Methods for determining the extent of added benefit of new drugs - empirical evidence from dossier assessments[1]

EXTRACT

Project: GA23-01 Version: 1.1 Status: 15 Nov 2024 DOI: 10.60584/GA23-01_V1.1_en

[1] Translation of Chapters 1 to 7 of the working paper GA 23-01 *Ausmaßmethodik Zusatznutzen – Empirie aus Dossierbewertungen* (Version 1.1; Status: 15 November 2024). 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

Methods for determining the extent of added benefit of new drugs - empirical evidence from dossier assessments

Commissioning agency

Prepared on the basis of IQWiG's general commission

Internal Project No.

GA23-01

DOI-URL

https://doi.org/10.60584/GA23-01_V1.1_en

Address of publisher

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Siegburger Str. 237
50679 Köln
Germany

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

This report was prepared without the involvement of external experts.

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

IQWiG employees

- Ivona Djuric
- Erika Baumbach
- Florina Kerekes
- Katrin Nink
- Guido Skipka

Keywords

Drug Approval, Treatment Outcome, Government Regulation, Benefit Assessment

Executive summary

As part of the general commission, the topic “methods for determining the extent of added benefit of new drugs - empirical evidence from dossier assessments” was addressed.

Background

To determine the added benefit of new drugs based on dossier assessments, the extent of added benefit (EAB) must be quantified in accordance with the German Regulation for Early Benefit Assessment of New Pharmaceuticals (Arzneimittel-Nutzenbewertungsverordnung, AM-NutzenV). The corresponding categories are major, considerable, minor or non-quantifiable added benefit; no added benefit proven, and less benefit. The methods for determining the EAB for relative effect measures (EAB methods) were first published with the dossier assessment on ticagrelor (A11-02) and subsequently adopted in the Institute's General Methods.

The EAB methods are based on assumptions of effect sizes for relative effect measures. The requirements for the effect sizes that lead to the categorization of an added benefit as "major", “considerable“ or “minor“ vary depending on the relevance of the outcomes. The outcome categories can be grouped into 3 categories:

1) Mortality
2) Serious (or severe) symptoms (or late complications)

   Serious (or severe) adverse effects

   Health-related quality of life (HrQoL)
3) Non-serious (or non-severe) symptoms (or late complications)

   Non-serious (or non-severe) adverse effects

For simpler presentation and better readability, the following shortened designations are used for the outcome categories in this working paper:

1) Mortality
2) Serious/severe/HrQoL
3) Non-serious/non-severe

For each outcome category, the categorization of a result into the different EAB categories (major, considerable, minor) was based in each case on a desired actual effect. Using a hypothetical sample size estimate, a threshold was calculated based on this desired effect. In the sense of a shifted hypothesis boundary, a 95% confidence interval (CI) must fall below this threshold for the result to be categorized as a major, considerable or minor added benefit.

Since the introduction of the German Act on the Reform of the Market for Medicinal Products (Arzneimittelmarktneuordnungsgesetz, AMNOG) and the dossier assessment on ticagrelor (A11-02), a total of 679 dossier assessments were completed by April 2023 and - if suitable data were available for the assessment - the respective EAB at the outcome level was determined according to the aforementioned methods. Based on these results, it can be described whether and how far the actually available effect estimates in the various outcome categories achieve the desired effect sizes on which the EAB methods were based.

Research question

Based on the results of completed dossier assessments, the aim of the present analysis is to empirically examine, for the various outcome categories, how far the desired effect sizes described in the development of the EAB methods in 2011 are actually achieved when these methods are applied. For this purpose, the actual effect estimates from the dossier assessments are compared with the desired effect sizes specified for threshold determination.

In the comparison with the actual effect estimates, the desired effect sizes specified in the development of the EAB methods are referred to below as “expected effect sizes”.

Methods

For this project, all dossier assessments from 2011 up to and including 2022 were used and the relevant information extracted. In addition to basic information (e.g. commission number, drug, indication and research questions), this includes in particular the information on the effect estimates. The effect estimates of the statistically significant effects that were considered in the EAB determination were extracted and analysed. For this purpose, the statistically significant effects from the respective table on the presentation of the EAB at the outcome level (EAB table), which is available in the dossier assessments, were used. The relative risk (RR), the hazard ratio (HR), the Peto odds ratio (Peto OR) and the rate ratio were considered as relevant effect measures. If there was an effect modification by subgroup characteristics, the results of the corresponding subgroup(s) were extracted from the EAB table. In addition to the information on the effects, information on the outcome category (e.g. HrQoL), the direction of the effect (positive/negative from the perspective of the assessed drug), the EAB, and the probability[2] (hint, indication, proof) were extracted from the EAB table. Additional relevant information, including the size of the population analysed, the type of evidence presented (direct, indirect comparison or meta-analysis) and the name of the studies included to derive the added benefit, was also extracted.

The dossier assessments, including any related addenda, were used for the assessment of the information included in the analysis. In the case of addenda and reassessments after the

[2] “The probability of the added benefit describes the certainty of conclusions on the added benefit.” [1]

assessment deadline or in the case of new scientific findings, the most recent results relevant to the dossier assessment were extracted and included in the analysis.

Information was synthesized and analysed. For this purpose, first the data pool was described by describing the included dossier assessments (e.g. indication, drug). The effect estimates actually available for the individual outcome categories were then compared with the expected effect sizes used in the EAB methods to derive the thresholds.

Results

Results of information retrieval

To answer the present research question, 667 dossier assessments and 266 addenda from the years 2011 up to and including 2022 were available. In 239 dossier assessments including addenda, an EAB determination was available for at least 1 outcome for 277 research questions. A total of 1747 outcomes were extracted.

Of the total of 1747 EAB determinations, 89% (n = 1557) originate from a direct comparison of one study and 10% (n = 174) from meta-analyses with several direct comparative studies. A large proportion of the outcomes with EAB determination (77.5%) originate from dossier assessments on oncological topics.

For all possible combinations of outcomes and EAB categories, EAB determinations are available at the outcome level, so that all combinations can be analysed. More than half of all EAB determinations (52% [n = 911]) referred to outcomes of the outcome category “serious/severe/HrQoL”, with the majority referring to outcomes on adverse effects.

Distribution of the actual effect estimates (all outcomes)

Overall, very different distributions of the actual effect estimates were found depending on outcome and EAB category.

The values of the actual effect estimates in the respective highest EAB categories ("major" for mortality and serious/severe/HrQoL; "considerable" for non-serious/non-severe) are in the median slightly above the expected effect sizes: median effect estimate at 0.54 (expected effect: 0.50) and 0.24 (expected effect: 0.17) for the EAB "major" and at 0.35 (expected effect: 0.33) for the EAB "considerable". This means that more than half of the outcomes for which the highest possible EAB was derived have a smaller effect than specified for the respective EAB in the development of the EAB methods.

