i.e. factors that are related to both treatment and outcomes and can therefore distort a treatment effect, must therefore be taken into account when estimating effects. A key aspect in comparative studies without randomization is therefore the adequate adjustment for confounders in order to obtain interpretable estimates of the effect of interest – despite the risk of bias through confounding. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has published good research practices

for this purpose that should be followed [112-114]. In order to achieve adequate confounder control, regardless of the methods used, it is particularly necessary to

- identify in advance all important confounders (including important interactions) and consider them in the model in an appropriate form
- completely collect data on these important confounders in the study
- plan the study with a sufficient sample size to be able to consider all confounders in the model
- describe the causal model exactly, e.g. by means of causal graphics
- present the assumptions of the causal model, and
- substantiate, e.g. on the basis of scientific literature, why these assumptions can be justified in the specific case of use.

If one or more of these important confounders are not contained in the data set, analysis results based on this data set are not suitable for a benefit assessment. In addition, when creating the model for data analysis, all other important characteristics of the existing evidence must also be adequately considered, such as an existing hierarchical structure of the data (e.g. region, clinic, patient).

Methodical approaches to confounder adjustment

In the statistical workshop, it was elaborated that for confounder adjustment for benefit assessments of drugs, only those procedures conducted using individual patient data (IPD) are generally meaningful. Various approaches are available for confounder adjustment using IPD. In general, especially the following 3 approaches can be distinguished [115-117]:

- The simplest method is direct adjustment for confounders in a multiple regression analysis. In principle, this type of adjustment involves the formation of strata that are similar in terms of the confounders (e.g. age, sex and severity of the disease). For each of these inherently homogeneous strata, the treatment effect is estimated separately and these estimates are then combined into a common estimate. This procedure has been established for a long time. However, it has the disadvantage that these analyses can only include a limited number of confounders, depending on the sample size or number of events, as otherwise they provide unreliable results or do not function mathematically at all. Furthermore, adjustments can only be made according to the confounders observed.
- A widespread class of methods that has also been established is based on propensity scores. In a first step, the probability of each person receiving 1 of the 2 treatments to be compared (e.g. the drug to be assessed) is calculated depending on the confounders observed. In this first step, the influence of a larger number of potential confounders per patient is condensed into a single score between 0 and 1, the propensity score. This propensity score is used for adjustment in a second step. There are various procedures for

this, e.g. (direct) adjustment in a regression analysis (see above), the formation of pairs of persons whose propensity scores are very similar (so-called matching), or weighting using the propensity score. The great advantage of procedures based on propensity scores is that, compared to direct adjustment, considerably more possible confounders can be included. In addition, propensity scores can be used to identify and, if necessary, exclude patients who are not theoretically eligible for both treatment options (see also below: "Propensity scores as a method for considering confounders"). However, they can also only adjust according to the confounders observed.

- Another possibility of confounder control is the use of instrument variables. In medical care, a specific treatment is usually chosen depending on prognostic factors (e.g. severity of the disease). In studies without randomization, this leads to imbalances between the groups to be compared with regard to these factors. The idea behind the method of instrument variables is that it describes the selection of a treatment, but is not associated with prognostic factors. For example, this could be the case for certain types of medical training (type of surgery) where the application of the particular surgical method depends only on the centre where the patient is (randomly) treated, but not on the severity of the disease. An instrument variable must therefore be highly correlated with the probability of receiving a particular treatment, but must neither directly nor by association with confounders influence the outcomes. In the above example, it would then be possible to compare the different surgical methods by adjustment with the appropriate instrument variable, assuming sufficient similarity of the treatment groups in terms of prognostic factors. The (theoretical) advantage of this method is that a suitable instrument variable can be used to adjust for both observed and unobserved confounders. In practice, however, the strict assumptions associated with the selection of instrument variables can often hardly be verified.

In the statistical workshop, it was elaborated that the use of instrument variables is of minor relevance for benefit assessments of drugs, among other things because of the associated hardly verifiable assumptions. In contrast, propensity scores are a frequently used method for considering confounders in comparative studies without randomization based on registries. This method will thus be described in more detail below.

Propensity scores as a method for considering confounders

As described above, when using propensity scores, a score aiming to describe the influence of a larger number of confounders is formed for each patient. There are several points to consider when selecting the model to determine the propensity score, estimating the treatment effects using the propensity score, and interpreting the data. Essential points for a general understanding of the use of propensity scores for benefit assessments of drugs are described in detail below.

Positivity, overlap and balance

When using propensity scores, 3 terms are essential: positivity, overlap and balance.

A prerequisite for a comparison of 2 patient groups using propensity scores is first of all, that the patients of both groups are theoretically eligible for both treatments of interest (new drug or comparator therapy). This means, on the one hand, that the patient groups are specified by the research question of the benefit assessment and, for example, patients with a contraindication to one of the treatments investigated must not be included in the analysis (positivity). On the other hand, this means that in the available database there must be sufficient overlap (measured by the propensity score) between the population that has received one treatment (e.g. new drug) and the population that has received the other treatment (comparator therapy). Sufficient overlap means that the distribution of patients among the different propensity scores must be similar. For example, if most patients receiving the new drug have a propensity score below 0.3, but those receiving the comparator therapy have a score above 0.7, then there is insufficient overlap between the populations.

In addition to sufficient overlap, it is necessary that the populations in the treatment groups are sufficiently balanced. This means that the treatment groups do not differ relevantly with regard to important confounders (e.g. age, severity of disease). This is not guaranteed alone by the strong overlap in the populations measured by the propensity score, since the overlap is determined by considering the propensity score as an overall measure, but not by similarity of the individual confounders in detail. Thus, in one group of patients there could be considerably more elderly patients, while in the other group, disease severity is on average considerably higher, without this leading to marked differences in the propensity score.

The degree of overlap and balance between the groups depends first of all on the model chosen to form the propensity score. However, it can also be influenced by "trimming" (excluding patients in non-overlapping areas of the propensity score) and the adjustment methods. The sufficiently overlapping and sufficiently balanced patient population is ultimately the population for whom the estimated effects apply using the propensity score. Therefore, this population should be described in detail and it should be investigated whether it sufficiently depicts the population selected for the original research question. If this is not the case, the estimated effects may only apply to a limited population. On the one hand, however, the postulated advantage of the collection of routine practice data (no relevant restriction of the population investigated) is then potentially lost. On the other hand, the effects may not be interpretable in a meaningful way, e.g. if the artificial population resulting from the overlap cannot be delimited in the German health care context. In a specific situation of use, the necessary confounder adjustment can therefore lead to a considerable limitation of the applicability and interpretability of the results.

Different methods of effect estimation using propensity scores

There are different methods for estimating the treatment effect using propensity scores [118]. When using propensity scores, the existing guidance on data analysis with this method should be followed [118-120]. Which method is the most suitable for a particular case of use can sometimes only be decided on the basis of the specific data, since different methods can lead to different degrees of overlap or balance [121]. However, the analysis plan can and should

describe the decision-making structure for method selection. This includes, for example, the necessary minimum degree of overlap and balance. In addition, sensitivity analyses should be conducted with different propensity score methods, provided that these also fulfil the necessary minimum degree of overlap and balance.

Propensity scores do not replace the need to measure all relevant confounders

Unmeasured confounders can also play a role in the use of propensity scores, so that this method also requires the definition and recording of important confounders and their inclusion in the analysis [122]. As a way out of this situation, the high-dimensional propensity score (hdPS) was suggested, which considers a very large number of confounders via an automatic search. This aims to ensure that important unmeasured confounders are considered via measured confounders (proxy variables) [123,124]. However, the automatic search contradicts the principle of systematically identifying important confounders on the basis of existing scientific literature and expert knowledge, and it is hardly possible to present the method applied in a transparent manner [125]. Irrespective of this, even when using the hdPS, the risk of unmeasured confounders and thus of potential bias in the results remains [125]. This is because the hdPS method can only consider the variables recorded in the data set as possible proxy variables, and it is unclear for the respective situation of use whether all known or unknown, unmeasured confounders can be considered via these variables. Therefore, the hdPS method cannot replace the above-described requirement that the relevant confounders be completely collected and considered in the model.

Qualitative certainty of results and replication of results

In practice, even if the above guidelines are strictly adhered to, the accuracy of the assumptions regarding confounder adjustment cannot be fully verified and unmeasured (or completely unknown) confounders may play a role. In comparative studies without randomization, the validity of the results therefore always depends on the assumptions made [126]. Therefore, there is a special obligation to justify these assumptions, supported by scientific literature.

Overall, according to the Institute's General Methods, the results of comparative studies without randomization as a rule provide at most a low qualitative certainty of results, which, in a meta-analysis of at least 2 such studies with a statistically significant result, leads at most to a hint of the existence of an effect [36]. Replication of the results can also be aimed for within the same data set, e.g. by analysis according to regions or within the same data source by a second independent sample.

Even if studies without randomization only show a low qualitative certainty of results, they can increase the certainty of results of the overall conclusion on added benefit when combined with other data. This must be assessed on a case-by-case basis. Examples are reliable data on important outcomes, which are supplemented by the study without randomization in other outcome categories, or the joint consideration of a (small) study with randomization with a (larger) study without randomization.

Quantification / magnitude of the effect

Even with the most careful analysis and fulfilment of the quality requirements mentioned above, due to potentially unknown confounders, a conclusion on the benefit or harm of an intervention should only be derived from the effects observed in the study if these effects exceed a certain effect size. A (positive or negative) conclusion on the benefit or harm can be drawn if the confidence interval for the effect observed exceeds a threshold that must be defined. Since the fulfilment of the above-mentioned quality requirements is a prerequisite for the observation of effects, this threshold value should be significantly below the value for the "dramatic effect" (relative risk of 5–10 [36]), e.g. in a range of 2–5 for the relative risk. The specific threshold depends on the quality of the data in the individual case, including knowledge of relevant confounders. Depending on the data, such a threshold can also be applied specifically to outcomes, e.g. due to the lack of blinding of treatments or a different direction of bias for positive or negative effects.

In benefit assessments of drugs according to §35a SGB V, the extent of the added benefit of an intervention must be quantified [17]. Different effect sizes are required for the classification into the extent categories "minor", "considerable" and "major". For this classification, the method currently used in IQWiG's dossier assessments cannot be adopted without change, as it is based on data with a higher certainty of results without the need for a threshold value for conclusions on benefit or harm [127].

Starting from the above-mentioned threshold for a conclusion on benefit or harm, if the threshold is exceeded, there is at least a minor added benefit for the respective outcome. Exceptions are outcomes in the category "non-serious/non-serious complications", because, according to the Regulation for Early Benefit Assessment of New Pharmaceuticals[17], a "not only marginal improvement" is additionally required for these outcomes [17]. For all outcome categories, classification into the extent categories "considerable" or "major" requires higher (i.e. above the above-mentioned threshold) effect sizes that are graded according to magnitude.

Special requirements for the analysis of comparative studies with randomization

The requirements for the analysis of comparative studies with randomization are described in numerous guidelines of the drug regulatory authorities and are not the focus of this report. Examples are the guidelines of the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) [11], in particular the guideline "Statistical Principles for Clinical Trials" [128], as well as the statistics guidelines of EMA [129].

It should be noted that in comparison to studies without randomization, comprehensive confounder adjustment is not required, since confounder control is ensured by randomization.

[17] *Arzneimittel-Nutzenbewertungsverordnung*

Likewise it is no longer required to increase the sample size of the study to enable comprehensive confounder adjustment in a model.

Merging analyses from several data sources

For rare diseases in particular, it may be useful and necessary to conduct studies in international cooperation. In this case, the following principles in particular must be observed:

- If several data sources are used for the analyses (e.g. registries from the participating countries), data must be harmonized at the level of content and structure. This requires a standardized procedure, e.g. according to Fortier 2016 (Maelstrom Research Guidelines [130]). The procedure described there for retrospective data harmonization can and should in principle also be applied to prospective studies before the start of the study.
- In the case of analyses from several registries, it may make sense not to form a common data pool for the joint project and then analyse it, but to plan and conduct identically planned studies in the individual registries and then to summarize these studies meta-analytically. On the one hand, this reduces data protection requirements, since the IPD only need to be available to those who are authorized to access it anyway. On the other hand, any differences in effects between the country-specific studies can be revealed and addressed in the meta-analysis. These advantages must be weighed against possible disadvantages compared to the statistical analysis of a single data pool [131].
- For analyses that use data generated outside the German health care context of interest, it must be justified that the data are routine practice data similar to those in Germany, or that deviations from these data are not relevant for the effect estimate. In this context, attention must be paid to any differences in health care pathways, concomitant therapies, resistance levels (in the case of antibiotic therapy), etc. This justification can be supported by means of heterogeneity tests in the meta-analytical summary of the studies from different countries (see previous point).

5.5.3 Reporting requirements

Adequate reporting on the methods and results of studies is an integral part of high-quality research [12]. The quality of the documentation of a study essentially determines the informative value of the results obtained. The complete documentation includes the study protocol (planning of the methods and conduct of the study), the analysis plan (planning of the data analysis), and the results report (description of the planned methods [including the analysis] and conduct of the study, deviations from this planning and reporting of complete results). In addition, there is a discussion, also in the regulatory area, as to what extent the controlled availability of anonymized IPD should be part of study documentation [132-134].

The study protocol and the analysis plan serve not only to describe the methods and conduct of the study, but also to prespecify the study planning. This means that procedures in the study are determined before the data are collected and thus without knowledge of the data. This prevents

influence on the results obtained (consciously or unconsciously) by choosing methods while knowing the data (see also Section 5.3.2).

While in the previous sections the requirements were characterized by whether a comparative study should be conducted with or without randomization, with regard to the possibilities of prespecification of the study planning as an essential characteristic of meaningful study results, it is primarily important whether a comparative study with prospective or retrospective data collection is available.

Prospective comparative studies

The prespecification of the methods of prospective comparative studies with or without randomization can be easily achieved and verified. Here, it is only necessary to finalize the study protocol and the analysis plan in a manner comprehensibly documented before the start of data collection. Documentation by means of study protocol amendments is an established standard for documenting controlled changes in the course of the study.

For comprehensible documentation, it is recommended to publish the study protocol and the analysis plan in a study registry before the start of data collection. With the ClinicalTrials.gov registry, a suitable database is available that allows registration with storage of study protocols and analysis plans for all study types [21].

Retrospective comparative studies (including studies with retrospective comparator arms)

It is more difficult to ensure transparency of study planning and analytical methods before the conduct of retrospective studies. Since the data are already available, it cannot be conclusively ensured that the study was planned without knowledge of the data.

In this context, the FDA also points out problems, especially in retrospective observational studies. For example, in its paper on “real-world evidence” [13], the FDA notes: “The potential lack of up-front transparency, especially in retrospective observational study design and conduct, coupled with the fact that retrospective analyses in electronic datasets can be conducted multiple times relatively inexpensively with varying study design elements, makes it possible to conduct numerous retrospective studies until the desired result is obtained and then submit only favorable results as if they were the result of a single study with a prespecified protocol.” Currently, the FDA has no solution to this problem either, and reporting requirements for such studies are still being developed [13].

Regardless of this fundamental problem, it is recommended that a study protocol and an analysis plan for retrospective comparative studies should also be published in a study registry. The registration and publication of the study protocol and analysis plan in ClinicalTrials.gov is also possible for this type of study.

Special requirements for the reporting of comparative studies without randomization

Methodologically, the planning of comparative studies without randomization to determine treatment effects should explicitly replicate the planning of a comparative study with randomization (see Section 5.3.2). This emulation of a target trial should be explicitly described in the study protocol. In addition, the study report should describe the extent to which this replication has been successful (e.g. with regard to the depiction of the inclusion/exclusion criteria).

As described in Sections 5.3.2 and 5.3.4, comparative studies without randomization require, in particular, approximation of the similarity of the treatment groups to be compared in terms of prognostic factors by adjusting for relevant confounders. For reporting purposes, this means that the relevant confounders must be identified in the study protocol and that the conduct of confounder control must be pre-specified in the analysis plan as far as possible. The results report of the study must contain information on the availability of data on relevant confounders. The results report must also include a transparent presentation of the analysis procedure and the unadjusted and adjusted comparisons of the treatment groups [135-137].

5.6 Assessment of the concepts for generating routine practice data and their analysis for benefit assessments according to §35a SGB V

Possible study designs for generation routine practice data for benefit assessments

The present report examines the generation of routine practice data for benefit assessments. From this objective, it follows directly that data must be collected that enable a comparison between patient groups treated with different interventions (see Section 5.2).

The generation of comparative routine practice data is not bound to a specific study design, but can be conducted on the basis of different study designs. In particular, routine practice data can be collected in comparative studies both without and with randomization [9,27,138-140]

Type and scope of data collection

The aim of collecting routine practice data for benefit assessments of drugs is to be able to draw sufficiently reliable conclusions on the benefit and harm of the drugs to be assessed, both under conditions of routine practice and under consideration of the research question of the benefit assessment. This aim does not mean that data collection has to be limited to those data collected in routine practice without the purpose of generating information for a benefit assessment. Rather, such an incorrectly understood limitation of data collection would jeopardize the aim of the benefit assessment. For the benefit assessment, data are often required that are not documented for all patients in routine practice (e.g. data on health-related quality of life, on symptoms or adverse effects, see also Sections 5.4.3 and 5.4.4). An analysis of the evidence gaps named by the G-BA in orphan drug assessments underlines the need to collect data on all outcome categories of the benefit assessment (see Section 5.7.2).

The extent of the supplementary data collection depends, among other things, on the study design chosen. As explained in Sections 5.3.2 and 5.5.2, supplementary data collection may especially be necessary for comparative studies without randomization, since, in addition to data on outcomes, information that allows for confounder control must also be collected.

The collection of the necessary data can be integrated into the daily treatment routine. The assessment of possible data collection tools has shown that registries, possibly supplemented by a study-specific collection of the data not available in the registry, currently and for the near future represent the most realistic possibility (see Section 5.4). Registries offer the most likely option for adaptation to the data collection required. This concerns both the specification of the data required and the data quality.

Analysis of data collected in routine practice

The requirements for the analysis of routine practice data are largely determined by the study design used. The primary challenge is to achieve sufficient similarity between the treatment groups in terms of prognostic factors by adjusting for confounders when comparative studies without randomization are chosen.

The need for confounder adjustment in the analysis leads to a number of requirements for the data set to be collected. These include, for example, the pre-specification and recording of the relevant confounders and a sufficient study size to adequately perform the adjustment. In particular, the specification of the sufficient study size for confounder adjustment is currently missing in the discussion about comparative studies without randomization, although particularly this requirement is important for studies in small populations for which this study type is frequently proposed. For example, adequate confounder control cannot usually be performed in studies with a sample size of less than 100 patients. Efficient study designs are particularly necessary for studies in small populations.

At present, the planning of comparative studies without randomization and the associated data collection is often insufficient for high-quality analyses [141]. The quality deficiencies are due, among other things, to the fact that the selected data sources do not contain the necessary information, e.g. on confounders or outcomes [142]. Section 6.2 therefore contains recommendations for action for registry operators and persons responsible for registry studies, which serve as preparation for future data collections and analyses for benefit assessments of drugs.

Informative value of data collected in routine practice

According to the standards of evidence-based medicine, the informative value of the data collected is based on a combination of qualitative and quantitative certainty of results, also in the case of the collection routine practice data.

With regard to the qualitative certainty of results, there is a scientific consensus that a comparative study without randomization (with at most moderate effects) cannot achieve the informative value of a comparative study with randomization. For this reason, the collection of

routine practice data according to GSAV by means of studies without randomization is not meaningful if, on the basis of the available evidence, it can be expected that there are no relevant differences between the new drug and the comparator therapy.

Rather, the question arises as to what extent the inherent uncertainty of comparative studies without randomization can be reduced by measures approximating the similarity of the treatment groups in terms of prognostic factors.

If, in certain cases, the inherently increased uncertainty of comparative studies without randomization is to be accepted in the benefit assessment, the following key points arise from the assessment of the concepts for generating routine practice data and their analysis for benefit assessments:

- No effects can be derived from comparative studies without randomization if the data quality in the data sources used and the quality of analysis and reporting is not high (exceptions are, under certain circumstances, effect sizes that are so large that they can no longer be plausibly explained by confounders alone).
- Even under high quality requirements (for data, analysis and reporting), no more than a hint of an effect can normally be derived from comparative studies without randomization.
- Due to the inherent uncertainty of the results from comparative studies without randomization through remaining unknown confounders, even under high quality requirements, a conclusion on the benefit or harm of an intervention can only be derived from the effects observed if a certain effect size is exceeded. Quantification of an added benefit according to the legally prescribed extent categories requires corresponding effect sizes graded according to magnitude.

5.7 Suggestions for a procedure for collection of routine practice data according to GSAV

The preceding chapters of this report describe concepts for the collection of routine practice data for benefit assessments according to §35a SGB V. In the following text, the results of the project will be discussed in connection with the option of collecting routine practice data, which was introduced by the GSAV.

5.7.1 Definition of the research questions for the collection of routine practice data according to GSAV

The basis for the collection of routine practice data according to GSAV is the definition of the research question to be answered by this data collection. This question contains at least the components of the PICO scheme and the necessary duration of the data collection. The exact specifics of the research question are derived from the evidence gap arising in the benefit assessment that is to be closed by the data collection. A research question defined in this way is also the starting point for the description of the necessary scope of the data collection

(including duration of observation and sample size estimation). Table 9 describes the components of the research question.

Table 9: Definition of the research question for the collection of routine practice data according to GSAV for benefit assessments according to §35a SGB V

Component of the research question	**Characteristics of the components depending on the evidence gap identified in the benefit assessment**
Patient population (P)	▪ patient population according to the approved therapeutic indication or ▪ limited patient population for which the benefit assessment identified an evidence gap
New drug (I)	▪ drug to be assessed
Appropriate comparator therapy (C)	▪ appropriate comparator therapy
Patient-relevant outcomes (O)	▪ patient-relevant outcomes for mortality, morbidity and health-related quality of life or ▪ specific patient-relevant outcomes that can close the evidence gap identified in the benefit assessment
Duration of data collection	▪ duration of data collection depending on the treatment situation (e.g. long-term vs. acute therapy) for which an evidence gap is to be closed
SGB: Sozialgesetzbuch (Social Code Book)	

Examples for evidence gaps include:

- **Patient population:** there is a particular need for the drug in patients with a severe form of the disease covered by the therapeutic indication, but only data on patients with a mild form of the disease were available for the benefit assessment; data for the patient population with severe disease are missing.
- **New drug:** in the studies available for the benefit assessment, a new drug was used in a different form of administration from the authorized use, thus lacking results on the comparison of the authorized form of administration with the appropriate comparator therapy.
- **Appropriate comparator therapy:** no data of informative value were available for the benefit assessment with regard to the comparison of the new drug with the appropriate comparator therapy.
- **Patient-relevant outcomes**: there is a lack of data on health-related quality of life, although the existing treatment situation (e.g. palliative care) or the results of the benefit assessment (e.g. survival benefit, but at the same time an increased incidence of serious adverse events) underline that results on health-related quality of life are relevant for treatment decisions.

- **Duration of data collection:** for a drug for a chronic disease that has to be taken longer-term or permanently, only data for a very limited treatment period are available for the benefit assessment; benefit and harm in long-term treatment remain unclear.

5.7.2 Evidence gaps in previous benefit assessments using orphan drugs as an example

Collections of routine practice data according to GSAV should be conducted for benefit assessments and should be aimed at quantifying the added benefit of an intervention [1]. A targeted collection of routine practice data according to GSAV must therefore address those evidence gaps that are important for the benefit assessment and the connected determination of the extent of added benefit.

In order to be able to better assess which evidence gaps exist in benefit assessments and are relevant for the determination of the extent of added benefit, G-BA decisions on benefit assessments of orphan drugs were examined in more detail. Decisions on orphan drugs with market access in the years 2014 to 2018 were examined, including decisions on new therapeutic indications for orphan drugs in this period. Re-assessments of the same therapeutic indications were not considered, e.g. after the expiry of a possible time limit of a decision on the drug or after exceeding an annual revenue threshold of €50 million euros.

In the following section, the G-BA decisions are first analysed in summary. The aim is to identify categories of evidence gaps and to describe any connections between evidence gaps and the data available at the time of market access and the prevalence of the disease.

Summary analysis of decisions on orphan drugs with market access 2014 to 2018

A total of 67 decisions on orphan drugs were relevant for the summary analysis during the period in question. In these 67 decisions, conclusions on the extent of added benefit were made for 85 different research questions. A list of the decisions and the related research questions can be found in Appendix D.

According to §35a (1) Sentence 11 SGB V, the added benefit for orphan drugs is deemed to be proven at market access. Accordingly, the categories “major”, “considerable” or “minor” (quantified added benefit) or “non-quantifiable” listed in the Regulation for Early Benefit Assessment of New Pharmaceuticals[18] can be used for determining the extent of the added benefit for orphan drugs [17]. Figure 5 shows the division of the research questions into those in which the added benefit was quantified and those in which the added benefit was classified as “non-quantifiable”. In each case, it was added how often different study designs were the basis for this classification, according to the decision by the G-BA.

[18] *Arzneimittelnutzenbewertungsverordnung*

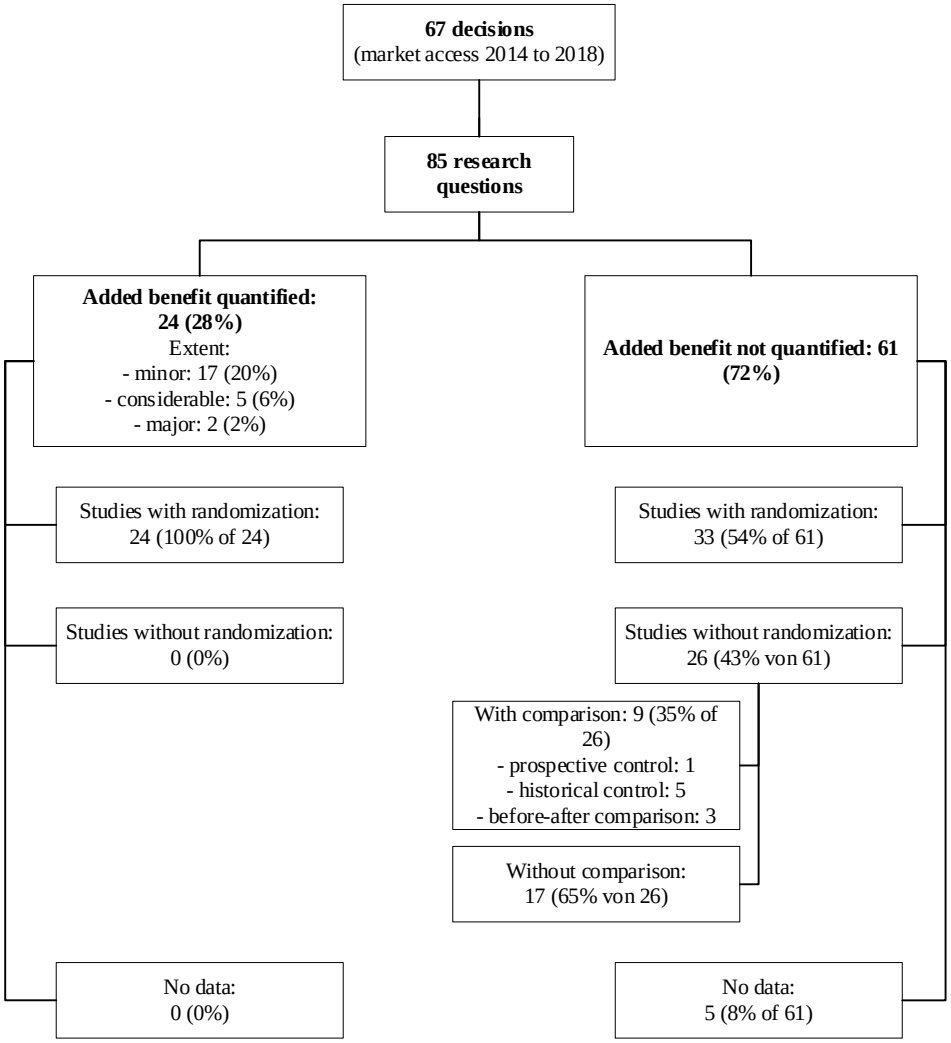


Figure 5: Quantification of the added benefit and available data – G-BA decisions on orphan drugs with market access 2014 to 2018

The added benefit was quantified only in about a quarter of the research questions (24 of 85 questions, 28%). In these cases, only studies with randomization provided the data basis. In none of the questions was the added benefit quantified on the basis of comparative studies without randomization.

The added benefit was not quantified in about three-quarters of the research questions considered (61 of 85 questions, 72%). Data were available for most of these questions (92%), the majority of which were studies with randomization (33 of 61 questions, 54%). For 26 of these 61 questions (43%), data from studies without randomization were available, in 3 cases in addition to a study with randomization. In many cases, these were data on orphan drugs alone (single-arm studies) and thus not comparative data. In some cases, data were also available for a control group, mostly as a historical control. For 5 of the 61 questions with non-quantifiable added benefit (8%), no data were available at all.

Table 10 below contrasts the size of the target population according to the G-BA decision and the data (study types) on which the G-BA decision is based, separately for research questions with quantified added benefit and those with non-quantifiable added benefit.

Table 10: Size of the target population and data basis for the benefit assessment – G-BA decisions on orphan drugs with market access 2014 to 2018

Study type	**Size of target population[a]**		
	Research questions with quantified added benefit N Median [IQR]	**Research questions with non-quantifiable added benefit N Median [IQR]**	**Any added benefit N Median [IQR]**
Studies with randomization	24 research questions 428 [144; 3255]	33 research questions 815 [350; 2300]	57 research questions 433 [255; 2450]
Studies without randomization	0 research questions	26 research questions 165 [65; 500]	26 research questions 165 [65; 500]
No data	0 research questions	5 research questions 250 [26; 455]	5 research questions 250 [26; 455]
Total (any study type)	24 research questions 428 [144; 3255]	61 research questions 433 [120; 865]	85 research questions 433 [120; 1400]

a: Information according to the G-BA decision; if a range was specified in the decision, the respective mean value was used; if the information was valid for several research questions, it was divided equally between the questions.
G-BA: Gemeinsamer Bundesausschuss (Federal Joint Committee); IQR: interquartile range.

From this comparison it can be derived on the one hand that studies without randomization were more often presented for research questions with a small target population, but that on the other hand, a small target population was not per se the cause of a lack of quantifiability of the added benefit.

Characterization of the evidence gaps for research questions with non-quantifiable added benefit

For the characterization of evidence gaps, the 61 research questions for which the G-BA identified a non-quantifiable added benefit are examined in more detail below. The background is that in these cases the scientific data basis was insufficient to quantify the added benefit [17]. According to §130b (3) of SGB V, the collection of routine practice data according to GSAV should be aimed at quantifying the added benefit, because if the added benefit cannot be quantified on the basis of the data obtained, a lower reimbursement price for an orphan drug must be agreed upon [1].

The characterization of the evidence gaps was conducted on the basis of a text analysis of the supporting reasons for the respective decision of the G-BA, by extracting and categorizing the reasons given primarily for the lack of quantifiability of the added benefit. The following aspects were considered in this context:

- Does the evidence gap described by the G-BA refer both to the drug to be assessed (orphan drug) and the control group (e.g. in the case of evidence gaps in studies with randomization) or only to the control group (e.g. if data on the control group were completely missing)? What was the extent of the evidence gap (data completely missing, data partially missing)?
- In which outcome category does the evidence gap exist (mortality, morbidity, health-related quality of life, adverse events)?
- What reason is given for the evidence gap (data were not available [e.g. because they were not collected at all], data were collected in insufficient quality or quantity, data analysis was inadequate)?
- For research questions where no relevant data gaps were named in the supporting reasons: What other reasons were there for the determination of a non-quantifiable added benefit?

Figure 6 shows how often evidence gaps were identified for the new drug (orphan drug) or the control group, and the extent of each of these gaps.

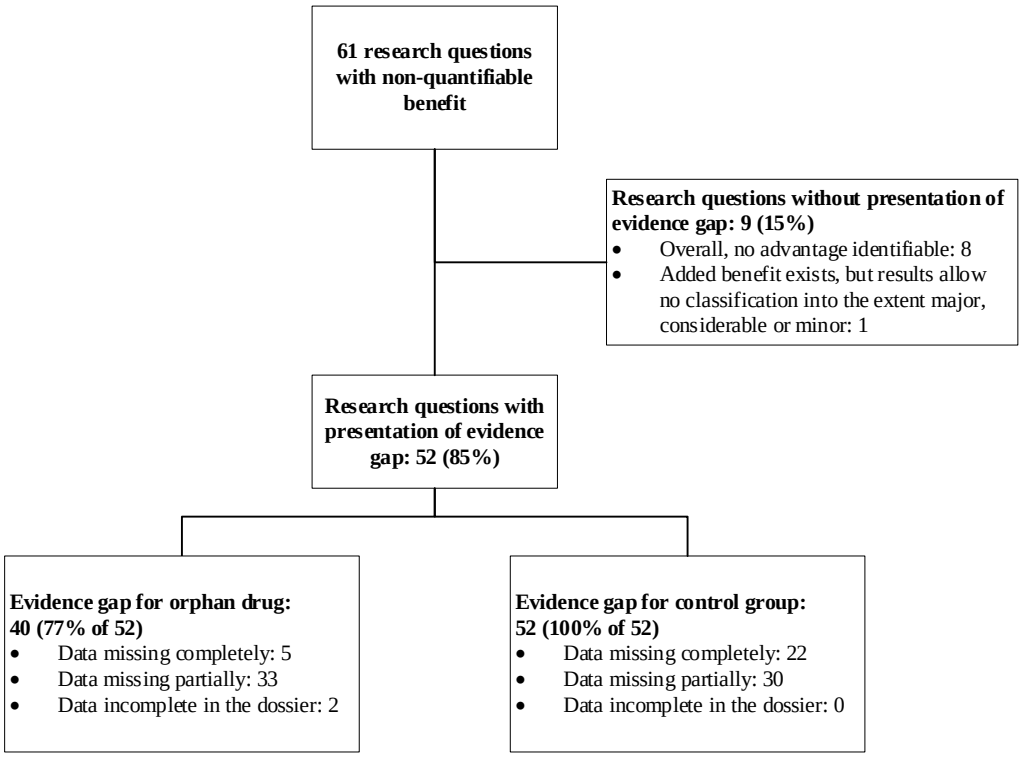


Figure 6: Overview of evidence gaps in questions with non-quantifiable added benefit – G-BA decisions on orphan drugs with market access 2014 to 2018

The analysis shows that the G-BA did not quantify the added benefit without presenting an evidence gap in only a few cases (9 of 61 questions, 15%). In 8 of these 9 questions, no advantage for the orphan drug could be identified in the overall weighing of positive and negative effects compared to the control group. In these cases, the determination of a “non-

quantifiable added benefit" obviously followed the legal specification of the existence of an added benefit for orphan drugs at market access (§35a [1] Sentence 11 SGB V). In one case, the G-BA considered the data to be sufficient for the determination of an added benefit, but did not allow for a classification into one of the categories "minor", "considerable" or "major", without specifically presenting gaps in the evidence.