In the respective second-highest EAB categories ("considerable" for the outcome categories “mortality” and “serious/severe/HrQoL”; "minor" for the outcome category “non-serious/non-severe”), the actual effect estimates are almost completely or mostly below the expected effect sizes: median effect estimate at 0.70 (expected effect: 0.83) and 0.49

(expected effect: 0.67) for the EAB "considerable" and at 0.58 (expected effect: 0.67) for the EAB "minor". This means that the expected effect sizes are achieved in the vast majority of cases for these EAB categories.

Overall, especially in the highest EAB category ("major" for mortality and serious/severe/HrQoL; "considerable" for non-serious/non-severe), it is shown that the effects are smaller in more than half of the EAB determinations than specified in the development of the EAB methods for the respective EAB category.

Distribution of actual effect estimates - by outcome and EAB category

Outcome category "mortality", EAB category "major"

Approximately 62% of the values of the actual effect estimates are above the expected effect size (0.50), i.e. they show smaller effects than specified in the development of the EAB methods for this EAB category. In addition, the lower CI limit is also above the expected effect size for approx. 29% of these outcomes. This means that the expected effect of 0.50 is not only larger in these cases, but also lies outside the 95% CI of the actual effect estimate.

The values of the actual effect estimates tend to increase with increasing population size (number of analysed patients): The larger the number of analysed patients, the smaller the actual effect estimate in this EAB category. This seems plausible, as the population size has an influence on the precision of the effect estimate of a study. Therefore, in larger populations, in the case of smaller effects the upper limit of the 95% CI more frequently falls below the threshold for an EAB category. This potential correlation between population size and effect size is also shown in the distribution of the actual effect estimates separated by subgroup populations (EAB determination by a subgroup characteristic due to a subgroup effect) and total populations, but only for this outcome category.

Outcome category "serious/severe/HrQoL", EAB category "major"

Approximately 61% of the values of the actual effect estimates for this outcome category are above the expected effect size (0.17), i.e. they show smaller effects than specified in the development of the EAB methods for this EAB category. Likewise, the lower limit of the CI is also above the expected effect size for more than half of these outcomes. The values of the actual effect estimates also tend to increase with increasing population size (number of analysed patients) for the outcomes "serious/severe symptoms" and "HrQoL". There is no recognizable effect dependence on the population size for the outcome "serious/severe adverse effects".

Outcome category "non-serious symptoms or adverse effects", EAB category "considerable"

Approximately 53% of the values of the actual effect estimates are above the expected effect size (0.33), i.e. they show smaller effects than specified in the development of the EAB

methods for this EAB category. Likewise, the lower limit of the CI is also above the expected effect size for 44% of these outcomes. In contrast to the outcome category “mortality” and the outcomes “serious and severe symptoms” and “HrQoL”, it is not shown in this category that the effect becomes smaller with increasing population size.

Conclusion

Based on the results of completed dossier assessments, the aim of the present analysis was to empirically examine, for the various outcome categories, how far the desired effect sizes specified in the development of the EAB methods in 2011 are actually achieved when these methods are applied. For this purpose, the actual effect estimates from the dossier assessments are compared with the desired effect sizes specified for threshold determination.

With a total of 1747 EAB determinations at the outcome level from 667 dossier assessments and 266 addenda from 2011 up to and including 2022, a comprehensive evidence base is available.

The empirically specified actual effect estimates only partially reach the effect sizes expected according to the development of the EAB methods:

- For the highest EAB categories ("major" for the outcome categories "mortality" and "severe/severe/HrQoL", "considerable" for the outcome category "non-serious/non-severe"), more than half of the actual effect estimates are smaller than the expected effect sizes.
- For the other EAB categories ("considerable" and "minor" for the outcome categories "mortality" and "serious/severe/HrQoL", "minor" for the outcome category "non-serious/non-severe"), the effect estimates are almost completely or mostly below the expected effect sizes.
- With the exception of population size, no other factors could be identified for which at least partially different distributions of the actual effect estimates were found.
- A possible adaptation of the EAB methods should focus on the highest EAB categories.

Table of contents

Page

List of tables ix
List of figures x
List of abbreviations xi
1 Background 1
2 Research question 5
3 Course of the project 6
3.1 Specifications and changes during the course of the project 6
4 Methods 7
4.1 Information retrieval 7
4.2 Assessment of information 7
4.3 Information synthesis and analysis 7
5 Results 8
5.1 Results of information retrieval 8
5.2 Distribution of the actual effect estimates (all outcomes) 11
5.3 Detailed analysis of the distribution of actual effect estimates - by outcome and EAB category 14
5.3.1 Outcome category “mortality”, EAB category "major" 14
5.3.2 Outcome category “serious/severe/HrQoL”, EAB category "considerable" 16
5.3.3 Outcome category “non-serious/non-severe”, EAB category "considerable" 19
6 Discussion 22
7 Conclusion 24
References for English extract 25

List of tables

Page

Table 1: Empirical results - distribution of outcomes with EAB determination by type of evidence and topic 9
Table 2: Empirical results - distribution of outcomes with EAB determination according to outcome categories and EAB categories 10
Table 3: Empirical results - distribution of outcomes with EAB determination according to outcomes 10
Table 4: Distribution of the available actual effect estimates (all indications) 12

List of figures

Page

Figure 1: Determination of extent of added benefit – quantitative operationalizations' 3
Figure 2: Result of information retrieval and assessment 8
Figure 3: Distribution of available actual effect estimates by outcome and EAB category – analysis of all outcomes 11
Figure 4: Effect estimates for all 45 EAB determinations for the outcomes of the outcome category “mortality” with the EAB “major" - distribution of effect estimates and related CIs (95% CI) in ascending order 15
Figure 5: Effect estimates for the outcomes of the outcome category “mortality” with the EAB “major" - distribution by population size 16
Figure 6: Effect estimates for all 350 EAB determinations for the outcomes of the outcome category “serious/severe/HrQoL” with the EAB “major" - distribution by effect estimates and the related CIs in ascending order 17
Figure 7: Effect estimates for “serious/severe symptoms” outcomes with the EAB “major" - distribution by population size 18
Figure 8: Effect estimates for HrQoL outcomes with the EAB “major" - distribution by population size 18
Figure 9: Effect estimates for “serious/severe adverse effects” outcomes with the EAB “major" - distribution by population size 19
Figure 10: Effect estimates of all 581 EAB determinations for the outcomes of the outcome category “non-serious/non-severe” with the EAB “considerable" - distribution according to effect estimates and related CIs in ascending order 20
Figure 11: Effect estimates for "non-serious/non-serious symptoms” outcomes with the EAB “considerable" - distribution by population size 21
Figure 12: Effect estimates for "non-serious/non-serious adverse effects” outcomes with the EAB “considerable" - distribution by population size 21

List of abbreviations

Abbreviation	Meaning
AMNOG	Arzneimittelmarktneuordnungsgesetz (Act on the Reform of the Market for Medicinal Products)
CI	confidence interval
EAB	extent of added benefit
EAB methods	methods for determining the extent of added benefit
HR	hazard ratio
HrQoL	health-related quality of life
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care)
RR	relative risk

1 Background

Determination of added benefit in accordance with the German Regulation for Early Benefit Assessment of New Pharmaceuticals

To determine the added benefit of new drugs based on dossier assessments, the extent of added benefit (EAB) must be quantified in accordance with the German Regulation for Early Benefit Assessment of New Pharmaceuticals (Arzneimittel-Nutzenbewertungsverordnung, AM-NutzenV). The corresponding categories are major, considerable, minor or non-quantifiable added benefit; no added benefit proven, and less benefit [1]. The methods for determining the EAB for relative effective measures (EAB methods) were first published with the dossier assessment on ticagrelor (A11-02 [2]) and subsequently adopted in the Institute's General Methods [1].