For the vast majority of the research questions without quantification of the added benefit (52 of 61 questions, 85%) there were evidence gaps. The extent of the evidence gaps was greater for the control groups than for the orphan drugs, which was mainly due to a complete lack of data for the control group. Gaps in evidence were also found for orphan drugs in the majority of cases. In summary, the following picture emerges:

- In all 52 research questions with evidence gaps (100%) data on the control group were missing, for the respective orphan drug this was the case in 40 questions (77%).
- The extent of the evidence gaps was also higher for the control group than for the orphan drug group: in 22 of the 52 cases, data on the control group were missing completely (42%), whereas this was the case for orphan drugs in only 5 of the 40 questions with evidence gaps (13%).

The following Table 11 shows for which outcome categories evidence gaps existed according to the G-BA and for which reasons.

Table 11: Presentation of the evidence gaps per outcome category – orphan drugs with market access 2014 to 2018, research questions with non-quantifiable added benefit

Group **Explanation for evidence gap**	**Mortality** **n (% of N)**	**Morbidity** **n (% of N)**	**Quality of life** **n (% of N)**	**Adverse effects** **n (% of N)**
Orphan drugs (N = 40)	20 (50%)	27 (68%)	26 (65%)	18 (45%)
No data available	*7 (18%)*	*8 (20%)*	*16 (40%)*	*7 (18%)*
Deficits in data collection[a]	*11 (28%)*	*12 (30%)*	*5 (13%)*	*5 (13%)*
Deficits in data analysis	*2 (5%)*	*7 (18%)*	*5 (13%)*	*6 (15%)*
Control groups (N = 52)	33 (63%)	40 (77%)	35 (67%)	34 (65%)
No data available	*22 (42%)*	*27 (52%)*	*30 (58%)*	*27 (52%)*
Deficits in data collection[a]	*11 (21%)*	*8 (15%)*	*2 (4%)*	*4 (8%)*
Deficits in data analysis	*0 (0%)*	*5 (10%)*	*3 (6%)*	*3 (6%)*

a: Including too short observation period or too small sample size.

Table 12 below shows the number of research questions for which evidence gaps existed in 1, 2, 3 or all 4 outcome categories.

Table 12: Number of outcome categories with an evidence gap – orphan drugs with market access 2014 to 2018, research questions with non-quantifiable added benefit

Number of outcome categories with evidence gaps	**Orphan drugs** n (% of 40)	**Control groups** n (% of 52)
1 outcome category	12 (30%)	12 (23%)
2 outcome categories	12 (30%)	12 (23%)
3 outcome categories	9 (23%)	6 (12%)
4 outcome categories	7 (18%)	22 (42%)

The detailed presentation of the evidence gaps leads in particular to the following conclusions:

- Evidence gaps were frequent and almost equally distributed in the 4 outcome categories "mortality", "morbidity", "health-related quality of life" and "adverse effects". This applies equally to the orphan drug groups and to the respective control groups.
- The evidence gaps were greater for the control groups than for the orphan drug groups across all outcome categories.
- While for the control groups, the complete lack of data was by far the most common cause of the evidence gap, for the orphan drug groups, deficiencies in data collection in the studies presented in the dossier were also a common reason.
- For the orphan drug and especially for the control groups, the evidence gaps were only in a few cases caused by deficiencies in the analysis of the data collected and presented in the dossier.
- For most research questions, evidence gaps existed in 2 or more of the 4 outcome categories.

Conclusion of the summary analysis of orphan drug assessments and existing evidence gaps

On the one hand, the summary analysis of orphan drug assessments for the years 2014 to 2018 shows that relevant data for a large part of the research questions were submitted for a benefit assessment in the corresponding dossier at the time of market access (80 out of 85 questions, 94%). In about two thirds of the cases, these were studies with randomization and in one third of the cases, studies without randomization. On the other hand, the analysis also shows that, despite this high rate of research questions with relevant data, the added benefit could only be quantified for about a quarter of questions. It could not be deduced from the available information that the added benefit could not be quantified, especially in small target populations.

The analysis of the research questions with non-quantifiable added benefit shows that in a total of 52 of the 85 research questions assessed by the G-BA in the years 2014 to 2018, evidence gaps were presented that were decisive for the lack of quantifiability (61%). Following this,

almost two-thirds of the orphan drug assessments are potential candidates for the collection of routine practice data according to GSAV if these are to serve the quantification of the added benefit. The evidence gaps exist for approx. 80% of the research questions for the respective orphan drug itself and for 100% of the control group, also because data on the control group were often completely missing. Therefore, a targeted collection of routine practice data according to GSAV must as a rule be planned and conducted in a comparative manner, involving a control group relevant from the point of view of the G-BA (comparator therapy).

Gaps in evidence are often present in several outcome categories (mortality, morbidity, health-related quality of life and adverse effects), although not always in all outcome categories. As evidence gaps often also exist in the outcome categories of morbidity and health-related quality of life, information on PROs will often be required for a targeted collection of routine practice data according to GSAV, as this is essential for addressing the evidence gaps in these outcome categories.

It is therefore overall foreseeable that data collection planned as a requirement for marketing authorization, which is intended in particular to identify rare or late-onset adverse effects of the respective orphan drug, will often not be suitable in unchanged form for targeted data collection according to GSAV for use in benefit assessments. This is because evidence gaps usually exist in several outcome categories and thus also in others (e.g. mortality and health-related quality of life) and they also always exist for the control group. What change or extension in the data collection required by the regulatory authorities is necessary for a targeted collection of routine practice data according to GSAV for benefit assessments (in particular: outcomes [e.g. PROs], control group, duration of observation, observation intervals) must be examined in each case based on the existing evidence gap for the quantification of added benefit.

5.7.3 Proposals for the procedure of the collection of routine practice data according to GSAV in the benefit assessment procedure according to §35a of SGB V

Based on the analysis of the orphan drug assessments for the years 2014 to 2018 and on the requirements of SGB V, Table 13 shows possible process steps for the collection of routine practice data according to GSAV in the benefit assessment procedure according to §35a SGB V.

Table 13: Process steps for routine practice data collection according to GSAV for benefit assessments according to §35a SGB V

Process step	Comment
Identification of an evidence gap in the G-BA decision on a benefit assessment according to §35a SGB V	▪ Evidence gap: relevant data gap for the comparison of the new drug with the (appropriate) comparator therapy with regard to patient-relevant outcomes (especially if the evidence gap does not allow quantification of the added benefit)
Description of the G-BA specifications for routine practice data collection according to GSAV and transmission to the pharmaceutical company	▪ Definition of the research question (see Sections 5.3.1 and 5.7.1) ▪ Duration, type and scope of data collection (duration of data collection per patient, sample size based on a sample size estimation) ▪ Type and scope of the analysis (depending on the study type used; see Sections 5.3.4 and 5.5.2) ▪ Specification of the time points for the evaluation of the data obtained (at least every 18 months) ▪ Specification of the requirements, taking into account ongoing and planned data collection, especially those resulting from requirements of the regulatory authorities (e.g. EMA)
Evaluation of the data collected and the obligation to collect data	▪ At the time of the first evaluation, the G-BA will check whether a (publicly available) study protocol including an analysis plan is available that reflects the routine practice data collection according to GSAV as requested ▪ At the first and each subsequent evaluation time point, the G-BA will evaluate the available data and decide whether the data collection can be stopped or should be continued
EMA: European Medicines Agency; G-BA: Gemeinsamer Bundesausschuss (Federal Joint Committee); GSAV: Gesetz für mehr Sicherheit in der Arzneimittelversorgung (Law for More Safety in the Supply of Medicines [own translation]); SGB: Sozialgesetzbuch (Social Code Book)	

With the decision on a benefit assessment according to §35a SGB V, the G-BA determines whether there is a gap in the evidence for the assessment of a drug and, if so, which one. This determination forms the basis for the description of the requirements for the collection of routine practice data according to GSAV. The specifications for the pharmaceutical company include the components listed in Table 13. In defining the specifications, the G-BA considers ongoing and planned data collections, especially those arising from requirements of the regulatory authorities (e.g., EMA). In this step, it is also determined at what point in time the collected data should be evaluated.

It should be discussed to what extent the G-BA or IQWiG should be involved in the development of study protocols for the data collection required. The goal of such involvement would be to ensure that the planned data collection is basically usable in a benefit assessment.

At the time of the first evaluation of the collection of routine practice data according to GSAV, the G-BA checks whether the study to be conducted is registered in a study registry and whether at least a study protocol (including the analysis plan) depicting the data collection according to the specifications is publicly available. This protocol is a prerequisite for conducting the data collection.

The available data is also evaluated at the specified evaluation times. In a first step, it can be checked whether the data are basically suitable for the benefit assessment. In a second step, the re-assessment can then be performed.

Depending on the available data, the G-BA can decide whether the data collection should be continued or whether it can be finalized, either because the existing evidence gap has been closed or because the data collection cannot provide sufficient evidence for the re-assessment of the added benefit.

Potential requirement for the collection of routine practice data according to GSAV in the run-up to a benefit assessment

In certain cases, it may be useful to start the collection of routine practice data according to GSAV at market access and thus at the beginning of a benefit assessment. This would be the case, for example, if it is to be expected that a large proportion of the patients for whom a new drug is suitable will be treated immediately after market access or if a drug is used once only. Data collection starting immediately at market access could help to considerably reduce the period during which insufficient evidence of the added benefit of the drug is available.

In order to already specify the collection of routine practice data according to GSAV in the run-up to a benefit assessment and then to request it from the time of market access, the G-BA can draw upon information from the approval procedure or on its own literature searches.

6 Discussion

6.1 Study designs for the use of routine practice data

A meaningful discussion of suitable study designs for the use of routine practice data must mandatorily consider the research question to be answered by the data collection. The decision as to whether a study should be conducted with a single study arm or as a comparative study with at least 2 arms is already determined by the research question. Thus, for example, data describing the type of health care provided for a disease or characterizing the patient population receiving certain interventions can be collected in single-arm studies. If, on the other hand, effects of an intervention are to be investigated for a benefit assessment, comparative studies are necessary (see Section 5.2). Therefore, only comparative studies will be discussed in the following text.

Even considering the arguments of the current intensive discussion on the use of routine practice data to support decisions in the health care system, it must be stated that conclusions on the benefit and harm of medical interventions based on comparative studies without randomization as a rule are more uncertain than studies with randomization. Furthermore, based on the concept of emulating a target trial (see Section 5.3.2), comparative studies without randomization are as a rule associated with considerably greater effort than those with randomization, provided that the requirements for data quality are the same. How relevant the difference in uncertainty is depends on the context of the decision. The greater uncertainty may be of secondary importance if interventions have a low risk of harm for individual patients or the health care system. Uncertainty may also be acceptable if the observed effects of an intervention are so large that they can no longer plausibly be explained by the influence of confounders. In these cases, it can be assumed that an effect is retained [143], even if it cannot then be quantified.

The increased uncertainty of results of comparative studies without randomization has led to an intensive discussion of methods [22,23,140,144]. The fundamental problem in comparative studies without randomization is the potential lack of similarity of the treatment groups in terms of prognostic factors, which leads to a questioning of the causal relationship between intervention and effect. The methodological approaches for the use of data from these studies therefore primarily aim to approximate the similarity of the treatment groups in terms of prognostic factors, e.g. by emulating a randomized target study and by adjusting for confounders (see Sections 5.3.2 and 5.5.2). These methodological approaches require detailed planning of the study to ensure the necessary data collection and are only possible after appropriately extensive and specific data collections and with sufficiently large samples. The conduct of a high-quality comparative study without randomization is therefore associated with a high level of effort. However, the methods of confounder control cannot fundamentally rule out bias. In particular, it remains unclear for the individual study to what extent confounder adjustment is successful, since even if the known relevant confounders are taken into account, additional unknown relevant confounders may be present.

Randomization, on the other hand, ensures similarity of the treatment groups in terms of prognostic factors with little effort and thus, depending on the planned outcomes, enables studies with a smaller sample size and less extensive data collection. This advantage is particularly relevant for studies in small patient populations.

Optimized studies for decision-making in health care

Despite the relationships between randomization and the efficiency of study design described in the previous section, there is a perception that comparative studies without randomization could be conducted more easily, more quickly and possibly with more relevant results than comparative studies with randomization. The common arguments describe comparative studies with randomization as studies including narrowly defined patient populations that do not depict the patients treated in routine practice. Furthermore, it is claimed that these studies are neither feasible in small patient populations (e.g. for rare diseases) nor suited to investigate rare or late-onset events. Finally yet importantly, it is also claimed that feasibility is limited by the great effort and costs involved.

Patient populations with relevance for routine practice

It is true that currently many comparative studies with randomization include narrowly defined patient populations via their inclusion and exclusion criteria. However, randomization is not mandatorily associated with such a restriction of the patient population. A comparative study with randomization can also include a population covering patients who are treated in routine practice with the interventions to be investigated. These so-called pragmatic randomized trials are already being conducted [145,146], and recommendations for the design of such trials have been developed [27,29].

Studies investigating rare and late events

The investigation of rare events requires large case numbers, while the observation of late-onset events requires long observation periods. It is precisely for these case constellations that comparative (non-interventional) studies without randomization are often proposed, instead of comparative studies with randomization.

On the other hand, so-called large simple trials are discussed as a solution to these situations. These are comparative studies with randomization that collect a limited data set specifically tailored to the existing research question and are often conducted in routine practice [29]. This means that there is an overlap with the pragmatic randomized trials described above. For these studies, the collection of data in registries, or in electronic patient records or using claims data from health insurance funds is also being discussed.

Efficient design for meaningful results in small populations

The feasibility of comparative studies with randomization is in particular being questioned for interventions to be used in small patient populations, such as those for rare diseases. Empirical investigations of studies in rare diseases invalidate this argument. For example, a systematic

review of studies registered in ClinicalTrials.gov showed that many comparative randomization studies are also conducted in rare and very rare diseases [147]. The same conclusion is reached in an IQWiG report examining underlying studies for orphan drug approvals by the EMA for the years 2001 to 2013 [148]. The review of the underlying studies in decisions by the G-BA on the early benefit assessment of orphan drugs in the present report also showed a high proportion of comparative studies with randomization (see Section 5.7.2). These results show that the feasibility of comparative studies with randomization in small populations is not in question.

The conduct of comparative studies without randomization appears questionable, particularly in small populations, because the available sample sizes do not usually allow adequate adjustment for relevant confounders. In small populations, the necessary methods for these studies without randomization may not be applicable at all. Decision-making based on such studies is therefore associated with a high degree of uncertainty for the patients concerned.

In the case of very small sample sizes, it therefore seems more reasonable to resort to a controlled increase in uncertainty. For example, the statistical error level can be increased above the usual value of (2-sided) 5% to 10% [148,149]. In descending order of priority, restrictions in the external validity could also be accepted, e.g. by including data from similar therapeutic indications or by using established surrogate outcomes within combined outcomes [148].

These compromises in the required precision of the results or the definition of the outcomes are more likely to lead to interpretable results than the performance of a comparative study without randomization that is insufficiently adjusted for confounders.

Minimization of effort and costs

Other reasons named for difficulties in conducting comparative studies with randomization are the costs, effort and duration required. However, it should be considered that costs and effort are not primarily caused by randomization, but by the measures required in these studies to ensure data quality. However, ensuring data quality is equally necessary for studies without randomization.

Here, too, the solution appears to lie in changing the way these studies are conducted rather than in abandoning this meaningful study design. New tools, such as registry-based RCTs, should be used to conduct studies within existing data structures and thus improve recruitment, reduce costs, and enable long-term data collection [13,42,44]. Registry-based RCTs can represent an option to reduce the effort required for comparative studies with randomization, especially in combination with the above-described approaches of pragmatic RCTs and large simple trials limited to the collection of data on essential outcomes.

Summary

In summary, it may be easier and more effective to conduct a comparative study with randomization, considering the adjustments described above, than to try to generate high quality results from a comparative study without randomization.

6.2 General recommendations for action based on the results of the registry interviews

The analysis of the possible concepts for generating routine practice data for benefit assessments has shown that in the German health care context, registries will probably be the most important tool for the collection of routine practice data for the near future. This applies regardless of whether the registry-based studies are comparative studies with or without randomization.

It cannot be answered in general terms whether and to what extent the various patient registries are already suitable at this point in time for answering future research questions in benefit assessments of drugs, as this depends on the respective registry and the type of registry study, but also in particular on the specific research questions. However, on the basis of the analyses and discussions with the registry operators, various fields of action can be described that serve to support the individual registries in particular (and the registry landscape in Germany in general) in the task of collecting routine practice data for benefit assessments of drugs. In this context, a distinction can be made between 3 (partially overlapping) target groups: registry operators, contracting parties or persons responsible for registry studies as well as decision-makers in health care and health policy. In the following text, recommendations for action are listed for these 3 target groups with the aim of supporting the collection of routine practice data for benefit assessments of drugs.

Registry operators

For the individual registries, the following 3 steps in particular could be used to examine the extent to which they represent a suitable data source with regard to the collection of routine practice data for benefit assessments of drugs:

- Examination of the registry with regard to the mandatory criteria for ensuring data quality mentioned in Section 5.5.1.3 as well as the general criteria that are as a rule relevant for registry studies for use in benefit assessments of drugs.
- Emulation of one or more target trials (see Section 5.3.2) in the registry on the basis of actually existing evidence gaps in completed benefit assessments of drugs. For this purpose, the decisions and the supporting reasons of the G-BA as well as the corresponding benefit assessments of IQWiG or the G-BA (for orphan drugs) can be used. In the case of orphan disease registries, it is useful to emulate target trials on research questions relating to orphan drugs with a non-quantifiable added benefit (see Appendix D). For the other diseases, it seems appropriate to identify relevant evidence gaps from decisions of initially limited duration or from research questions for which the added benefit has not been proven.

- Identification of the relevant confounders for these target trials or the research questions investigated with them. This should be done as systematically as possible, e.g. on the basis of scientific literature in consultation with experts.
- Examination of whether studies with randomization can in principle be conducted in the registry or which (organizational) hurdles exist in this respect and how these can be removed.

With these steps, the suitability of the individual registries for the collection of routine practice data for benefit assessments of drugs can be estimated and any need for adaptation described. At the same time, the specific procedure of emulating a target trial and the systematic compilation of the relevant confounders can be piloted, unless experience in this regard is already available elsewhere. Furthermore, the registries can be prepared for the conduct of registry-based studies with randomization.

The proposed procedure does not mean that an expansion of each registry (which would possibly be very costly and reach organizational limits) is to be initiated, with the aim of being able to answer every foreseeable research question in the future. Rather, such an inventory can serve the purpose of targeted adaptation in individual areas, or even a conscious decision against such an adaptation if it contradicted other registry goals.

It would be useful to publish the specific procedure and the corresponding results, on the one hand, to inform the G-BA, IQWiG, and potential contracting parties for registry studies about evaluating the suitability of the respective registry, and on the other hand, to provide mutual support for the registry operators for suitability testing.

Contracting parties or other parties responsible for registry studies

Standard protocols and analysis plans should be developed for the conduct of registry studies for benefit assessments of drugs within the context of §35a SGB V. It can be assumed that existing standard documents, e.g. from the regulatory area, can be used to a large extent for this purpose. However, the following points in particular should be specifically implemented:

- development of generic methods for the systematic identification of all confounders relevant to the respective research question; description of these methods in the standard study protocol
- emulation of the target trial, with examples of typical PICO questions for benefit assessments of drugs
- generic description of how the research-question-related data quality of the registry is to be presented (including information on completeness and correctness for each relevant data field)

- generic description of the analysis according to Section 5.5.2, including prospective specification of the decision structure in procedures for the adjustment of confounders, depending on the specific data

These standard documents should be developed together with registry operators, since if they are designed for a specific research question, some elements require detailed knowledge of the respective data source (the respective registry). Furthermore, a reporting structure should be developed for a standard report of registry studies for benefit assessments of drugs according to §35a SGB V.

Decision-makers in health care and health policy

In Germany, the collection of routine practice data according to GSAV for benefit assessments of drugs is a political aim reflected in the corresponding legislation [1]. Discussions with the registry operators have identified various obstructive factors to the establishment and operation of a registry and thus to the achievement of the above-mentioned goal, namely, the use of collections of routine practice data for benefit assessments of drugs. Some of these factors are due to general conditions that can only be addressed by health care and health policy decision-makers across registries. For the 3 topics of data protection, IT landscape and funding, the main obstacles and the resulting options for action are described below.

Data protection

Variations in data protection requirements at different levels, e.g. between federal states, different universities or different hospitals, even within a federal state, hinder the establishment and the potentially necessary expansion of a registry for the collection of routine practice data for benefit assessments of drugs. It is not understandable why patients in hospital A in a northern federal state should be subject to different data protection regulations from patients in hospital B in the south or university C in the east.

It is recommended to create a uniform and binding framework for the purposes of patient registries for all parties involved (especially registry operators, centres, patients, and data protection experts).

IT landscape

According to the registry operators, the variations in the IT landscape in the centres and in particular the different HIS, combined with partially outdated hardware, represent a major obstacle to the operation of registries.

As fields of action, this results on the one hand in possible (special) investments in the technical infrastructure on the basis of the minimum standards to be defined. On the other hand, consideration could be given to legally anchoring requirements for HIS, like those for medical practice software in the outpatient sector (§73 (9) SGB V), with the aim of creating uniform interfaces and uniform core data sets, taking into account semantic interoperability.

Registry funding

With legally justified exceptions (e.g. cancer registries), there is no sustainable public funding for patient registries, which in the view of the registry operators can be an obstacle to the establishment and operation of a registry, but especially to ensuring a high level of data quality for benefit assessments of drugs.

It can be argued that, like clinical studies, registry studies on drugs should be funded by the contracting parties of the studies. In contrast to clinical studies, however, it is expected that a technical and organizational infrastructure would be maintained in the long term, which could be used in the short term for the planning and conduct of registry studies. Irrespective of the question as to whether the funding of registry studies by the pharmaceutical industry endangers the independence of a registry, such funding in any case cannot ensure the continuous availability and operation of registries.

From the point of view of some registry operators, it is therefore hardly feasible to target the registry towards the future conduct of registry studies for benefit assessments of drugs if this is associated with high and non-counter-funded costs. This could potentially jeopardize other objectives of the registry, because the necessary funds would then no longer be available.

In view of the evidently great public interest in registry studies for benefit assessments of drugs, it therefore seems sensible to consider public funding of registries under certain conditions. This could, for example, include (additional) investments to meet the mandatory criteria for ensuring data quality. Due to the objective of the GSAV (benefit assessments according to §35a SGB V), the extent to which funding via the SHI system with participation of private health insurance funds would be considered would be a matter for discussion.

Funding of interventional registry studies

The conduct of registry-based interventional studies is only possible within the legal framework of the German Drug Act. However, funding of such studies by the registry operators as study sponsors is currently unrealistic due to the often high drug costs, which makes it difficult to conduct independent interventional registry studies. It could therefore be useful to regulate the coverage of costs for interventional routine practice studies within the framework of SGB V. For example, a model for regulating cost-coverage applications would be conceivable as in 35c (2) SGB V for the cost-coverage for off-label studies.

7 Conclusion

Study design and data collection

- The use of routine practice data for benefit assessments of drugs according to §35a SGB V requires a comparison between the new drug and the appropriate comparator therapy specified by the G-BA; this requires the conduct of comparative studies.
- The collection of routine practice data from electronic patient records and from claims data of health insurance funds for benefit assessments according to §35a SGB V is currently and foreseeably not considered realistic; rather, a study-specific data collection or data collection from patient registries is necessary.

Routine practice comparative studies without randomization

- If comparative studies without randomization are to be used for the benefit assessment, it must be ensured at the stage of study planning that the study conduct and the data collected are of the quality required to generate interpretable results.
- Essential components of such a study planning are a study protocol including an analysis plan, the emulation of a target trial that deals with the relevant research question, and ensuring that sufficient data are collected for confounder control.
- A key aspect of the analysis of a comparative study without randomization is adequate confounder adjustment; this adjustment must be pre-specified as far as possible and the assumptions made (e.g. the definition of the relevant confounders) must be substantiated.
- No effects can be derived from comparative studies without randomization if the data quality in the data sources used and the quality of analysis and reporting is not high.
- Even under high quality requirements (for data, analysis and reporting), no more than a hint of an effect can normally be derived from comparative studies without randomization.
- Due to the inherent uncertainty of the results from comparative studies without randomization, because of potentially unknown confounders, a conclusion on the benefit or harm of an intervention should only be derived from the effects observed in the study if these effects exceed a certain effect size. Quantification of an added benefit according to the legally prescribed extent categories requires corresponding effect sizes graded according to magnitude.
- The possibility to consider retrospective study designs depends on whether the available data sources already contain the necessary data in the required quality; comparisons of a new drug with historical controls only appear realistic if the same data source (e.g. a disease-specific clinical registry) is used for the new drug and the historical control.

Routine practice comparative studies with randomization

- Routine practice comparative studies can also be randomized (pragmatic clinical trials).

- The effort required for a routine practice comparative study with randomization will generally – with comparable data quality – be less than the effort required for a study without randomization, as confounder data collection and confounder adjustment can be omitted.
- Routine practice comparative studies with randomization are of higher informative value than those without randomization, and the quantification of added benefit is more reliable.
- Especially after market authorization, depending on the existing research question, routine practice comparative studies with randomization can be conducted with limited data collection ([large] simple trials); conducting them in registries has an additional potential to accelerate the conduct of the studies and make them less complex and resource-intensive (registry-based comparative studies with randomization).

References

1. Bundesministerium für Gesundheit. Gesetz für mehr Sicherheit in der Arzneimittelversorgung (GSAV). Bundesgesetzblatt 2019; Teil 1(30): 1202-1220.

2. Müller D, Augustin M, Banik N, Baumann W, Bestehorn K, Kieschke J et al. Memorandum Register für die Versorgungsforschung. Gesundheitswesen 2010; 72(11): 824-839.

3. Stausberg J, Maier B, Bestehorn K, Gothe H, Gröne O, Jacke C et al. Memorandum Register für die Versorgungsforschung: Update 2019 [manuscript accepted]. Gesundheitswesen.

4. Gliklich R, Dreyer N, Leavy M. Registries for evaluating patient outcomes: a user´s guide; 3rd edition; AHRQ pub. no. 13(14)-EHC111. Rockville: Agency for Healthcare Research and Quality; 2014. URL: https://effectivehealthcare.ahrq.gov/sites/default/files/related_files/registries-guide-3rd-edition-vol-2-140430.pdf

5. European Medicines Agency. Discussion paper: use of patient disease registries for regulatory purposes; methodological and operational considerations [online]. 05.11.2018 [Accessed: 23.05.2019]. URL: https://www.ema.europa.eu/documents/other/discussion-paper-use-patient-disease-registries-regulatory-purposes-methodological-operational_en.docx.

6. European Network for Health Technology Assessment. REQueST (Registry Evaluation and Quality Standards Tool) [online]. 30.09.2019 [Accessed: 18.12.2019]. URL: https://eunethta.eu/wp-content/uploads/2019/10/Registry-Evaluation-and-Quality-Standards-Tool-REQueST-1.xlsm.

7. Mandeville KL, Valentic M, Ivankovic D, Pristas I, Long J, Patrick HE. Quality assurance of registries for Health Technology Assessment. Int J Technol Assess Health Care 2018; 34(4): 360-367.

8. Clinical Trials Transformation Initiative. CTTI recommendations: registry trials [online]. [Accessed: 18.12.2019]. URL: https://www.ctti-clinicaltrials.org/files/recommendations/registrytrials-recs.pdf.

9. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL et al. Real-world evidence: what is it and what can it tell us? N Engl J Med 2016; 375(23): 2293-2297.

10. Senn S. Statistical issues in drug development. Chichester: Wiley; 1997.

11. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH guidelines [online]. [Accessed: 18.12.2019]. URL: https://www.ich.org/page/ich-guidelines.

12. EQUATOR Network. Webauftritt [online]. [Accessed: 28.11.2019]. URL: https://www.equator-network.org.

13. Food and Drug Administration. Framework for FDA's Real-World Evidence Program [online]. 12.2018 [Accessed: 16.07.2019]. URL: https://www.fda.gov/media/120060/download.

14. Patient-Centered Outcomes Research Institute. PCORI methodology standards [online]. 26.02.2019 [Accessed: 18.12.2019]. URL: https://www.pcori.org/sites/default/files/PCORI-Methodology-Standards.pdf.

15. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ et al (Ed). Cochrane handbook for systematic reviews of interventions, version 6.0. Hoboken: Wiley-Blackwell; 2019.

16. Bundesministerium für Gesundheit. Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz - AMNOG). Bundesgesetzblatt Teil 1 2010; (67): 2262-2277.

17. Bundesministerium für Gesundheit. Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung - AM-NutzenV) [online]. 09.08.2019 [Accessed: 18.12.2019]. URL: http://www.gesetze-im-internet.de/am-nutzenv/index.html.

18. Schneeweiss S. Real-world evidence of treatment effects: the useful and the misleading. Clin Pharmacol Ther 2019; 106(1): 43-44.

19. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Sofosbuvir (chronische Hepatitis C bei Jugendlichen): Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A17-55 [online]. 11.01.2018 [Accessed: 22.01.2018]. (IQWiG-Berichte; Volume 582). URL: https://www.iqwig.de/download/A17-55_Sofosbuvir_Nutzenbewertung-35a-SGB-V_V1-0.pdf.

20. Wellek S, Donner-Banzhoff N, König J, Mildenberger P, Blettner M. Studienplanung und -auswertung nach dem Stepped-Wedge-Design: Teil 26 der Serie zur Bewertung wissenschaftlicher Publikationen. Dtsch Arztebl 2019; 116(26): 453-458.

21. U.S. National Library of Medicine. ClinicalTrials.gov [online]. 2019 [Accessed: 28.11.2019]. URL: https://clinicaltrials.gov.

22. Hernan MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016; 183(8): 758-764.

23. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016; 355: i4919.

24. Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. J Clin Epidemiol 1995; 48: 23-40.

25. Eapen ZJ, Lauer MS, Temple RJ. The imperative of overcoming barriers to the conduct of large, simple trials. JAMA 2014; 311(14): 1397-1398.

26. Ford I, Norrie J. Pragmatic trials. N Engl J Med 2016; 375(5): 454-463.

27. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015; 350: h2147.

28. Patient-Centered Outcomes Research Institute. Guidance on the design and conduct of trials in real-world settings: factors to consider in pragmatic patient centered outcomes research [online]. [Accessed: 18.12.2019]. URL: https://www.pcori.org/sites/default/files/PCORI-Guidance-Design-and-Conduct-of-Trials-Real-World-Settings-Factors-to-Consider-Pragmatic-PCOR.pdf.

29. Zuidgeest MGP, Goetz I, Groenwold RHH, Irving E, Van Thiel G, Grobbee DE. Series: pragmatic trials and real world evidence; paper 1: introduction. J Clin Epidemiol 2017; 88: 7-13.

30. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic Reviews 2015; 4: 1.

31. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23(20): 3105-3124.

32. Rücker G. Network meta-analysis, electrical networks and graph theory. Res Syn Meth 2012; 3(4): 312-324.

33. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50(6): 683-691.

34. March S, Andrich S, Drepper J, Horenkamp-Sonntag D, Icks A, Ihle P et al. Gute Praxis Datenlinkage (GPD). Gesundheitswesen 2019; 81(8-9): 636-650.

35. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Biotechnologisch hergestellte Wirkstoffe bei rheumatoider Arthritis: Abschlussbericht; Auftrag A16-70 [online]. 23.07.2019 [Accessed: 30.10.2019]. URL: https://www.iqwig.de/download/A16-70_Biotechnologisch-hergestellte-Wirkstoffe-bei-rheumatoider-Arthritis_Abschlussbericht_V1-0.pdf.

36. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 5.0. Köln: IQWiG; 2017. URL: https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf.

37. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Linagliptin: erneute Nutzenbewertung gemäß § 35a Absatz 5b SGB V; Dossierbewertung; Auftrag A12-11 [online]. 29.11.2012 [Accessed: 11.03.2013]. (IQWiG-Berichte; Volume 144). URL: https://www.iqwig.de/download/A12-11_Linagliptin_Erneute_Nutzenbewertung_35a_Absatz_5b_SGB_V.pdf.

38. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009; 338: b1147.

39. Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect comparisons: a review of reporting and methodological quality. PLoS One 2010; 5(11): e11054.

40. Kiefer C, Sturtz S, Bender R. Indirekte Vergleiche und Netzwerk-Metaanalysen: Schätzung von Effekten bei fehlenden direkten Vergleichsstudien; Teil 22 der Serie zur Bewertung wissenschaftlicher Publikationen. Dtsch Arztebl 2015; 112(47): 803-808.