EAB methods

The EAB methods are based on assumptions of effect sizes for relative effect measures. The requirements for the effect sizes that lead to the categorization of an added benefit as "major", “considerable“ or “minor“ vary depending on the relevance of the outcomes. The outcome categories can be grouped into 3 categories according to their relevance:

1) Mortality
2) Serious (or severe) symptoms (or late complications)

   Serious (or severe) adverse effects

   Health-related quality of life (HrQoL)
3) Non-serious (or non-severe) symptoms (or late complications)

   Non-serious (or non-severe) adverse effects

For simpler presentation and better readability, the following shortened designations are used for the outcome categories in this working paper:

1) Mortality
2) Serious/severe/HrQoL
3) Non-serious/non-severe

For each outcome category, the categorization of a result into the different EAB categories (major, considerable, minor) was based in each case on a desired actual effect. Using a hypothetical sample size estimate, a threshold was calculated based on this desired effect. In the sense of a shifted hypothesis boundary, a 95% confidence interval (CI) must fall below this threshold for the result to be categorized as major, considerable or minor added benefit.

For the outcome “mortality”, an increase in survival time is assessed as "major" if the threshold of 0.85 derived from a desired effect of 0.50 (relative risk, RR) is undercut by the upper limit of the 95% CI. An RR of 0.50 was postulated by Djulbegovic et al. 2008 as a requirement for a "breakthrough" [3]. Based on this, the desired effect sizes for the various aforementioned outcome categories and, derived from this, the corresponding thresholds of the respective 95% CIs for the derivation of the different EAB categories were first specified (see Figure 1 and dossier assessment A11-02) [1,2,4].

		Outcome category			
		Survival time (mortality)	**Serious *(or severe)* symptoms *(or late complications)* and adverse effects**	***Health-related quality of life***	**Non-serious *(or non-severe)* symptoms *(or late complications)* and adverse effects**
Added benefit	**Major** **sustained and great improvement** in the therapy-relevant benefit, which has not previously been achieved versus the appropriate comparator therapy	Major increase in survival time **CI_s: 0.85** (RR_1: 0.50)	Long-term freedom or extensive avoidance **CI_s: 0.75** (RR_1: 0.17) **and risk ≥ 5%[2]**	*Major improvement*[1] **CI_s: 0.75** (RR_1: 0.17) **and risk ≥ 5%[2]**	Not applicable
	Considerable **marked improvement** in the therapy-relevant benefit, which has not previously been achieved versus the appropriate comparator therapy	Moderate increase in survival time **CI_s: 0.95** (RR_1: 0.83)	Alleviation or relevant avoidance **CI_s: 0.90** (RR_1: 0.67)	*Important improvement*[1] **CI_s: 0.90** (RR_1: 0.67)	Important avoidance **CI_s: 0.80** (RR_1: 0.33)
	Minor **moderate and not only marginal improvement** in the therapy-relevant benefit, which has not previously been achieved versus the appropriate comparator therapy	*Any (statistically significant) increase in survival time* **CI_s: 1.00**	*Any (statistically significant) reduction* **CI_s: 1.00**	*Relevant improvement*[1] **CI_s: 1.00**	Relevant avoidance **CI_s: 0.90** (RR_1: 0.67)

Amendments to the ANV in *italics*.
1: The precondition is the use of a validated instrument as well as a validated response criterion. Values apply for non-response.
2: For at least one of the two groups to be compared.
ANV: Arzneimittel-Nutzenbewertungsverordnung (Regulation for Early Benefit Assessment of New Pharmaceuticals); CI_s: threshold for upper limit of the 95% confidence interval; RR_1: actual relative risk.

Figure 1: Determination of extent of added benefit – quantitative operationalizations[3,4]

[3] Table extracted from [2] and translated.

[4] The effect sizes desired for the respective EAB in the development of the EAB methods were originally specified on the basis of the RR. The abbreviation “RR1: actual RR” corresponds to the expected effect sizes described in the working paper.

Dossier assessments since the introduction of the EAB methods

Since the introduction of the German Act on the Reform of the Market for Medicinal Products (Arzneimittelmarktneuordnungsgesetz, AMNOG) and the dossier assessment for ticagrelor (A11-02), a total of 679 dossier assessments were prepared by April 2023 and - if suitable data were available for the assessment - the respective EAB was determined at the outcome level according to the aforementioned EAB methods.

On the basis of these results, it can be described whether and how far the actual effect estimates in the various outcome categories achieve the desired effect sizes on which the EAB methods were based [1,4]. The comparison between desired and actual effects can also provide the starting point for a discussion on the further development of the EAB methods.

2 Research question

Based on the results of completed dossier assessments, the aim of the present analysis is to empirically examine, for the various outcome categories, how far the desired effect sizes specified in the development of the EAB methods in 2011 are actually achieved when these methods are applied. For this purpose, the actual effect estimates from the dossier assessments are compared with the desired effect sizes specified for threshold determination.

In the comparison with the actual effect estimates, the desired effect sizes specified in the development of the EAB methods are referred to below as “expected effect sizes”.

3 Course of the project

As part of the general commission, the topic “methods for determining the extent of added benefit of new drugs - empirical evidence from dossier assessments” was addressed. The work on the project started on 8 May 2023.

A working paper was prepared based on an internal project outline. This report was submitted to the Federal Joint Committee (G-BA) on 21 October 2024 and published on IQWiG's website four weeks later.

3.1 Specifications and changes during the course of the project

Changes compared to version 1.0

The present version 1.1 of 15 November 2024 replaces version 1.0 of the working paper of 21 October 2024. The following changes are included in version 1.1 compared to version 1.0:

- The references to IQWiG's General Methods were updated to the current version 7.0. The reference for Figure 1 was corrected.

The changes have no impact on the conclusion of the working paper.

4 Methods

4.1 Information retrieval

For this project, all dossier assessments from 2011 up to and including 2022 were used and the relevant information extracted. In addition to basic information (e.g. commission number, drug, indication and research questions), this includes in particular the information on the effect estimates. The effect estimates of the statistically significant effects that were considered in the EAB determination were extracted and analysed. For this purpose, the statistically significant effects from the respective table on the presentation of the EAB at the outcome level (EAB table), which is available in the dossier assessments, were used. The RR, the hazard ratio (HR), the Peto odds ratio (Peto OR) and the rate ratio were considered as relevant effect measures. If there was an effect modification by subgroup characteristics, the results of the corresponding subgroup(s) were extracted from the EAB table. In addition to the information on the effects, information on the outcome category (e.g. HrQoL), the direction of the effect (positive/negative from the perspective of the assessed drug), the EAB, and the probability[5] (hint, indication, proof) were extracted from the EAB table. Additional relevant information, including the size of the population analysed, the type of evidence presented (direct, indirect comparison or meta-analysis) and the name of the studies included to derive the added benefit, was also extracted.

Information from dossier assessments from 2011 up to and including 2022 (commissions A11-02 to A22-137, including the related addenda) was extracted and considered in the analysis.

4.2 Assessment of information

The dossier assessments, including any related addenda, were used for the assessment of the information included in the analysis. Data were extracted from these documents in a standardized manner. In the case of addenda and reassessments after the assessment deadline or in the case of new scientific findings, solely the most recent results relevant to the dossier assessment were extracted and included in the analysis.

4.3 Information synthesis and analysis

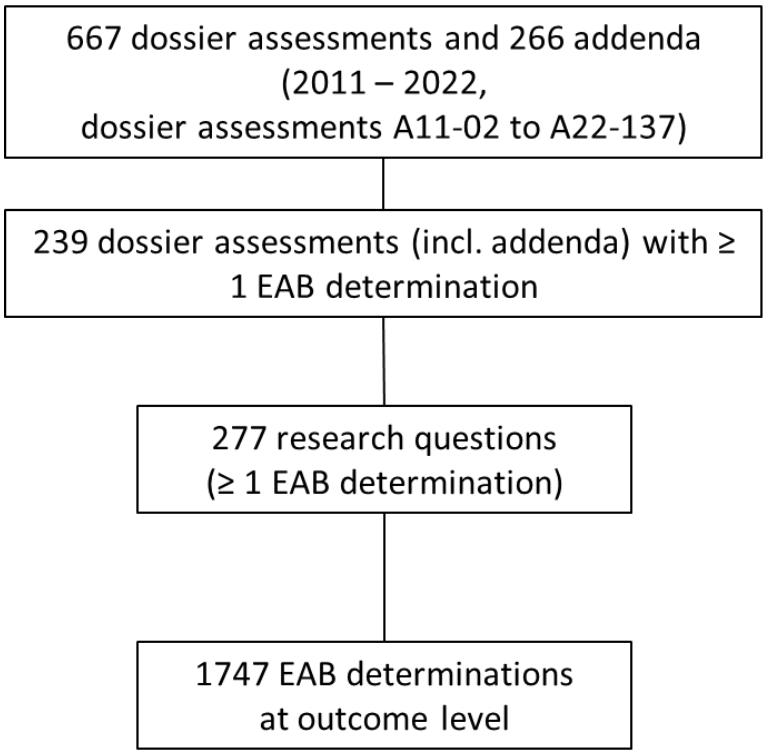
Information was synthesized and analysed. For this purpose, the data pool was first described by describing the included dossier assessments (e.g. indication, drug). The effect estimates actually available for the individual outcome categories were then compared with the expected effect sizes used in the EAB methods to derive the thresholds [1,4].

[5] "The probability of the added benefit describes the certainty of conclusions on the added benefit." [1]

5 Results

5.1 Results of information retrieval

To answer the present research question, 667 dossier assessments and 266 addenda from the years 2011 up to and including 2022 were available. In 239 dossier assessments including addenda, an EAB determination was available for at least 1 outcome for 277 research questions. A total of 1747 outcomes were extracted.



EAB: extent of added benefit.

Figure 2: Result of information retrieval and assessment

Description of the outcomes with EAB determination

Table 1 describes the outcomes with EAB determination by type of evidence and topic.

Table 1: Empirical results - distribution of outcomes with EAB determination by type of evidence and topic

Characteristic **Category**	**Outcomes with EAB determination** **N = 1747**
Type of evidence, n (%)	
Direct comparison	1557 (89.1)
Meta-analysis	174 (10.0)
Indirect comparison ≥ 2 studies[a]	16 (0.9)
Topic, n (%)	
Diabetes	42 (2.4)
Cardiovascular/pulmonary	65 (3.7)
Infectious diseases	77 (4.4)
Oncology	1353 (77.5)
Psychiatry/neurology	48 (2.8)
Others[b]	162 (9.3)

a. Including meta-analyses.
b. Indications that cannot be categorized in the other categories listed.
EAB: extent of added benefit; n: number of outcomes.

Of the total of 1747 EAB determinations, 89% (n = 1557) originate from a direct comparison of one study and 10% (n = 174) from meta-analyses with several direct comparative studies. A large proportion of the outcomes with EAB determination (77.5%) originate from dossier assessments on oncological topics (see Table 1). The remaining outcomes with EAB determination originate from dossier assessments on diabetes (2.4%), cardiovascular/pulmonary (3.7%), infectious diseases (4.4%), psychiatry/neurology (2.8%) and others (9.3%).

Table 2 and Table 3 show the distribution of the outcomes with EAB determination to the various outcome and EAB categories.

Table 2: Empirical results - distribution of outcomes with EAB determination according to outcome categories and EAB categories

Outcome category	EAB category			
	Outcome with EAB "major" n (%[a])	Outcome with EAB "considerable" n (%[a])	Outcome with EAB "minor" n (%[a])	Total n (%[a])
Mortality	45 (2.6)	42 (2.4)	19 (1.1)	106 (6.1)
Serious/severe/HrQoL	350 (20.0)	282 (16.1)	279 (16.0)	911 (52.1)
Non-serious/non-severe	-[b]	581 (33.3)	149 (8.5)	730 (41.8)

a. Percentages refer to the population of all extracted outcomes (N = 1747).
b. According to the General Methods of the Institute [1] no EAB "major" is determined for the category "non-serious/non-severe".
EAB: extent of added benefit; HrQoL: health-related quality of life; n: number of outcomes;

Table 3: Empirical results - distribution of outcomes with EAB determination according to outcomes

Outcome category / Outcome	EAB category		
	Outcome with EAB "major" n (%[a])	Outcome with EAB "considerable" n (%[a])	Outcome with EAB "minor" n (%[a])
Mortality	45 (2.6)	42 (2.4)	19 (1.1)
Morbidity			
Serious (or severe) symptoms (or late complications)	25 (1.4)	55 (3.1)	33 (1.9)
Non-serious (or non-severe) symptoms (or late complications)	-[b]	137 (7.8)	75 (4.3)
Health-related quality of life	40 (2.3)	52 (3.0)	75 (4.3)
Adverse effects			
Serious (or severe) adverse effects	285 (16.3)	175 (10.0)	171 (9.8)
Non-serious (or non-severe) adverse effects	-[b]	444 (25.4)	74 (4.2)

a. Percentages refer to the population of all extracted outcomes (N = 1747).
b. According to the General Methods of the Institute [1] no EAB "major" is determined for the category "non-serious/non-severe".
EAB: extent of added benefit; n: number of outcomes.