41. Meinecke AK, Welsing P, Kafatos G, Burke D, Trelle S, Kubin M et al. Series: pragmatic trials and real world evidence; paper 8: data collection and management. J Clin Epidemiol 2017; 91: 13-22.

42. Lauer MS, D'Agostino RB. The Randomized Registry Trial: the next disruptive technology in clinical research? N Engl J Med 2013; 369(17): 1579-1581.

43. Li G, Sajobi TT, Menon BK, Korngut L, Lowerison M, James M et al. Registry-based randomized controlled trials: what are the advantages, challenges, and areas for future research? J Clin Epidemiol 2016; 80: 16-24.

44. Foroughi S, Wong HI, Gately L, Lee M, Simons K, Tie J et al. Re-inventing the randomized controlled trial in medical oncology: the registry-based trial. Asia Pac J Clin Oncol 2018; 14(6): 365-373.

45. Lagerqvist B, Fröbert O, Olivecrona GK, Gudnason T, Maeng M, Alström P et al. Outcomes 1 year after thrombus aspiration for myocardial infarction. N Engl J Med 2014; 371(12): 1111-1120.

46. Glinatsi D, Heiberg MS, Rudin A, Nordstrom D, Haavardsholm EA, Gudbjornsson B et al. Head-to-head comparison of aggressive conventional therapy and three biological treatments and comparison of two de-escalation strategies in patients who respond to treatment: study protocol for a multicenter, randomized, open-label, blinded-assessor, phase 4 study. Trials 2017; 18(1): 161.

47. Mathes T, Buehn S, Prengel P, Pieper D. Registry-based randomized controlled trials merged the strength of randomized controlled trails and observational studies and give rise to more pragmatic trials. J Clin Epidemiol 2018; 93: 120-127.

48. Yndigegn T, Hofmann R, Jernberg T, Gale CP. Registry-based randomised clinical trial: efficient evaluation of generic pharmacotherapies in the contemporary era. Heart 2018; 104(19): 1562.

49. Eichler HG, Bloechl-Daum B, Broich K, Kyrle PA, Oderkirk J, Rasi G et al. Data rich, information poor: can we use electronic health records to create a learning healthcare system for pharmaceuticals? Clin Pharmacol Ther 2019; 105(4): 912-922.

50. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. Annu Rev Public Health 2016; 37: 61-81.

51. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. Clin Pharmacol Ther 2019; 106(1): 36-39.

52. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. CMAJ 2016; 188(8): E158-E164.

53. Horenkamp-Sonntag D, Linder R, Engel S, Verheyen F. Brachytherapie beim Prostata-Ca: Möglichkkeiten und Limitationen von GKV-Routinedaten für eine Nutzenbewertung. Gesundheitswesen 2016; 78(5): 298-305.

54. Medicines & Healthcare products Regulatory Agency. MHRA position statement and guidance: electronic health records [online]. 16.09.2015 [Accessed: 18.12.2019]. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/470228/Electronic_Health_Records_MHRA_Position_Statement.pdf.

55. Food and Drug Administration. Use of electronic health record data in clinical investigations: guidance for industry [online]. 07.2018 [Accessed: 21.11.2019]. URL: https://www.fda.gov/media/97567/download.

56. European Medicines Agency. Reflection paper on expectations for electronic source data and data transcribed to electronic data collection tools in clinical trials [online]. 09.06.2010 [Accessed: 21.11.2019]. URL: https://www.ema.europa.eu/documents/regulatory-procedural-guideline/reflection-paper-expectations-electronic-source-data-data-transcribed-electronic-data-collection_en.pdf.

57. Bertelsmann Stiftung. Elektronische Patientenakten: Einrichtungsübergreifende Elektronische Patientenakten als Basis für integrierte patientenzentrierte Behandlungsmanagement-Plattformen [online]. 04.2017 [Accessed: 18.12.2019]. URL: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/VV_eEPA_Expertise_final.pdf.

58. Garets D, Davis M. Electronic medical records vs. electronic health records: yes, there is a difference; a HIMSS Analytics white paper. Chicago: HIMSS Analytics; 2006.

59. Organisation for Economic Co-operation and Development. Using routinely collected data to inform pharmaceutical policies: analytical report for OECD and EU countries [online]. 2019 [Accessed: 24.05.2019]. URL: https://www.oecd.org/health/health-systems/Using-Routinely-Collected-Data-to-Inform-Pharmaceutical-Policies-Analytical-Report-2019.pdf.

60. Pacurariu A, Plueschke K, McGettigan P, Morales DR, Slattery J, Vogl D et al. Electronic healthcare databases in Europe: descriptive analysis of characteristics and potential for use in medicines regulation. BMJ Open 2018; 8(9): e023090.

61. De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B et al. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015; 53: 162-173.

62. Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc 2014; 21(4): 578-582.

63. Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: shining light on a new research paradigm. Ann Intern Med 2015; 163(8): 635-636.

64. Caumanns J. Zur Diskussion: Stand der Digitalisierung im deutschen Gesundheitswesen. Z Evid Fortbild Qual Gesundhwes 2019; 143: 22-29.

65. Stephani V, Busse R, Geissler A. Benchmarking in der Krankenhaus-IT: Deutschland im internationalen Vergleich. In: Klauber J, Geraedts M, Friedrich J, Wasem J (Ed). Krankenhaus-Report 2019. Berlin: Springer; 2019. URL: https://link.springer.com/content/pdf/10.1007%2F978-3-662-58225-1_2.pdf.

66. Krüger-Brand HE. Medizininformatik-Initiative: Impulse für die digitale Medizin. Dtsch Arztebl 2019; 116(42): A1883-A1888.

67. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 2013; 51(8 Suppl 3): S30-S37.

68. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013; 20(1): 144-151.

69. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F et al. Electronic health records to facilitate clinical research. Clin Res Cardiol 2017; 106(1): 1-9.

70. Ehrenstein V, Petersen I, Smeeth L, Jick SS, Benchimol EI, Ludvigsson JF et al. Helping everyone do better: a call for validation studies of routinely recorded health data. Clin Epidemiol 2016; 8: 49-51.

71. Mc Cord KA, Al-Shahi Salman R, Treweek S, Gardner H, Strech D, Whiteley W et al. Routinely collected data for randomized trials: promises, barriers, and implications. Trials 2018; 19(1): 29.

72. Najafzadeh M, Schneeweiss S. From trial to target populations: calibrating real-world data. N Engl J Med 2017; 376(13): 1203-1205.

73. Mc Cord KA, Ewald H, Ladanie A, Briel M, Speich B, Bucher HC et al. Current use and costs of electronic health records for clinical trial research: a descriptive study. CMAJ Open 2019; 7(1): E23-E32.

74. Brennan L, Watson M, Klaber R, Charles T. The importance of knowing context of hospital episode statistics when reconfiguring the NHS. BMJ 2012; 344: e2432.

75. Green SM. Congruence of disposition after emergency department intubation in the National Hospital Ambulatory Medical Care Survey. Ann Emerg Med 2013; 61(4). 423-426.e8.

76. Bayley KB, Belnap T, Savitz L, Masica AL, Shah N, Fleming NS. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. Med Care 2013; 51(8 Suppl 3): S80-S86.

77. Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, LoCasale R. Guidelines for good database selection and use in pharmacoepidemiology research. Pharmacoepidemiol Drug Saf 2012; 21(1): 1-10.

78. Guinn D, Wilhelm EE, Lieberman G, Khozin S. Assessing function of electronic health records for real-world data generation. BMJ Evid Based Med 2019; 24(3): 95-98.

79. Wasserman RC. Electronic medical records (EMRs), epidemiology, and epistemology: reflections on EMRs and future pediatric clinical research. Acad Pediatr 2011; 11(4): 280-287.

80. Pokora RM, Cornet LL, Daumke P, Mildenberger P, Zeeb H, Blettner M. Validierung von semantischen Analysen von unstrukturierten medizinischen Daten fur Forschungszwecke. Gesundheitswesen 09.10.2019 [Epub ahead of print].

81. Köpcke F, Trinczek B, Majeed RW, Schreiweis B, Wenk J, Leusch T et al. Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence. BMC Med Inform Decis Mak 2013; 13: 37.

82. Lau EC, Mowat FS, Kelsh MA, Legg JC, Engel-Nitz NM, Watson HN et al. Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. Clin Epidemiol 2011; 3: 259-272.

83. Moore TJ, Furberg CD. Electronic health data for postmarket surveillance: a vision not realized. Drug Saf 2015; 38(7): 601-610.

84. PMV Forschungsgruppe. Gutachten: Daten für die Versorgungsforschung; Zugang und Nutzungsmöglichkeiten; Datengutachten für das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI) [online]. 08.10.2014 [Accessed: 18.12.2019]. URL: https://www.dimdi.de/static/.downloads/deutsch/dimdi-sekundaerdaten-expertise.pdf.

85. Horenkamp-Sonntag D, Linder R, Wenzel F, Gerste B, Ihle P. Prüfung der Datenqualität und Validität von GKV-Routinedaten. In: Swart E, Ihle P, Gothe H, Matusiewicz D (Ed). Routinedaten im Gesundheitswesen. Bern: Huber; 2014. p. 314-330.

86. Erler A, Beyer M, Muth C, Gerlach FM, Brennecke R. Garbage in - Garbage out? Validität von Abrechnungsdiagnosen in hausärztlichen Praxen. Gesundheitswesen 2009; 71(12): 823-831.

87. Munch C, Gottschall M, Hübsch G, Köberlein-Neu J, Schubel J, Bergmann A et al. Qualitat der hausärztlichen Diagnosedokumentation in Patientenakten: eine Analyse am Beispiel von Schilddrüsenerkrankungen. Z Evid Fortbild Qual Gesundhwes 2016; 115-116: 56-62.

88. Czwikla J, Domhoff D, Giersiepen K. ICD-Codierqualität ambulanter Krebsdiagnosen in GKV-Routinedaten. Z Evid Fortbild Qual Gesundhwes 2016; 118-119: 48-55.

89. Gerste B, Gutschmidt S. Datenqualität von Diagnosedaten aus dem ambulanten Bereich: kritische Anmerkungen am Beispiel Diabetes. Gesundheits- und Sozialpolitik 2006; (3-4): 10-24.

90. Schubert I, Ihle P, Koster I. Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. Gesundheitswesen 2010; 72(6): 316-322.

91. Walker J, Buchwald V. Plausibilität der Diagnosekodierung in der ambulanten Versorgung: Analyse auf Basis eines Ärztepanels. In: Repschläger U, Schulte C, Osterkamp N (Ed). Gesundheitswesen aktuell 2011. Barmer GEK; 2011. p. 128-151.

92. AG medizinische Grouperanpassung. Analysen der Diagnoseeinträge in der Datengrundlage des Bewertungsausschusses für die Ermittlung der diagnosebezogenen Veränderungsraten für das Jahr 2016 (Geburtstagsstichprobe Version 7.0) [online]. 14.06.2017 [Accessed: 08.01.2020]. URL: https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/aerztliche_versorgung/verguetung_und_leistungen/klassifikationsverfahren/7_Analysen_der_Diagnoseeintraege_in_der_Datengrundlage_des_BA_fuer_die_Ermittlung_der_diagnosebezogenen_Veraenderungsraten_fuer_das_Jahr_2016_Stand_14.06.2017.pdf.

93. Drösler S, Garbe E, Hasford J, Schubert I, Ulrich V, Van de Ven W et al. Sondergutachten zu den Wirkungen des morbiditätsorientierten Risikostrukturausgleichs: erstellt durch den Wissenschaftlichen Beirat zur Weiterentwicklung des Risikostrukturausgleichs im Auftrag des Bundesministeriums für Gesundheit [online]. 27.11.2017 [Accessed: 08.01.2020]. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Gesundheit/Berichte/Sondergutachten_Wirkung_RSA_2017.pdf.

94. Gerste B, Günster C. Erkrankungshäufigkeiten und Inanspruchnahme von Gesundheitsleistungen. In: Günster C, Klose J, Schmacke N (Ed). Versorgungsreport 2012. Stuttgart: Schattauer; 2012. p. 315-384.

95. Fiori W, Brüning K, Bunzemeier H, Roeder N. Unberechtigte Vorwürfe. Dtsch Arztebl 2009; 106(33): A1596-A1598.

96. Maass C, Kuske S, Lessing C, Schrappe M. Are administrative data valid when measuring patient safety in hospitals? A comparison of data collection methods using a chart review and administrative data. Int J Qual Health Care 2015; 27(4): 305-313.

97. Hoffmann F, Andersohn F, Giersiepen K, Scharnetzky E, Garbe E. Validierung von Sekundärdaten: Grenzen und Möglichkeiten. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2008; 51(10): 1118-1126.

98. GKV-Spitzenverband. Anzahl der Krankenkassen im Zeitverlauf: Konzentrationsprozess durch Fusionen (Angaben zum Stichtag 1.1.) [online]. [Accessed: 18.12.2019]. URL: https://www.gkv-spitzenverband.de/media/grafiken/krankenkassen/Grafik_Krankenkassenanzahl_Konzentrationsprozess_300dpi_2019-03-29.jpg.

99. Hoffmann F, Icks A. Unterschiede in der Versichertenstruktur von Krankenkassen und deren Auswirkungen fur die Versorgungsforschung: Ergebnisse des Bertelsmann-Gesundheitsmonitors. Gesundheitswesen 2012; 74(5): 291-297.

100. Hoffmann F, Koller D. Verschiedene Regionen, verschiedene Versichertenpopulationen? Soziodemografische und gesundheitsbezogene Unterschiede zwischen Krankenkassen. Gesundheitswesen 2017; 79(1): e1-e9.

101. Schafer I, Reusch M, Siebert J, Hilbring C, Augustin M. Assoziation von Krankenversicherung und soziodemografischen Faktoren mit der Versorgung maligner Melanome. Gesundheitswesen 2017; 79(1): 21-27.

102. Ohlmeier C. Versorgungsepidemiologische Analysen kardiovaskulärer Erkrankungen auf Basis von Routinedaten der gesetzlichen Krankenkassen [Dissertation] Bremen: Universität Bremen; 2015.

103. Marx MM, Dulas FM, Schumacher KM. Verbesserung der Sichtbarkeit seltener Erkrankungen in Gesundheitssystemen durch spezifische Routinekodierung. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2017; 60(5): 532-536.

104. Schröder H. Arzneimittelverordnungen. In: Swart E, Ihle P, Gothe H, Matusiewicz D (Ed). Routinedaten im Gesundheitswesen. Bern: Huber; 2014. p. 74-87.

105. Kuklik N, Stausberg J, Jockel KH. Adverse drug events in German hospital routine data: a validation of International Classification of Diseases, 10th revision (ICD-10) diagnostic codes. PLoS One 2017; 12(11): e0187510.

106. Schellschmidt H, Mansky T, Heller G, Robra B. Indikatoren und Messinstrumente zur Bewertung der medizinischen Ergebnisqualität im Krankenhaus. In: Stock J, Szecsenyi J (Ed). Stichwort: Qualitätsindikatoren. Bonn: Kompart Verlagsgesellschaft; 2007. p. 173-200.

107. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. J Clin Epidemiol 2004; 57(2): 131-141.

108. Nabhan C, Klink A, Prasad V. Real-world evidence: what does it really mean? JAMA Oncol 2019; 5(6): 781-783.

109. Motheral BR, Fairman KA. The use of claims databases for outcomes research: rationale, challenges, and strategies. Clin Ther 1997; 19(2): 346-366.

110. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q 1966; 44(3 Suppl): 166-206.

111. Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Stat Med 2007; 26(1): 20-36.

112. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources; the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report; part I. Value Health 2009; 12(8): 1044-1052.

113. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources; the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report; part II. Value Health 2009; 12(8): 1053-1061.

114. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources; the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report; part III. Value Health 2009; 12(8): 1062-1073.

115. Agoritsas T, Merglen A, Shah ND, O'Donnell M, Guyatt GH. Adjusted analyses in studies addressing therapy and harm: users' guides to the medical literature. JAMA 2017; 317(7): 748-759.

116. D' Agostino RB Jr, D' Agostino RB Sr. Estimating treatment effects using observational data. J Am Med Assoc 2007; 297(3): 314-316.

117. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. J Am Med Assoc 2007; 297(3): 278-285.

118. Kuss O, Blettner M, Borgermann J. Propensity Score: an alternative method of analyzing treatment effects. Dtsch Arztebl Int 2016; 113(35-36): 597-603.

119. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. Stat Med 2014; 33(7): 1242-1258.

120. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998; 17(19): 2265-2281.

121. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. BMJ 2019; 367: l5657.

122. Austin PC, Mamdani MM, Stukel TA, Anderson GM, Tu JV. The use of the propensity score for estimating treatment effects: administrative versus clinical data. Stat Med 2005; 24(10): 1563-1578.

123. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 2009; 20(4): 512-522.

124. Wyss R, Fireman B, Rassen JA, Schneeweiss S. Erratum: high-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 2018; 29(6): e63-e64.

125. Austin PC, Wu CF, Lee DS, Tu JV. Comparing the high-dimensional propensity score for use with administrative data with propensity scores derived from high-quality clinical data. Stat Methods Med Res 11.04.2019 [Epub ahead of print].

126. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Stat Sci 1999; 14(1): 29-46.

127. Skipka G, Wieseler B, Kaiser T, Thomas S, Bender R, Windeler J et al. Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs. Biom J 2016; 58(1): 43-58.

128. International Conference on Harmonisation E9 Expert Working Group. ICH harmonised tripartite guideline: statistical principles for clinical trials. Stat Med 1999; 18(15): 1905-1942.

129. European Medicines Agency. Clinical efficacy and safety guidelines: biostatistics [online]. 2019 [Accessed: 28.11.2019]. URL: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/clinical-efficacy-safety/biostatistics.

130. Fortier I, Raina P, Van den Heuvel ER, Griffith LE, Craig C, Saliba M et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. Int J Epidemiol 2017; 46(1): 103-115.

131. Mathew T, Nordström K. Comparison of one-step and two-step meta-analysis models using individual patient data. Biom J 2010; 52(2): 271-287.

132. Taichman DB, Backus J, Baethge C, Bauchner H, De Leeuw PW, Drazen JM et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. PLoS Med 2016; 13(1): e1001950.

133. Bierer BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. N Engl J Med 2016; 374(25): 2411-2413.

134. European Medicines Agency. European Medicines Agency policy on publication of clinical data for medicinal products for human use (Policy/0070) [online]. 21.03.2019 [Accessed: 08.01.2020]. URL: https://www.ema.europa.eu/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf.

135. Elm Ev, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007; 335(7624): 806.

136. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015; 12(10): e1001885.

137. Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ 2018; 363: k3532.

138. Freemantle N, Blonde L, Bolinder B, Gerber RA, Hobbs FD, Martinez L et al. Real-world trials to answer real-world questions. Pharmacoeconomics 2005; 23(8): 747-754.

139. Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. J Clin Epidemiol 2010; 63(10): 1053-1058.

140. Eichler HG, Koenig F, Arlett P, Enzmann H, Humphreys A, Petavy F et al. Are novel, non-randomised analytic methods fit for decision-making? The need for prospective, controlled and transparent validation. Clin Pharmacol Ther 01.10.2019 [Epub ahead of print].

141. Dhiman P, Lee H, Kirtley S, Collins GS. More consideration was needed when conducting non-randomised studies of interventions. J Clin Epidemiol 2019; 117: 99-108.

142. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of using real-world data to replicate clinical trial evidence. JAMA Netw Open 2019; 2(10): e1912869.

143. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. BMJ 2007; 334(7589): 349-351.

144. Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. Nat Rev Clin Oncol 2019; 16(5): 312-325.

145. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA 2018; 320(9): 867-868.

146. Woodcock A, Bakerly ND, New JP, Gibson JM, Wu W, Vestbo J et al. The Salford Lung Study protocol: a pragmatic, randomised phase III real-world effectiveness trial in asthma. BMC Pulm Med 2015; 15(1): 160.

147. Hee SW, Willis A, Tudur Smith C, Day S, Miller F, Madan J et al. Does the low prevalence affect the sample size of interventional clinical trials of rare diseases? An analysis of data from the aggregate analysis of clinicaltrials.gov. Orphanet J Rare Dis 2017; 12(1): 44.

148. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Bewertung und Auswertung von Studien bei seltenen Erkrankungen: Rapid Report; Auftrag MB13-01 [online]. 05.09.2014 [Accessed: 14.10.2014]. (IQWiG-Berichte; Volume 241). URL: https://www.iqwig.de/download/MB13-01_Rapid-Report_Studien-bei-seltenen-Erkrankungen.pdf.

149. Renfro LA, Ji L, Piao J, Onar-Thomas A, Kairalla JA, Alonzo TA. Trial design challenges and approaches for precision oncology in rare tumors: experiences of the Children's Oncology Group. JCO Precision Oncology 2019; 3: 1-13.

Appendix A – Questionnaire for the interviews on patient registries

Background

IQWiG has been commissioned by the Federal Joint Committee (G-BA) to develop scientific concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs. For your information, you can find the commission and the details of the G-BA's specifications here (in German):

- https://www.g-ba.de/downloads/39-261-3773/2019-05-02_IQWIG-Beauftragung_Konzepte-Daten-Nutzenbewertung.pdf
- https://www.g-ba.de/downloads/40-268-5720/2019-05-02_IQWIG-Beauftragung_Konzepte-Daten-Nutzenbewertung_Auftragskonkretisierung.pdf

One element of the concept will be the collection of data in registries. The commission includes the specification of quality criteria and methodological requirements for the data collected in the context of the respective data generation, as well as the description of measures to ensure data quality. For these aspects, we would also like to benefit from your experience in the operation of registries. In this context, the commission, and thus also our interview, is limited to data collection and analysis for the purpose of benefit assessments of drugs.

Content of the questionnaire

The present questionnaire comprises 3 parts:

- In Part I, we ask you some questions about your own registry to get to know it better. Here we ask you to provide answers in note form to the free-text questions or to give specific information on quantities, time points etc.
- Part II lists general quality criteria in a table. Here we would like to ask you for your assessment of the relevance and feasibility of the quality criteria mentioned. In addition, we ask you to assess the degree of fulfilment of the individual quality criteria for your registry.
- Part III names 2 research questions for the benefit assessment (scenarios). On the basis of these scenarios, we would like to discuss with you in an interview whether and how you could imagine the corresponding research questions being answered in your registry. Written feedback on these scenarios is not necessary in advance.

The questionnaire is used to prepare our interview. We therefore ask you to send us the completed questionnaire 2 working days before the interview date. If you are unable to complete and send us both Parts I and II before the interview due to time constraints, please give priority to Part II of the questionnaire (quality criteria). You are then welcome to complete Part I after the interview and send it to us.

Interview procedure

In the first part of the interview, we will address specific quality criteria and your evaluation of relevance and effort based on your feedback on Part II of the questionnaire. If you have submitted Part I of the questionnaire in advance, we will address individual points mentioned there, if necessary.

In the second part of the interview, we will discuss with you your evaluation of the suitability of the registry for answering these research questions on the basis of the benefit assessment scenarios mentioned above, as well as the specific procedure in this regard.

No written minutes of the interview will be prepared (e.g. as a progress or results minutes). However, we will record the interview in order to ensure the comprehensibility of the content of the interview. This recording is only required for internal purposes. It will not be published and the recording will not be used outside of the work on the commission.

Depiction of the contents of the questionnaire and the interview in the report

In our report, we will identify the registries examined and the discussion partners. In the report we will not include information on the individual registries that are not publicly accessible, nor will we include verbatim quotes from the interviews.

We will document the feedback on Part II of the questionnaire, including the corresponding exchange in the interview in aggregated form in the report (e.g. information on which quality criteria are most frequently considered particularly relevant by the respondents, which quality criteria are considered to be particularly associated with great effort by the respondents and why, etc.).

Part I: Information on your registry

1) Which patient group(s) is/are included in your registry?
2) Since which year does the patient registry exist?
3) What are the main objectives or research questions of your patient registry?
4) Who is the owner of the patient registry?
5) How is the patient registry funded?
6) Which health care institutions (e.g. acute-care hospitals, rehabilitation clinics, and office-based medical practices) are included in the patient registry for data collection?
7) What was the initial number of data collection centres involved and what is this number currently?
8) How are data reported from the data collection centres to the patient registry?
9) Is there feedback to the data collection centres involved and in what form?
10) What personnel infrastructure does the patient registry have (full-time staff, qualifications)?
11) What kind of data are collected in the patient registry (multiple answers possible)?
    - ☐sociodemographic data
    - ☐ anthropometric data
    - ☐ clinical data
    - ☐ genetic data
    - ☐ patient-reported outcomes (PROs) / health-related quality of life
    - ☐ routine practice data
    - ☐ laboratory parameters
12) Approximately how many variables do you collect per patient and data collection time point?
13) How often or on what occasions do you collect data from the registry patients?
14) What main measures do you use to ensure the quality of the data collected in your patient registry?
15) Is there a written description of the registry or a corresponding protocol for the patient registry?
16) Have you already had experience with registry studies to assess medical interventions?
    a) If so, have you also performed comparative assessments of interventions or outcome studies?

b) If so, have they included randomized comparative assessments of interventions?

c) If so, have you also conducted such studies in national or international cooperation with other registry operators?

d) If so, please name the most important projects from your point of view, including publications (if available).

17) We would be grateful if you could provide us with references to written, publicly available information on your patient registry, e.g. on the registry's website or in the literature.

Part II: General quality criteria for registries

By ticking [x] of the quality criteria listed in the table, please name the criteria to which the characteristic listed in columns a. to d. applies from your point of view (great relevance, implementation difficulties, easy implementation, great effort). Please name only 5 (possibly different) criteria per column a. to d.

In addition, with regard to your own patient registry, we would like to ask you for a rough evaluation of the degree of fulfilment of all 29 criteria, on a four-level scale from 1 = not fulfilled to 4 = fully fulfilled (last column).

No.	Quality criterion	Great relevance	Implementation difficulties in practice	Easy implementability in practice	Great effort (time, resources)	Degree of fulfilment in your own registry 1 = not fulfilled to 4 = fully filled
Systematics						
1	Detailed registry description (protocol)					
Standardization						
2	Precise definition / operationalization of exposures, clinical events, outcomes and confounders					
3	Current data plan / coding manual					
4	Use of standard classifications (e.g. ICD-10) and terminology (e.g. MedDRA)					
5	Use of validated standard data collection tools (questionnaire, scales)					
6	Training courses on data collection and recording					
7	Implementation of a consensual disease-specific core data set					
8	Use of exact dates for the patient (e.g. birth, death, pregnancy)					

No.	Quality criterion	Great relevance	Implementation difficulties in practice	Easy implementability in practice	Great effort (time, resources)	Degree of fulfilment in your own registry 1 = not fulfilled to 4 = fully filled
9	Use of exact dates of disease (e.g. definitive diagnosis, clinically relevant events)					
10	Use of exact dates for important examinations					
11	Use of exact dates for treatments / interventions (e.g. for drugs: start / stop date, dose, dose changes)					
Validity of sample size generation						
12	Clearly defined inclusion and exclusion criteria for registry patients					
13	Completeness of registry patients					
14	Strategies to avoid selection bias in patient inclusion					
Validity of data collection						
15	Completeness of data					
16	Completeness of data collection time points (drop-outs)					
17	Registry monitoring through internal audits					
18	Registry monitoring through external audits					
19	QM system through routine recording of quality indicators					
20	SOPs (standard operating procedures) for data collection					
21	Source data verification (e.g. for 10% of randomly chosen patients per data collection period)					

No.	Quality criterion	Great relevance	Implementation difficulties in practice	Easy implementability in practice	Great effort (time, resources)	Degree of fulfilment in your own registry 1 = not fulfilled to 4 = fully filled
22	Collection and handling of safety-relevant data (adverse events) as for post-authorization safety studies (PASS)					
Validity of statistical analyses and reports						
23	Prespecification of analysis methods					
24	Description of handling of missing data					
25	Full report of results on all variables recorded					
26	Adjustment of results for potentially confounding or effect-modifying variables					
Superordinate quality criteria						
27	Transparency of registry (funding, decision paths, conflicts of interest)					
28	Scientific independence					
29	Use of existing spontaneous reporting systems in the event of suspected adverse events / adverse drug reactions					

MedDRA: Medical Dictionary for Regulatory Activities; ICD: International Statistical Classification of Diseases and Related Health Problems; QM: quality management; SOP: standard operating procedure

If important quality criteria are missing in the tabular list, please name them in the interview.

Part III: Scenario for the benefit assessment of drugs

Scenario 1

- Disease: malignant skin changes without spontaneous remission
- Intervention: new immunotherapy
- Appropriate comparator therapy of the G-BA: no therapy options so far except best supportive care (BSC)
- Approval, among others, with a single-arm study with the following main results:
  - response in 30% of patients, remission in 5% of patients; measured with imaging techniques, no assessment of relevance from the patient's perspective
  - serious adverse events in 30% of patients
- Result of the benefit assessment: non-quantifiable added benefit, justification for "non-quantifiable":
  - patient relevance of response unclear
  - Weighing of serious adverse events compared with the comparator therapy (BSC) is not possible because no data are available for BSC and the observed serious adverse events may also be caused by the disease itself (and not by the treatment)

Scenario 2

- Disease: chronic lung disease
- Intervention: inhaled double-combination with a new and a known drug
- Comparator therapy: inhaled double-combination with known drugs
- Approval, among others, with a direct comparative RCT with the following main results:
  - improvement in symptoms in 30% with new drug vs. 20% with comparator therapy
  - data on health-related quality of life not interpretable due to large amount of missing values
  - no relevant difference for adverse events
- Result of the benefit assessment: non-quantifiable added benefit, justification for "non-quantifiable":
  - escalation of the comparator therapy prior to study initiation not in accordance with German guidelines (study location: North America and Asia), comparator therapy underdosed in the RCT performed; effect size for symptom improvement and for adverse events therefore cannot be conclusively evaluated
  - data on health-related quality of life missing

Appendix B – Aspects of discussion for the workshop with statistical experts ("statistical workshop")

Background

IQWiG has been commissioned by the Federal Joint Committee (G-BA) to develop scientific concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs. For your information, you can find the commission and the details of the G-BA's specifications here (in German):

- https://www.g-ba.de/downloads/39-261-3773/2019-05-02_IQWIG-Beauftragung_Konzepte-Daten-Nutzenbewertung.pdf
- https://www.g-ba.de/downloads/40-268-5720/2019-05-02_IQWIG-Beauftragung_Konzepte-Daten-Nutzenbewertung_Auftragskonkretisierung.pdf

One element of the concept will be the collection of data in registries. The commission includes the specification of quality criteria and methodological requirements for the data collected in the context of the respective data generation, as well as the description of measures to ensure data quality. For these aspects, we would also like to benefit from your experience in the operation of registries. In this context, the commission, and thus also our interview, is limited to data collection and analysis for the purpose of benefit assessments of drugs.

Aspects of discussion for the statistical workshop

We would like to discuss the following points with you on 19 July 2019:

General issues:

- Which type of registries in Germany do you consider suitable in principle for the purpose of the benefit assessment of drugs and which type do you not consider suitable?
- What are the requirements for data management (data import/export, access, authorizations, etc.)?
- What requirements apply with regard to data corrections (initiation, monitoring, logging)?
- What are the requirements for the documentation of the data structure (coding, categories, scale etc.), also with regard to adjustments over time?
- What are the specific challenges for the multi-centre organization of registries?

Missing data:

- What about the completeness and comprehensiveness of the basically suitable registers?
- Apart from the aspect of completeness, are there other problems with missing data (e.g. at individual time points)?
- What are the requirements for handling missing data in statistical data analysis?
- Can these requirements be met in practice?

Statistical aspects:

- What are the requirements for handling confounding in statistical data analysis?
- What effects do the application of different methods of confounder adjustment (e.g. propensity scores) have on the population for whom a conclusion can be drawn?
- Are measures needed to consider multiple testing?
- Is there an obligation for sample size planning in statistical analyses of registry data?
- What requirements apply to the consideration of site effects (e.g. if exposure is monitored on an institution-related basis)?

Other:

- What are the requirements for performing registry-based RCTs?
- How would you plan a new registry suitable for the purpose of the benefit assessment of drugs?

These aspects are only a basic framework for the discussion on 19 July 2019. You are welcome to name further important aspects to be discussed.

At the end of the statistical workshop, we would like to go through a case study with you in order to make the collection and analysis of data on the basis of registries for the benefit assessment of drugs a little more specific.

Case study for discussion in the statistical workshop

- Disease: malignant skin changes without spontaneous remission
- Intervention: new immunotherapy
- Appropriate comparator therapy of the G-BA: no therapy options so far except best supportive care (BSC)
- Approval, among others, with a single-arm study with the following main results:
  - response in 30% of patients, remission in 5% of patients; measured with imaging techniques, no assessment of relevance from the patient’s perspective
  - serious adverse events in 30% of patients
- Result of the benefit assessment: non-quantifiable added benefit, justification for “non-quantifiable”:
  - patient relevance of response unclear
  - weighing of serious adverse events compared with the comparator therapy (BSC) is not possible because no data are available for BSC and the observed serious adverse events may also be caused by the disease itself (and not by the treatment)
  - no data on quality of life collected, no data on quality of life under BSC presented

Procedure of the statistical workshop

No meeting minutes of the statistical workshop will be prepared (progress or results minutes). However, to ensure that the content of the discussion is comprehensible, we will record the discussion in the workshop. This recording is only required for internal purposes. It will not be published, and the recording will not be used outside the work on the project.

Presentation of the content of the statistical workshop in the report

In our report, we will summarize the aspects discussed concerning the methodological requirements for statistical data analysis (e.g. information on what methodological requirements were considered particularly relevant by the experts, what methodological requirements can and cannot be met in practice, etc.). We will not include verbatim quotations from the individual experts in the report.