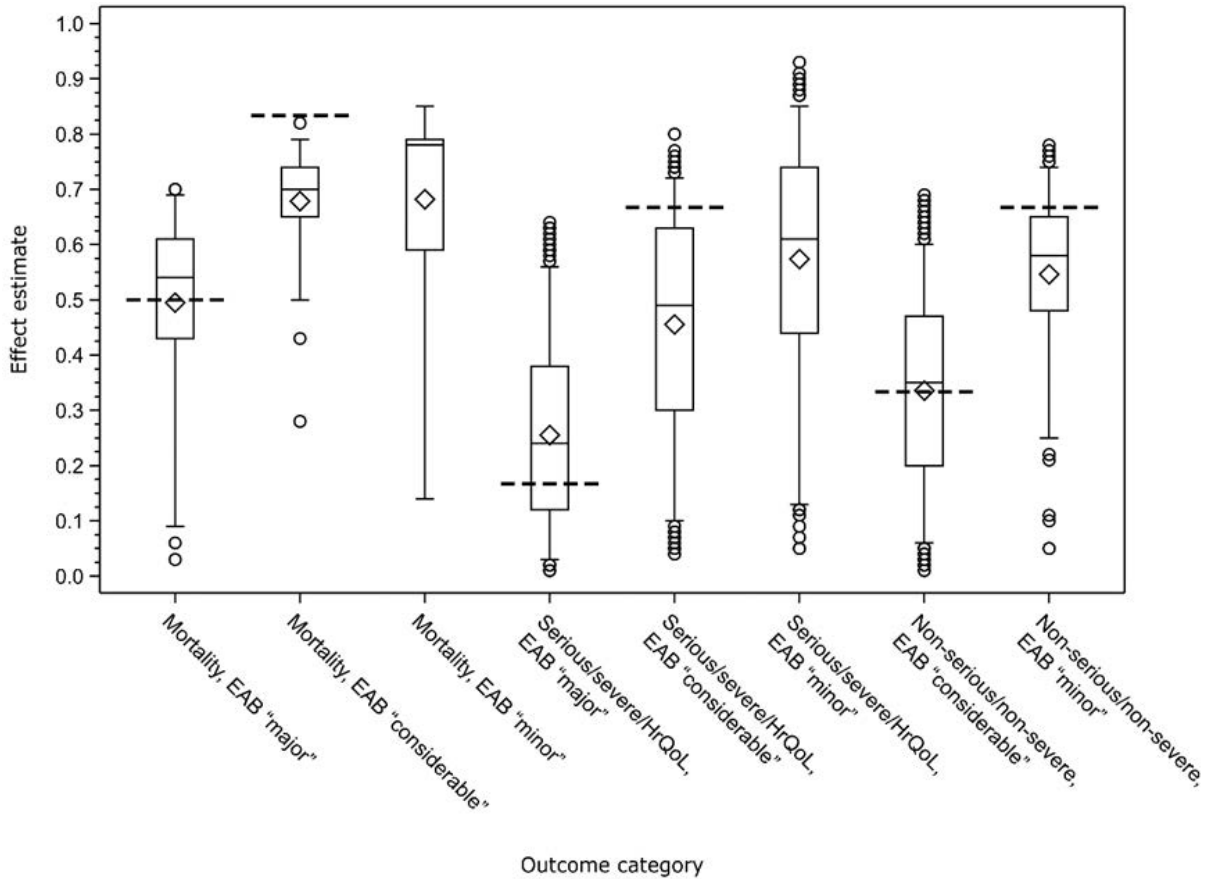
It can be seen from the tables that EAB determinations are available at the outcome level for all possible combinations of outcomes and EAB category, so that all combinations can be evaluated. More than half of all EAB determinations (52% [n = 911]) referred to outcomes of

the outcome category “serious/severe/HrQoL”, with the majority referring to outcomes on adverse effects.

5.2 Distribution of the actual effect estimates (all outcomes)

As described in Chapter 1, different thresholds (upper limit of the 95% CI) were specified based on the expected effect sizes to derive the EAB for the different outcome categories. For the following analyses, the available actual effect estimates from the dossier assessments were compared with the effect sizes expected for threshold determination.

The distribution of the actual effect estimates extracted from the dossier assessments, broken down by outcome and EAB category, is shown in Figure 3 and Table 4.



Explanation of boxplot:
The box corresponds to the range of the 25% to 75% quantile, the solid horizontal line within the box corresponds to the median, the diamond within the box corresponds to the mean value. The lines outside the box correspond to the 5% and 95% quantile. Values outside these quantiles are shown as circles.
The dashed horizontal lines correspond to the expected effect sizes.
EAB: extent of added benefit; HrQoL: health-related quality of life.

Figure 3: Distribution of available actual effect estimates by outcome and EAB category – analysis of all outcomes

Table 4: Distribution of the available actual effect estimates (all indications)

Outcome category	EAB category		
	Outcome with EAB "major"	**Outcome with EAB "considerable"**	**Outcome with EAB "minor"**
Mortality	n = 45	n = 42	n = 19
Expected effect size (RR)[a]	0.50	0.83	–[b]
Actual effect estimate[c]			
Median [Q1; Q3]	0.54 (0.43; 0.61)	0.70 (0.65; 0.74)	0.78 (0.59; 0.79)
Mean value	0.49	0.68	0.68
Serious/severe/HrQoL	n = 350	n = 282	n = 279
Expected effect size (RR)[a]	0.17	0.67	–[b]
Actual effect estimate[c]			
Median [Q1; Q3]	0.24 (0.12; 0.38)	0.49 (0.30; 0.63)	0.61 (0.44; 0.74)
Mean value	0.26	0.46	0.57
Non-serious/non-severe		n = 581	n = 149
Expected effect size (RR)[a]	–[d]	0.33	0.67
Actual effect estimate[c]			
Median [Q1; Q3]	No data	0.35 (0.20; 0.47)	0.58 (0.48; 0.65)
Mean value	No data	0.34	0.55

a. Originally developed on the basis of the RR, but related to all relative effect measures (HR, RR, Peto OR and rate ratio).
b. According to the General Methods of the Institute [1] any statistically significant effect at the usual error level of 5% is categorized as at least a minor added benefit.
c. Includes the effect estimates HR, RR, Peto OR and rate ratio.
d. According to the General Methods of the Institute [1] no EAB "major" is determined for the outcome category "non-serious/non-severe").

EAB: extent of added benefit; HrQoL: health-related quality of life; HR: hazard ratio; n: number of outcomes; OR: odds ratio; Q1: 1st quartile; Q3: 3rd quartile; RR: relative risk

Overall, very different distributions of the actual effect estimates were found depending on the outcome and EAB category (see Figure 3).

Comparison of the different EAB categories: The higher the category, the greater the effect

A comparison of the different EAB categories shows, as expected, that the actual effect estimate increases as the EAB category increases. This is also expressed by a lower value of the median or mean effect estimate in the higher EAB categories (see Table 4).