Appendix C – Quality criteria for registries and registry studies in the literature

Table 14: Comparison of quality criteria for registries and registry studies in the literature (multi-page table)

No.	Quality criterion	DNVF e. V. 2010/19[a]	EMA Discussion Paper 2018	AHRQ Manual 2014	REQueST EUnetHTA 2019[b]	CTTI Recom-mendations 2017
Systematics						
1	Detailed registry description (protocol)	X	X	X	X	X
Standardization						
2	Precise definition / operationalization of exposures, clinical events, outcomes and confounders	X	X	X	X	X
3	Current data plan / coding manual	X		X	X	X
4	Use of standard classifications (e.g. ICD-10) and terminology (e.g. MedDRA)	X	X	X	X	X
5	Use of validated standard data collection tools (questionnaire, scales, tests)	X		X		
6	Training courses on data collection and recording	X	X	X	X	X
7	Implementation of a consensual disease-specific core data set	X	X		X	
8	Use of exact dates for the patient (e.g. birth, death, pregnancy)		X			
9	Use of exact dates of disease (e.g. definitive diagnosis, clinically relevant events)		X			X
10	Use of exact dates for important examinations		X			X
11	Use of exact dates for treatments / interventions (e.g. for drugs: start / stop date, dose, dose changes)		X	X		X

Table 14: Comparison of quality criteria for registries and registry studies in the literature (multi-page table)

No.	Quality criterion	DNVF e. V. 2010/19[a]	EMA Discussion Paper 2018	AHRQ Manual 2014	REQueST EUnetHTA 2019[b]	CTTI Recom-mendations 2017
Achievement of the recruitment goal / sample composition						
12	Clearly defined inclusion and exclusion criteria for registry patients	X	X	X	X	X
13	Completeness of registry patients (complete recording or representative sample)	X	X	X	X	X
14	Strategies to avoid selection bias in patient inclusion to achieve representativeness	X	X	X	X	X
Validity of data collection						
15	Completeness of data per time point of data collection	X	X	X	X	X
16	Completeness of data collection time points (loss-to-follow-up, drop-outs)	X	X	X	X	X
17	Accuracy of data	X	X	X	X	X
18	Data consistency over time		X			X
19	Source data verification (e.g. for 10% randomly selected patients per study centre)		X	X		X
20	Registry monitoring by internal audits	X	X	X	X	X
21	Registry monitoring by external audits	X	X	X	X	X
22	QM system (if necessary, with regular collection of quality indicators)	X	X	X	X	X
23	SOPs for data collection	X	X	X		X
Superordinate quality criteria						
24	Registry transparency (e.g. funding, decision paths, conflicts of interest)	X	X	X	X	X
25	Scientific independence	X	X			

Table 14: Comparison of quality criteria for registries and registry studies in the literature (multi-page table)

No.	Quality criterion	DNVF e. V. 2010/19[a]	EMA Discussion Paper 2018	AHRQ Manual 2014	REQueST EUnetHTA 2019[b]	CTTI Recom-mendations 2017
26	Secure funding (for planned data collection period)		X		X	
27	Steering committee, executive committee	X	X	X	X	
28	Currency of the registry documents (e.g. protocol, data plan, statistical analysis plan, declaration of consent etc.)				X	
29	Respect of patient rights and data protection, consideration of ethical aspects	X	X	X	X	X
30	Timeliness (currentness and rapid availability of the required results)	X	X			X
31	Flexibility and adaptability (e.g. for embedding studies, for further data collection, in the event of changes in the health care situation)	X	X			X
32	Documentation trail - documentation of all process and definition changes in the registry	X		X		
33	Audit trail - documentation and attributability of all data transactions			X		X
34	Linkability with other data sources	X	X	X		X
Validity of statistical analyses and reports on registry studies						
35	Public registration of the planned registry study					
36	Preparation of a study protocol and a statistical analysis plan for the planned registry study		X			
37	Prespecification of the analysis methods in the statistical analysis plan	X	X	X		X

Table 14: Comparison of quality criteria for registries and registry studies in the literature (multi-page table)

No.	Quality criterion	DNVF e. V. 2010/19[a]	EMA Discussion Paper 2018	AHRQ Manual 2014	REQueST EUnetHTA 2019[b]	CTTI Recom-mendations 2017
38	Explanation of the handling of missing values	X	X	X	X	X
39	Adjudication committee for key outcomes			X		X
40	Adjustment of results of comparisons with regard to potentially confounding variables and consideration of effect modifying variables	X	X	X	X	
41	Sensitivity analyses (e.g. for different case definitions or consideration of confounders)		X	X		
42	Analysis / control of site effects	X				
43	Report on measures to avoid bias (e.g. selection bias, unmeasured confounding)	X	X	X	X	X
44	Full report of the results on all variables recorded	X		X		
45	Publication of the results report including study protocol and analysis plan					
Other possible criteria from a regulatory perspective						
46	Recording and handling of adverse events according to regulatory requirements		X	X		

a: The version of the memorandum on which this analysis was initially based was the publication from 2010; the update 2019, which has been submitted for publication, was also available before the rapid report was completed thanks to the friendly cooperation of DNVF e.V. There have been no relevant changes to the previous version for this comparative analysis.

b: The version of the REQueST tools on which this analysis is based was the version still undergoing the commenting procedure of EUnetHTA in May 2019. There have been no relevant changes to the currently published final version with regard to this comparative analysis.

AE: adverse event; AHRQ: Agency for Healthcare Research and Quality; CTTI: Clinical Trial Transformation; DNVF: German Network for Health Services Analysis (Deutsches Netzwerk Versorgungsforschung); EMA: European Medicines Agency; EUnetHTA: European Network for Health Technology Assessment; ICD: International Statistical Classification of Diseases and Related Health Problems; MedDRA: Medical Dictionary for Regulatory Activities; QM: quality management; REQueST: Registry Evaluation and Quality Standards Tool; SAP: statistical analysis plan; SOP: standard operating procedure

Appendix D – Research question and decisions on orphan drugs with market access 2014 to 2018