Effects in the highest EAB category often smaller than specified in the development of the EAB methods

The values of the actual effect estimates in the respective highest EAB categories ("major" for the outcome categories “mortality” and “severe/severe/HrQoL”; "considerable" for the outcome category “non-serious/non-severe”) are in the median slightly above the expected effect sizes: For the outcome categories “mortality” and “serious/severe/HrQoL”, the values of the respective actual median effect estimate are 0.54 (expected effect: 0.50) and 0.24 (expected effect: 0.17) for the EAB "major". For the outcome category “non-serious/non-severe”, the values of the actual median effect estimate are 0.35 (expected effect: 0.33) for the EAB "considerable" (see Figure 3 and Table 4). This means that more than half of the outcomes for which the highest possible EAB was derived have a smaller effect than specified for the respective EAB in the development of the EAB methods.

It can also be seen that the values of the actual effect estimates in the respective second-highest EAB categories ("considerable" for the outcome categories “mortality” and “serious/severe/HrQoL”; "minor" for the outcome category “non-serious/non-severe”) are almost completely or largely below the expected effect sizes: For the outcome categories “mortality” and “serious/severe/HrQoL”, the values of the respective median effect estimate are 0.70 (expected effect: 0.83) and 0.49 (expected effect: 0.67) for the EAB "considerable". For the outcome category “non-serious/non-severe”, the values of the median effect estimate are 0.58 (expected effect: 0.67) for the EAB "minor" (see Figure 3 and Table 4). This means that the expected effect sizes are achieved in the vast majority of outcomes for these outcome categories.

For the two outcome categories “mortality” and “serious/severe/HrQoL”, to achieve a categorization as a minor added benefit, there is no assumption in the sense of a shifted hypothesis boundary; rather, any statistically significant effect at the usual error level of 5% is categorized as "minor".

Overall, especially in the highest EAB category ("major" for mortality and serious/severe/HrQoL; "considerable" for non-serious/non-severe), it can be seen that the actual effect estimates in more than half of the EAB determinations are smaller than specified in the development of the EAB methods.

Investigation of cluster effects

In the present analyses, all effect estimates that led to a EAB determination were considered equally. The fact that effect estimates of outcomes from the same study may be correlated is not taken into account. This could lead to bias in the results. Sensitivity analyses were conducted to investigate whether the so-called cluster effect plays a relevant role. As the majority of dossier assessments are based on results from only one study, only one outcome

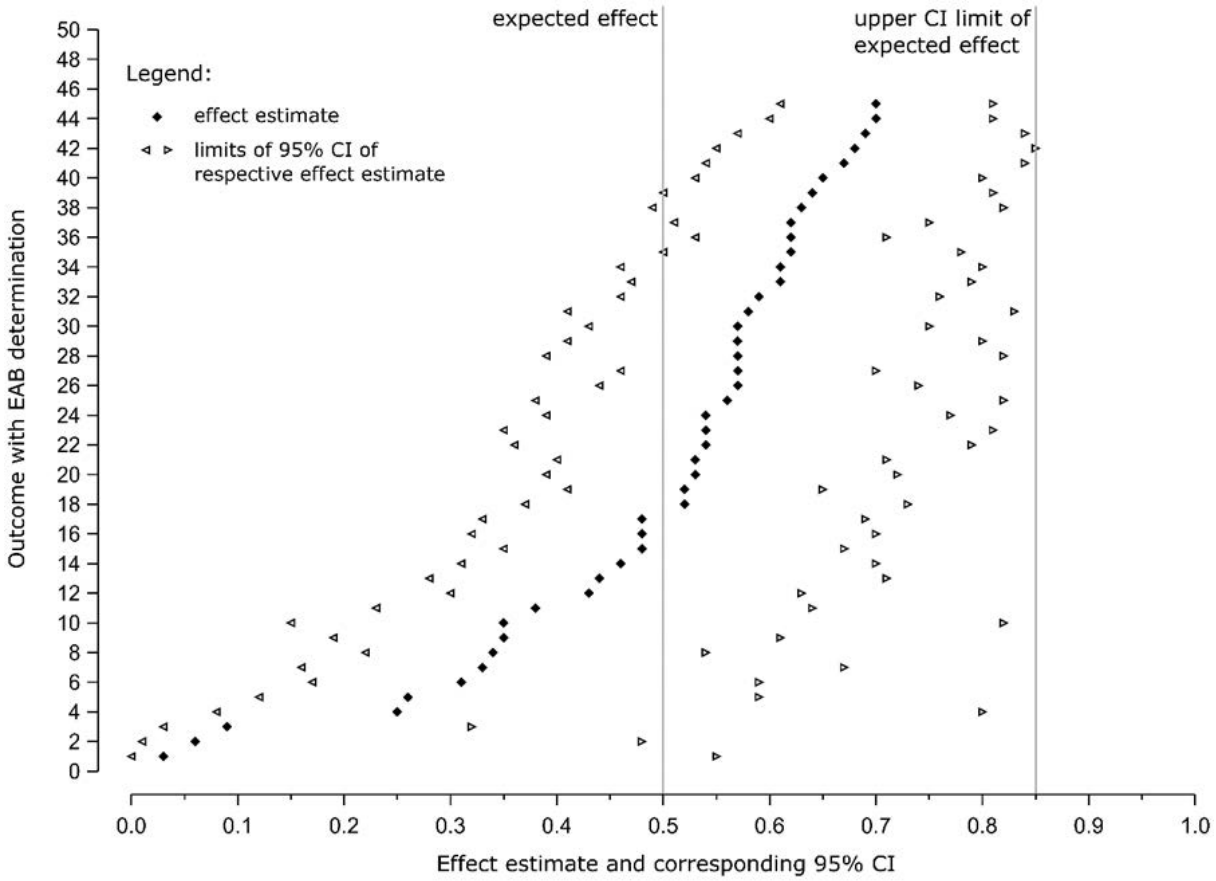
per dossier assessment was randomly selected and evaluated for these analyses. The comparison between the main and sensitivity analyses showed no notable differences. For comparison, the sensitivity analysis for Figure 3 is shown in Appendix A of the full report.

5.3 Detailed analysis of the distribution of actual effect estimates - by outcome and EAB category

Based on the results for the distribution of the actual effect estimates in Figure 3 and Table 4, the results for the outcome and EAB categories “mortality” and “serious/severe/HrQoL” with the EAB "major" as well as “non-serious/non-severe” with the EAB "considerable" were analysed in more detail, as the greatest deviations from the respective expected effect size in terms of smaller effect estimates are present here. In a further step, it was also analysed for these outcome and EAB categories whether the results differ when differentiated according to population size, certainty of conclusions, and topics. The differentiated analysis by population size shows different results for the outcomes, which are described in the following sections 5.3.1, 5.3.2 and 5.3.3. The differentiated analysis according to the certainty of conclusions and topics shows no relevant differences in the distribution of the actual effect estimates (see Appendix B and C of the full report). In this context, it should also be taken into account that in some cases only relatively few EAB determinations are available, such as for the outcome “mortality” in non-oncological topics or for various analyses differentiated according to the certainty of conclusions.

5.3.1 Outcome category “mortality”, EAB category "major"

With a median of 0.54, the values of the actual effect estimates in the outcome category “mortality” with the EAB "major" are above the expected effect size of 0.50 in more than half of the cases . This is also reflected in Figure 4, which shows the effect estimates and related CIs for all outcomes in this outcome category.

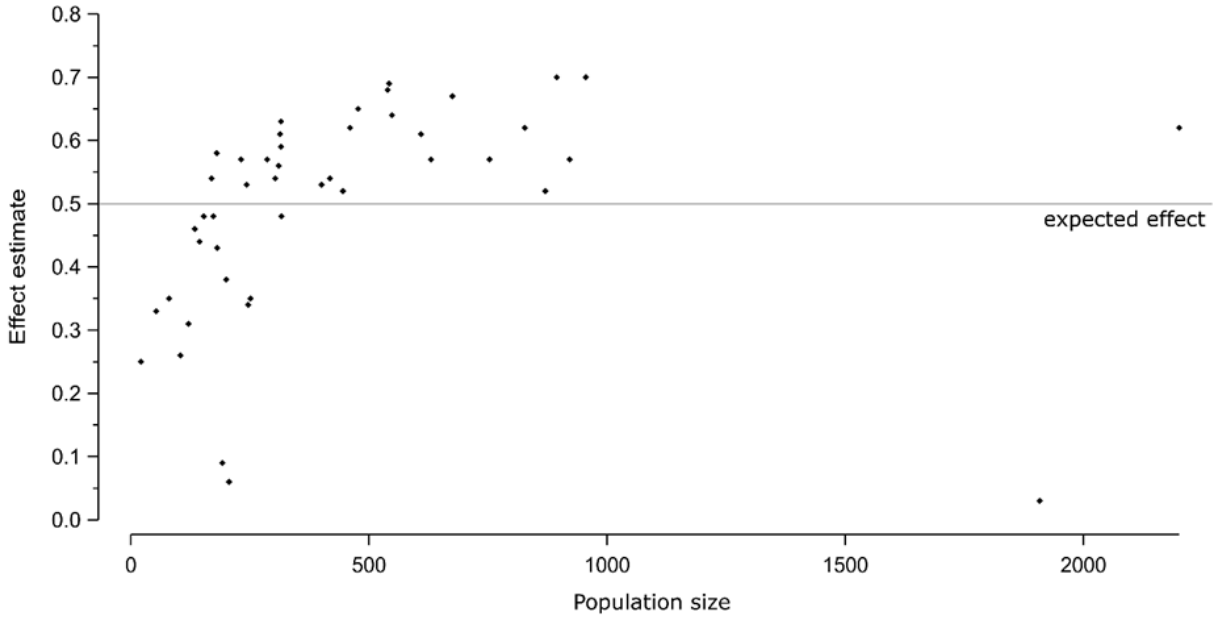


CI: confidence interval; EAB: extent of added benefit.

Figure 4: Effect estimates for all 45 EAB determinations for the outcomes of the outcome category “mortality” with the EAB “major" - distribution of effect estimates and related CIs (95% CI) in ascending order

Around 62% of the values of the actual effect estimates are above the expected effect size (0.50), i.e. they show smaller effects than specified in the development of the EAB methods for this EAB category. In addition, the lower CI limit is also above the expected effect size for approx. 29% of these outcomes. This means that the expected effect of 0.50 is not only larger in these cases, but also lies outside the 95% CI of the actual effect estimate.

Figure 5 shows the distribution of effect estimates depending on population size.



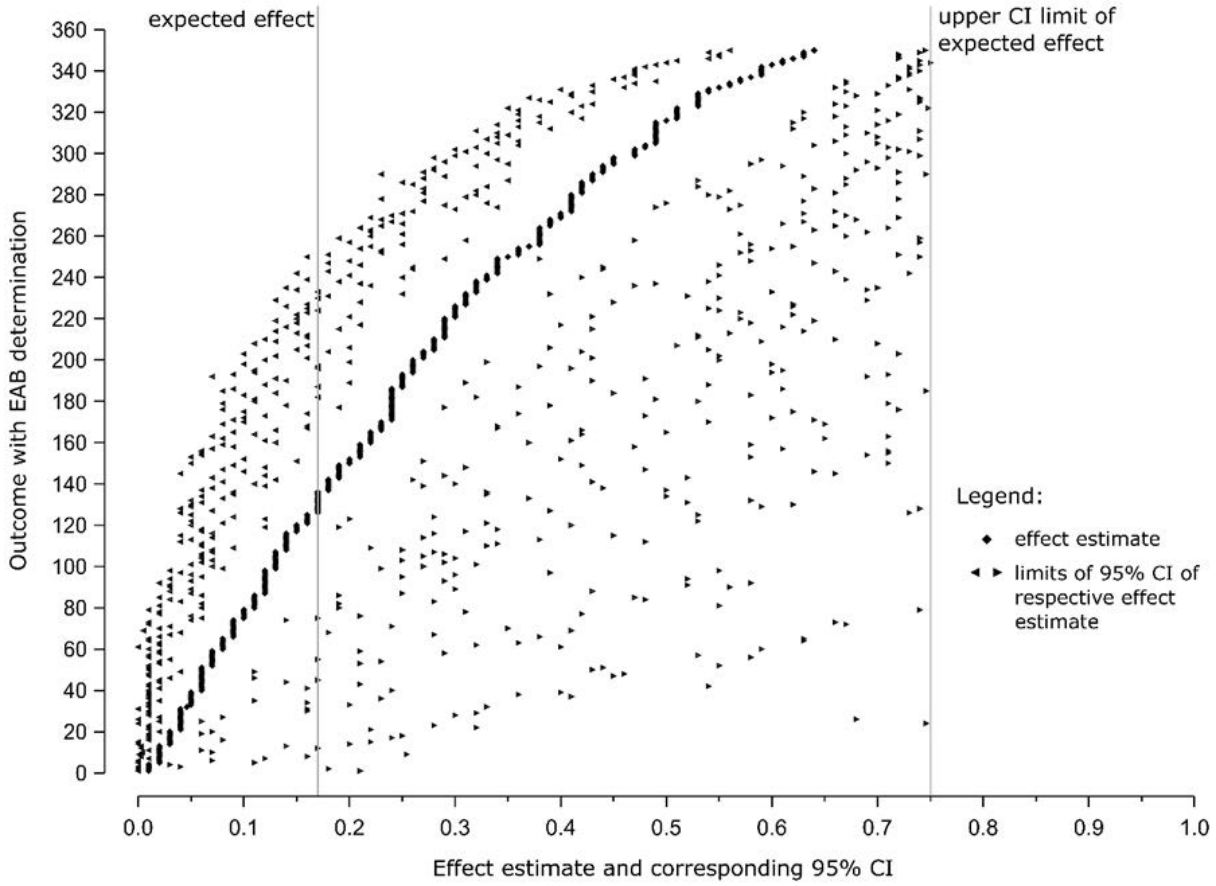
EAB: extent of added benefit.