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Macitentan	Pulmonary arterial hypertension (PAH)	01.02.2014	17.07.2014	Yes	https://www.g-ba.de/downloads/39-261-2030/2014-07-17_AM-RL-XII_Macitentan_2014-02-01-D-096_BAnz.pdf	https://www.g-ba.de/downloads/40-268-2888/2014-07-17_AM-RL-XII_Macitentan_2014-02-01-D-096_TrG.pdf
Riociguat	Chronic thromboembolic pulmonary hypertension (CTEPH)	01.05.2014	16.10.2014	Yes	https://www.g-ba.de/downloads/39-261-2076/2014-10-16_AM-RL-XII_Riociguat_2014-05-01-D-103_BAnz.pdf	https://www.g-ba.de/downloads/40-268-2978/2014-10-16_AM-RL-XII_Riociguat_2014-05-01-D-103_TrG.pdf
Riociguat	Pulmonary arterial hypertension (PAH)	01.05.2014	16.10.2014	Yes	https://www.g-ba.de/downloads/39-261-2076/2014-10-16_AM-RL-XII_Riociguat_2014-05-01-D-103_BAnz.pdf	https://www.g-ba.de/downloads/40-268-2978/2014-10-16_AM-RL-XII_Riociguat_2014-05-01-D-103_TrG.pdf
Cholic acid	Congenital disorders of the primary bile acid synthesis	15.05.2014	06.11.2014	No	https://www.g-ba.de/downloads/39-261-2088/2014-11-06_AM-RL-XII_Cholsaeure_2014-05-15-D-105_BAnz.pdf	https://www.g-ba.de/downloads/40-268-2995/2014-11-06_AM-RL-XII_Cholsaeure_2014-05-15-D-105_TrG.pdf
Elosulfase alfa	Type IVA mucopoly-saccharidose	01.06.2014	20.11.2014	Yes	https://www.g-ba.de/downloads/39-261-2100/2014-11-20_AM-RL-XII_Elosulfase%20alfa_2014-06-01-D-114_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3008/2014-11-20_AM-RL-XII_Elosulfase%20alfa_2014-06-01-D-114_TrG.pdf
Siltuximab	Multicentric Castleman disease	15.06.2014	04.12.2014	No	https://www.g-ba.de/downloads/39-261-2118/2014-12-04_AM-RL-XII_Siltuximab_2014-06-15-D-119_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3021/2014-12-04_AM-RL-XII_Siltuximab_2014-06-15-D-119_TrG.pdf
Cabozantinib	Thyroid neoplasms	01.08.2014	22.01.2015	Yes	https://www.g-ba.de/downloads/39-261-2147/2015-01-22_AM-RL-XII_Cabozantinib_2014-08-01-D-121_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3084/2015-01-22_AM-RL-XII_Cabozantinib_2014-08-01-D-121_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Obinutuzumab	CLL	15.08.2014	05.02.2015	No	https://www.g-ba.de/downloads/39-261-2167/2015-02-05_AM-RL-XII_Obinutuzumab_2014-08-15-D-120_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3098/2015-02-05_AM-RL-XII_Obinutuzumab_2014-08-15-D-120_TrG.pdf
Ivacaftor	Cystic fibrosis (gating mutations in the CFTR gene)	01.09.2014	19.02.2015	Yes	https://www.g-ba.de/downloads/39-261-2178/2015-02-19_AM-RL-XII_Ivacaftor-nAWG_2014-09-01-D-133_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3114/2015-02-19_AM-RL-XII_Ivacaftor-nAWG_2014-09-01-D-133_TrG.pdf
Teduglutide	Malabsorption syndrome	01.09.2014	19.02.2015	Yes	https://www.g-ba.de/downloads/39-261-2182/2015-02-19_AM-RL-XII_Teduglutid_2014-09-01-D-130_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3118/2015-02-19_AM-RL-XII_Teduglutid_2014-09-01-D-130_TrG.pdf
Alipogenti-parvovec	Hyperlipoprotein-aemia type I	01.11.2014	21.05.2015	No	https://www.g-ba.de/downloads/39-261-2250/2015-05-21_AM-RL-XII_Alipogentiparvovec_2014-11-01-D-138_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3224/2015-05-21_AM-RL-XII_Alipogentiparvovec_2014-11-01-D-138_TrG.pdf
Ibrutinib	CLL (at least 1 previous therapy)	01.11.2014	16.04.2015	No	https://www.g-ba.de/downloads/39-261-2229/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3187/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_TrG.pdf
Ibrutinib	CLL (17p mutation or TP53 mutation, first line)	0 1.11.2014	16.04.2015	No	https://www.g-ba.de/downloads/39-261-2229/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3187/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_TrG.pdf
Ibrutinib	Mantel cell lymphoma	01.11.2014	16.04.2015	No	https://www.g-ba.de/downloads/39-261-2229/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3187/2015-04-16_AM-RL-XII_Ibrutinib_2014-11-01-D-141_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Ataluren	Duchenne muscular dystrophy	01.12.2014	21.05.2015	Yes	https://www.g-ba.de/downloads/39-261-2252/2015-05-21_AM-RL-XII_Ataluren_2012-01-D-149_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3226/2015-05-21_AM-RL-XII_Ataluren_2012-01-D-149_TrG.pdf
Pasireotide	Acromegaly	01.01.2015	18.06.2015	Yes	https://www.g-ba.de/downloads/39-261-2263/2015-06-18_AM-RL-XII_Pasireotid_nAWG_2015-01-01-D-148_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3250/2015-06-18_AM-RL-XII_Pasireotid_nAWG_2015-01-01-D-148_TrG.pdf
Ramucirumab	Stomach cancer (combination with paclitaxel)	01.02.2015	16.07.2015	Yes	https://www.g-ba.de/downloads/39-261-2292/2015-07-16_AM-RL-XII_Ramucirumab_2015-02-01-D-150_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3275/2015-07-16_AM-RL-XII_Ramucirumab_2015-02-01-D-150_TrG.pdf
Ramucirumab	Stomach cancer (monotherapy)	01.02.2015	16.07.2015	Yes	https://www.g-ba.de/downloads/39-261-2292/2015-07-16_AM-RL-XII_Ramucirumab_2015-02-01-D-150_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3275/2015-07-16_AM-RL-XII_Ramucirumab_2015-02-01-D-150_TrG.pdf
Nintendanib	Idiopathic pulmonary fibrosis	15.03.2015	03.09.2015	Yes	https://www.g-ba.de/downloads/39-261-2322/2015-09-03_AM-RL-XII_Nintedanib_nAWG_2015-03-15-D-156_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3339/2015-09-03_AM-RL-XII_Nintedanib_nAWG_2015-03-15-D-156_TrG.pdf
Eliglustat	Gaucher disease type 1	01.04.2015	01.10.2015	No	https://www.g-ba.de/downloads/39-261-2350/2015-10-01_AM-RL-XII_Eliglustat_2015-04-01-D-159_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3381/2015-10-01_AM-RL-XII_Eliglustat_2015-04-01-D-159_TrG.pdf
Olaparib	Ovarian neoplasm	01.06.2015	27.11.2015	No	https://www.g-ba.de/downloads/39-261-2383/2015-11-27_AM-RL-XII_Olaparib_2015-06-01-D-166_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3451/2015-11-27_AM-RL-XII_Olaparib_2015-06-01-D-166_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Lenvatinib	Thyroid neoplasm	01.07.2015	17.12.2015	No	https://www.g-ba.de/downloads/39-261-2413/2015-12-17_AM-RL-XII_Lenvatinib_2015-07-01-D-164_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3495/2015-12-17_AM-RL-XII_Lenvatinib_2015-07-01-D-164_TrG.pdf
Asfotase alfa	Hypophosphatasia (patients ≤ 5 years)	01.10.2015	17.03.2016	No	https://www.g-ba.de/downloads/39-261-2526/2016-03-17_AM-RL-XII_Asfotase-alfa_D-188_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3662/2016-03-17_AM-RL-XII_Asfotase-alfa_D-188_TrG.pdf
Asfotase alfa	Hypophosphatasia (patients > 5 years)	01.10.2015	17.03.2016	No	https://www.g-ba.de/downloads/39-261-2526/2016-03-17_AM-RL-XII_Asfotase-alfa_D-188_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3662/2016-03-17_AM-RL-XII_Asfotase-alfa_D-188_TrG.pdf
Idebenone	Leber's hereditary optic neuropathy	01.10.2015	17.03.2016	No	https://www.g-ba.de/downloads/39-261-2527/2016-03-17_AM-RL-XII_Idebenon_D-191_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3663/2016-03-17_AM-RL-XII_Idebenon_D-191_TrG.pdf
Panobinostat	Multiple myeloma	01.10.2015	17.03.2016	No	https://www.g-ba.de/downloads/39-261-2529/2016-03-17_AM-RL-XII_Panobinostat_D-180_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3666/2016-03-17_AM-RL-XII_Panobinostat_D-180_TrG.pdf
Sebelipase alfa	LAL deficiency (not already rapidly progressing in infancy)	01.10.2015	17.03.2016	No	https://www.g-ba.de/downloads/39-261-2528/2016-03-17_AM-RL-XII_Sebelipase-alfa_D-187_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3664/2016-03-17_AM-RL-XII_Sebelipase-alfa_D-187_TrG.pdf
Sebelipase alfa	LAL deficiency (already rapidly progressing in infancy)	01.10.2015	17.03.2016	No	https://www.g-ba.de/downloads/39-261-2528/2016-03-17_AM-RL-XII_Sebelipase-alfa_D-187_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3664/2016-03-17_AM-RL-XII_Sebelipase-alfa_D-187_TrG.pdf
Isavuconazole	Aspergillosis	15.11.2015	04.05.2016	No	https://www.g-ba.de/downloads/39-261-2574/2016-05-04_AM-RL-XII_Isavuconazol_2015-11-15_D-192_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3754/2016-05-04_AM-RL-XII_Isavuconazol_2015-11-15_D-192_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Isavuconazole	Mucormycosis	15.11.2015	04.05.2016	No	https://www.g-ba.de/downloads/39-261-2574/2016-05-04_AM-RL-XII_Isavuconazol_2015-11-15_D-192_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3754/2016-05-04_AM-RL-XII_Isavuconazol_2015-11-15_D-192_TrG.pdf
Blinatumomab	ALL	15.12.2015	02.06.2016	No	https://www.g-ba.de/downloads/39-261-2605/2016-06-02_AM-RL-XII_Blinatumomab_D-201_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3801/2016-06-02_AM-RL-XII_Blinatumomab_D-201_TrG.pdf
Carfilzomib	Multiple myeloma	15.12.2015	02.06.2016	No	https://www.g-ba.de/downloads/39-261-2606/2016-06-02_AM-RL-XII_Carfilzomib_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3802/2016-06-02_AM-RL-XII_Carfilzomib_TrG.pdf
Ivacaftor	Cystic fibrosis (patients from 18 years of age with R117H mutation)	15.12.2015	02.06.2016	Yes	https://www.g-ba.de/downloads/39-261-2601/2016-06-02_AM-RL-XII_Ivacaftor_nAWG_2015-12-15-D-200_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3797/2016-06-02_AM-RL-XII_Ivacaftor_nAWG_2015-12-15-D-200_TrG.pdf
Ivacaftor	Cystic fibrosis (children 2 to 5 years with gating mutation)	15.12.2015	02.06.2016	No	https://www.g-ba.de/downloads/39-261-2601/2016-06-02_AM-RL-XII_Ivacaftor_nAWG_2015-12-15-D-200_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3797/2016-06-02_AM-RL-XII_Ivacaftor_nAWG_2015-12-15-D-200_TrG.pdf
Afamelanotide	Erythropoietic protoporphyria	15.02.2016	04.08.2016	No	https://www.g-ba.de/downloads/39-261-2674/2016-08-04_AM-RL-XII_Afamelanotid_D-218_BAnz.pdf	https://www.g-ba.de/downloads/40-268-3923/2016-08-04_AM-RL-XII_Afamelanotid_D-218_TrG.pdf
Albutrepeno-nacog alfa	Haemophilia B	01.06.2016	01.12.2016	No	https://www.g-ba.de/downloads/39-261-2776/2016-12-01_AM-RL-XII_Albutrepenonacog_D-227_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4075/2016-12-01_AM-RL-XII_Albutrepenonacog_D-227_TrG.pdf
Daratumumab	Multiple myeloma	01.06.2016	01.12.2016	No	https://www.g-ba.de/downloads/39-261-2772/2016-12-01_AM-RL-XII_Daratumumab_D-238_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4071/2016-12-01_AM-RL-XII_Daratumumab_D-238_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Migalastat	Fabry's disease	01.06.2016	01.12.2016	No	https://www.g-ba.de/downloads/39-261-2777/2016-12-01_AM-RL-XII_Migalastat_D-225_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4076/2016-12-01_AM-RL-XII_Migalastat_D-225_TrG.pdf
Eftrenocog alfa	Haemophilia B	15.06.2016	15.12.2016	No	https://www.g-ba.de/downloads/39-261-2804/2016-12-15_AM-RL-XII_Eftrenonacog-alfa_D-233_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4108/2016-12-15_AM-RL-XII_Eftrenonacog-alfa_D-233_TrG.pdf
Obinutuzumab	Follicular non-Hodgkin lymphoma	01.07.2016	15.12.2016	No	https://www.g-ba.de/downloads/39-261-2801/2016-12-15_AM-RL-XII_Obinutuzumab_D-229_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4105/2016-12-15_AM-RL-XII_Obinutuzumab_D-229_TrG.pdf
Brentuximab vedotin	Hodgkin lymphoma	01.08.2016	19.01.2017	No	https://www.g-ba.de/downloads/39-261-2835/2017-01-19_AM-RL-XII_Brentuximab-Vedotin_D-253_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4156/2017-01-19_AM-RL-XII_Brentuximab-Vedotin_D-253_TrG.pdf
Carfilzomib	Multiple myeloma	01.08.2016	19.01.2017	Yes	https://www.g-ba.de/downloads/39-261-2836/2017-01-19_AM-RL-XII_Carfilzomib_nAWG-Kombi-Dexamethason_D-255_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4157/2017-01-19_AM-RL-XII_Carfilzomib_nAWG-Kombi-Dexamethason_D-255_TrG.pdf
Pitolisant	Narcolepsy	01.08.2016	19.01.2017	No	https://www.g-ba.de/downloads/39-261-2838/2017-01-19_AM-RL-XII_Pitolisant_D-250_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4159/2017-01-19_AM-RL-XII_Pitolisant_D-250_TrG.pdf
Tasimelton	Insomnia	01.08.2016	19.01.2017	No	https://www.g-ba.de/downloads/39-261-2837/2017-01-19_AM-RL-XII_Tasimelteon_D-242_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4158/2017-01-19_AM-RL-XII_Tasimelteon_D-242_TrG.pdf
Teduglutide	Malabsorption syndrome	01.08.2016	19.01.2017	No	https://www.g-ba.de/downloads/39-261-2839/2017-01-19_AM-RL-XII_Teduglutid_D-254_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4160/2017-01-19_AM-RL-XII_Teduglutid_D-254_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Olaratumab	Sarcoma	01.12.2016	18.05.2017	Yes	https://www.g-ba.de/downloads/39-261-2949/2017-05-18_AM-RL-XII_Olaratumab_D-265_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4391/2017-05-18_AM-RL-XII_Olaratumab_D-265-TrG.pdf
Venetoclax	CLL (17p deletion or TP53 mutation)	01.01.2017	15.06.2017	No	https://www.g-ba.de/downloads/39-261-2970/2017-06-15_AM-RL-XII_Venetoclax_D-266_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4425/2017-06-15_AM-RL-XII_Venetoclax_D-266_TrG.pdf
Venetoclax	CLL (without 17p deletion or TP53 mutation)	01.01.2017	15.06.2017	No	https://www.g-ba.de/downloads/39-261-2970/2017-06-15_AM-RL-XII_Venetoclax_D-266_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4425/2017-06-15_AM-RL-XII_Venetoclax_D-266_TrG.pdf
Ixazomib	Multiple myeloma	15.01.2017	06.07.2017	No	https://www.g-ba.de/downloads/39-261-2994/2017-07-06_AM-RL-XII_Ixazomib_D-272_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4454/2017-07-06_AM-RL-XII_Ixazomib_D-272_TrG.pdf
Obeticholic acid	Biliary cirrhosis	15.01.2017	06.07.2017	No	https://www.g-ba.de/downloads/39-261-2995/2017-07-06_AM-RL-XII_Obeticholsaeure_D-269_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4455/2017-07-06_AM-RL-XII_Obeticholsaeure_D-269_TrG.pdf
Cerliponase alfa	Neuronal ceroid lipofuscinosis	01.07.2017	21.12.2017	No	https://www.g-ba.de/downloads/39-261-3168/2017-12-21_AM-RL-XII_Cerliponase-alfa_D-298_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4721/2017-12-21_AM-RL-XII_Cerliponase-alfa_D-298_TrG.pdf
Nusinersen	Spinal muscular atrophy type 2	01.07.2017	21.12.2017	Yes	https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4722/2017-12-21_AM-RL-XII_Nusinersen_D-294_TrG.pdf
Nusinersen	Spinal muscular atrophy type 1	01.07.2017	21.12.2017	Yes	https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4722/2017-12-21_AM-RL-XII_Nusinersen_D-294_TrG.pdf
Nusinersen	Spinal muscular atrophy type 3	01.07.2017	21.12.2017	No	https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4722/2017-12-21_AM-RL-XII_Nusinersen_D-294_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Nusinersen	Spinal muscular atrophy type 4	01.07.2017	21.12.2017	No	https://www.g-ba.de/downloads/39-261-3169/2017-12-21_AM-RL-XII_Nusinersen_D-294_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4722/2017-12-21_AM-RL-XII_Nusinersen_D-294_TrG.pdf
Inotozumab ozogamicin	ALL	15.07.2017	18.01.2018	Yes	https://www.g-ba.de/downloads/39-261-3192/2018-01-18_AM-RL-XII_Inotuzumab-Ozogamicin_D-297_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4763/2018-01-18_AM-RL-XII_Inotuzumab-Ozogamicin_D-297_TrG.pdf
Avelumab	Merkel cell carcinoma (not pretreated)	01.10.2017	16.03.2018	No	https://www.g-ba.de/downloads/39-261-3252/2018-03-16_AM-RL-XII_Avelumab_D-308_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4873/2018-03-16_AM-RL-XII_Avelumab_D-308_TrG.pdf
Avelumab	Merkel cell carcinoma (pretreated)	01.10.2017	16.03.2018	No	https://www.g-ba.de/downloads/39-261-3252/2018-03-16_AM-RL-XII_Avelumab_D-308_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4873/2018-03-16_AM-RL-XII_Avelumab_D-308_TrG.pdf
Midostaurin	AML	15.10.2017	05.04.2018	Yes	https://www.g-ba.de/downloads/39-261-3277/2018-04-05_AM-RL-XII_Midostaurin_AML_D-319_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4916/2018-04-05_AM-RL-XII_Midostaurin_AML_D-319_TrG.pdf
Midostaurin	Mastocytosis	15.10.2017	05.04.2018	No	https://www.g-ba.de/downloads/39-261-3279/2018-04-05_AM-RL-XII_Midostaurin_ASM_D-319_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4918/2018-04-05_AM-RL-XII_Midostaurin_ASM_D-319_TrG.pdf
Obinutuzumab	Follicular lymphoma	15.10.2017	05.04.2018	No	https://www.g-ba.de/downloads/39-261-3275/2018-04-05_AM-RL-XII_Obinutuzumab_nAWG_D-305_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4914/2018-04-05_AM-RL-XII_Obinutuzumab_nAWG_D-305_TrG.pdf
Telotristat ethyl	Carcinoid syndrome	15.10.2017	05.04.2018	No	https://www.g-ba.de/downloads/39-261-3276/2018-04-05_AM-RL-XII_Telotristatethyl_D-318_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4915/2018-04-05_AM-RL-XII_Telotristatethyl_D-318_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Cenegermin	Neurotrophic keratitis	15.11.2017	03.05.2018	No	https://www.g-ba.de/downloads/39-261-3305/2018-05-03_AM-RL-XII_Cenegermin-D-329_BAnz.pdf	https://www.g-ba.de/downloads/40-268-4967/2018-05-03_AM-RL-XII_Cenegermin-D-329_TrG.pdf
Niraparib	Ovarian cancer	15.12.2017	07.06.2018	No	https://www.g-ba.de/downloads/39-261-3361/2018-06-07_AM-RL-XII_Niraparib_D-331_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5050/2018-06-07_AM-RL-XII_Niraparib_D-331_TrG.pdf
Allogenic, genetically modified T cells	Graft-versus-host disease	15.01.2018	05.07.2018	No	https://www.g-ba.de/downloads/39-261-3396/2018-07-05_AM-RL-XII_allogene_genetisch_modifizierte_T-Zellen_D-333_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5102/2018-07-05_AM-RL-XII_allogene_genetisch_modifizierte_T-Zellen_D-333_TrG.pdf
Brentuximab vedotin	T-cell lymphoma	15.01.2018	05.07.2018	Yes	https://www.g-ba.de/downloads/39-261-3397/2018-07-05_AM-RL-XII_Brentuximab-Vedotin_D-340_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5103/2018-07-05_AM-RL-XII_Brentuximab-Vedotin_D-340_TrG.pdf
Letermovir	Prophylaxis of CMV reactivation	15.02.2018	02.08.2018	No	https://www.g-ba.de/downloads/39-261-3431/2018-08-02_AM-RL-XII_Letermovir_D-342_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5164/2018-08-02_AM-RL-XII_Letermovir_D-342_TrG.pdf
Glycerol phenylbutyrate	Urea cycle disturbance	01.03.2018	16.08.2018	No	https://www.g-ba.de/downloads/39-261-3449/2018-08-16_AM-RL-XII_Glycerolphenylbutyrat_D-303_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5184/2018-08-16_AM-RL-XII_Glycerolphenylbutyrat_D-303_TrG.pdf
Burosumab	Hypophosphat-aemia	15.04.2018	04.10.2018	No	https://www.g-ba.de/downloads/39-261-3515/2018-10-04_AM-RL-XII_Burosumab_D-349_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5295/2018-10-04_AM-RL-XII_Burosumab_D-349_TrG.pdf
Darvadstrocel	Crohn's disease	01.06.2018	22.11.2018	No	https://www.g-ba.de/downloads/39-261-3571/2018-11-22_AM-RL-XII_Darvadstrocel_D-366_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5411/2018-11-22_AM-RL-XII_Darvadstrocel_D-366_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Velmanase alfa	α-Mannosidosis	01.07.2018	20.12.2018	No	https://www.g-ba.de/downloads/39-261-3617/2018-12-20_AM-RL-XII_Velmanase-alfa_D-365_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5500/2018-12-20_AM-RL-XII_Velmanase-alfa_D-365_TrG.pdf
Gemtuzumab ozogamicin	AML	01.09.2018	21.02.2019	No	https://www.g-ba.de/downloads/39-261-3679/2019-02-21_AM-RL-XII_Gemtuzumab-Ozogamicin_D-380_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5585/2019-02-21_AM-RL-XII_Gemtuzumab-Ozogamicin_D-380_TrG.pdf
Tisagenlecleu-cel	Diffuse large cell B-cell lymphoma	15.09.2018	07.03.2019	No	https://www.g-ba.de/downloads/39-261-3700/2019-03-07_AM-RL-XII_Tisagenlecleucel-DLBCL_D-375_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5622/2019-03-07_AM-RL-XII_Tisagenlecleucel-DLBCL_D-375_TrG.pdf
Tisagenlecleu-cel	Acute B-cell leukaemia	15.09.2018	07.03.2019	No	https://www.g-ba.de/downloads/39-261-3701/2019-03-07_AM-RL-XII_Tisagenlecleucel-ALL_D-376_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5623/2019-03-07_AM-RL-XII_Tisagenlecleucel-ALL_D-376_TrG.pdf
Caplacizumab	Thrombotic-thrombocytopenic purpura	01.10.2018	22.03.2019	No	https://www.g-ba.de/downloads/39-261-3726/2019-03-22_AM-RL-XII_Caplacizumab_D-387_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5655/2019-03-22_AM-RL-XII_Caplacizumab_D-387_TrG.pdf
Daunorubicin / cytarabine	AML	01.10.2018	22.03.2019	Yes	https://www.g-ba.de/downloads/39-261-3716/2019-03-22_AM-RL-XII_Daunorubicin-Cytarabin_D-382_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5645/2019-03-22_AM-RL-XII_Daunorubicin-Cytarabin_D-382_TrG.pdf
Inotersen	Amyloidosis	01.10.2018	22.03.2019	No	https://www.g-ba.de/downloads/39-261-3717/2019-03-22_AM-RL-XII_Inotersen_D-381_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5646/2019-03-22_AM-RL-XII_Inotersen_D-381_TrG.pdf
Metreleptin	Generalized lipodystrophy	01.10.2018	22.03.2019	No	https://www.g-ba.de/downloads/39-261-3712/2019-03-22_AM-RL-XII_Metreleptin_D-385_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5641/2019-03-22_AM-RL-XII_Metreleptin_D-385_TrG.pdf

Table 15: Orphan drug assessments of the G-BA with market access 2014 to 2018 (multi-page table)

Drug	Research question	Start of procedure	Date of decision	Added benefit quantified	G-BA decision (URL) (in German)	Supporting reasons for G-BA decision (URL) (in German)
Metreleptin	Partial lipodystrophy	01.10.2018	22.03.2019	No	https://www.g-ba.de/downloads/39-261-3712/2019-03-22_AM-RL-XII_Metreleptin_D-385_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5641/2019-03-22_AM-RL-XII_Metreleptin_D-385_TrG.pdf
Patisiran	Amyloidosis	01.10.2018	22.03.2019	Yes	https://www.g-ba.de/downloads/39-261-3718/2019-03-22_AM-RL-XII_Patisiran_D-391_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5647/2019-03-22_AM-RL-XII_Patisiran_D-391_TrG.pdf
Vestronidase alfa	Sly syndrome	01.10.2018	22.03.2019	No	https://www.g-ba.de/downloads/39-261-3723/2019-03-22_AM-RL-XII_Vestronidase%20alfa_D-392_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5730/2019-03-22_AM-RL-XII_Vestronidase%20alfa_D-392_TrG.pdf
Axicabtagen-ciloleucel	Diffuse large cell B-cell lymphoma	01.11.2018	02.05.2019	No	https://www.g-ba.de/downloads/39-261-3771/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5741/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_TrG.pdf
Axicabtagen-ciloleucel	Primarily mediastinal B-cell lymphoma	01.11.2018	02.05.2019	No	https://www.g-ba.de/downloads/39-261-3772/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5742/2019-05-02_AM-RL-XII_Axicabtagen-Ciloleucel_D-406_D-416_TrG.pdf
Tezacaftor / ivacaftor	Cystic fibrosis (from 12 years, F508del-homozygous)	01.12.2018	16.05.2019	Yes	https://www.g-ba.de/downloads/39-261-3785/2019-05-16_AM-RL-XII_Tezacaftor-Ivacaftor_D-408_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5745/2019-05-16_AM-RL-XII_Tezacaftor-Ivacaftor_D-408_TrG.pdf
Tezacaftor / ivacaftor	Cystic fibrosis (from 12 years, F508del-heterozygous, further mutations in the CFTR gene)	01.12.2018	16.05.2019	Yes	https://www.g-ba.de/downloads/39-261-3785/2019-05-16_AM-RL-XII_Tezacaftor-Ivacaftor_D-408_BAnz.pdf	https://www.g-ba.de/downloads/40-268-5745/2019-05-16_AM-RL-XII_Tezacaftor-Ivacaftor_D-408_TrG.pdf

ALL: acute lymphoblastic leukaemia; AML: acute myeloid leukaemia; CFTR: cystic fibrosis transmembrane conductance regulator; CLL: chronic lymphoblastic leukaemia; CMV: cytomegalovirus; CTEPH: chronic thromboembolic pulmonary hypertension; G-BA: Gemeinsamer Bundesausschuss (Federal Joint Committee); LAL: lysosomal acid lipase; PAH: pulmonary arterial hypertension; URL: Uniform Resource Locator

The German version is published under
https://www.iqwig.de/en/projects-results/projects/drug-assessment/a19-43-development-of-scientific-concepts-for-the-generation-of-routine-practice-data-and-their-analysis-for-the-benefit-assessment-of-drugs-according-to-35a-social-code-book-v-rapid-report.11901.html