Figure 5: Effect estimates for the outcomes of the outcome category “mortality” with the EAB “major" - distribution by population size

In Figure 5 it can be seen that the values of the actual effect estimates tend to increase with increasing population size (number of patients analysed): The larger the number of patients analysed, the smaller the actual effect estimate in this EAB category. This seems plausible, as the population size has an influence on the precision of the effect estimate of a study. Therefore, in larger populations, in the case of smaller effects the upper limit of the 95% CI more frequently falls below the threshold for an EAB category. This potential correlation between population size and effect estimate is also shown in the results of the subgroup populations (EAB determination by subgroup characteristic due to subgroup effect) and total populations in the distribution of the actual effect estimates (see Appendix D of the full report), but only for this outcome category.

5.3.2 Outcome category “serious/severe/HrQoL”, EAB category "considerable"

In the outcome category “serious/severe/HrQoL” with the EAB "considerable", with a median of 0.24 the values of the actual effect estimates are above the expected effect size of 0.17 in more than half of the cases. Figure 6 shows the effect estimates and related CIs for all outcomes in this outcome category.

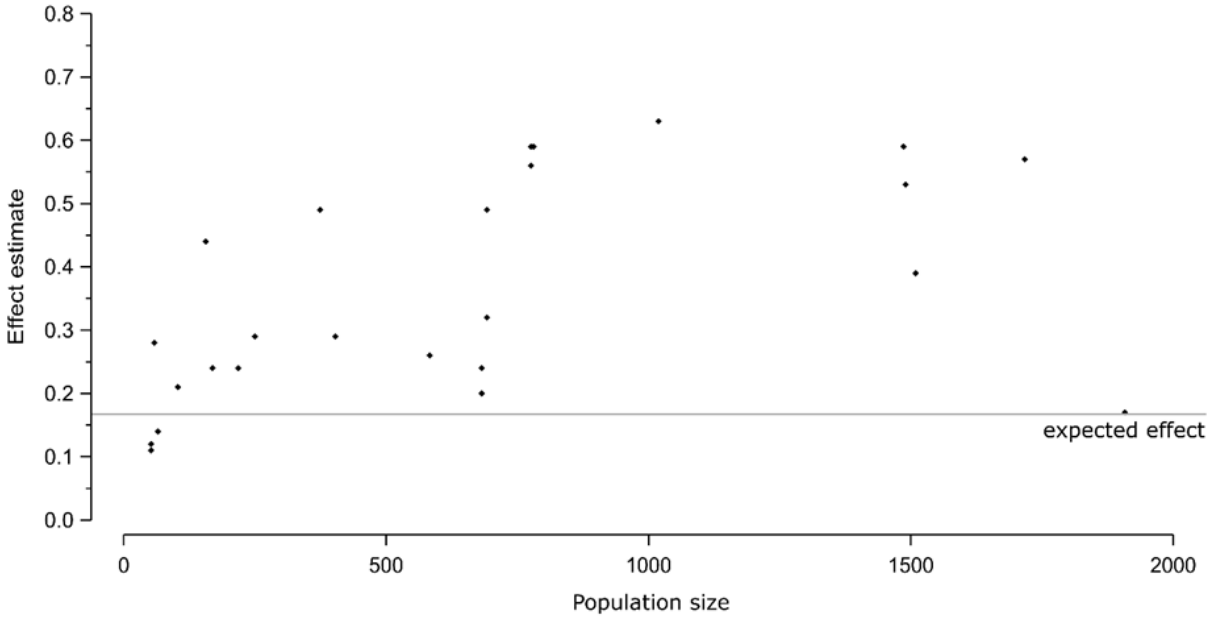


CI: confidence interval; EAB: extent of added benefit, HrQoL: health related quality of life.

Figure 6: Effect estimates for all 350 EAB determinations for the outcomes of the outcome category “serious/severe/HrQoL” with the EAB “major" - distribution by effect estimates and the related CIs in ascending order

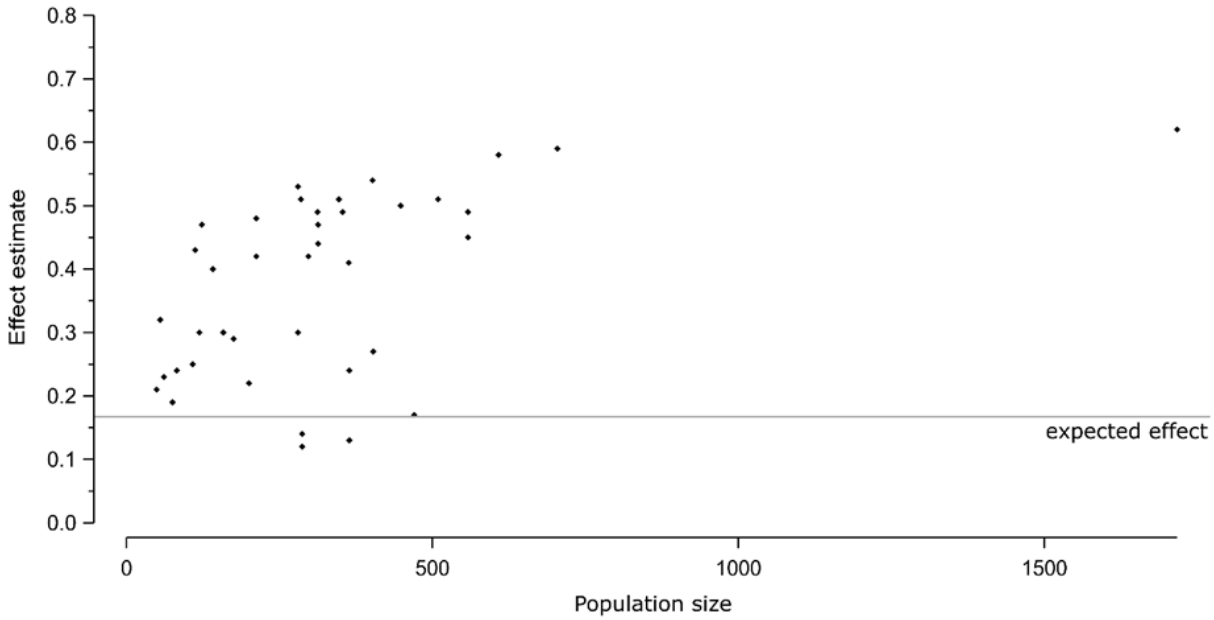
Around 61% of the values of the actual effect estimates for this outcome category are above the expected effect size (0.17), i.e. they show smaller effects than specified in the development of the EAB methods for this EAB category. Likewise, for more than half of these outcomes, the lower limit of the CI is also above the expected effect size.

Figure 7, Figure 8 and Figure 9 show the distribution of effect estimates per outcome depending on the population size.



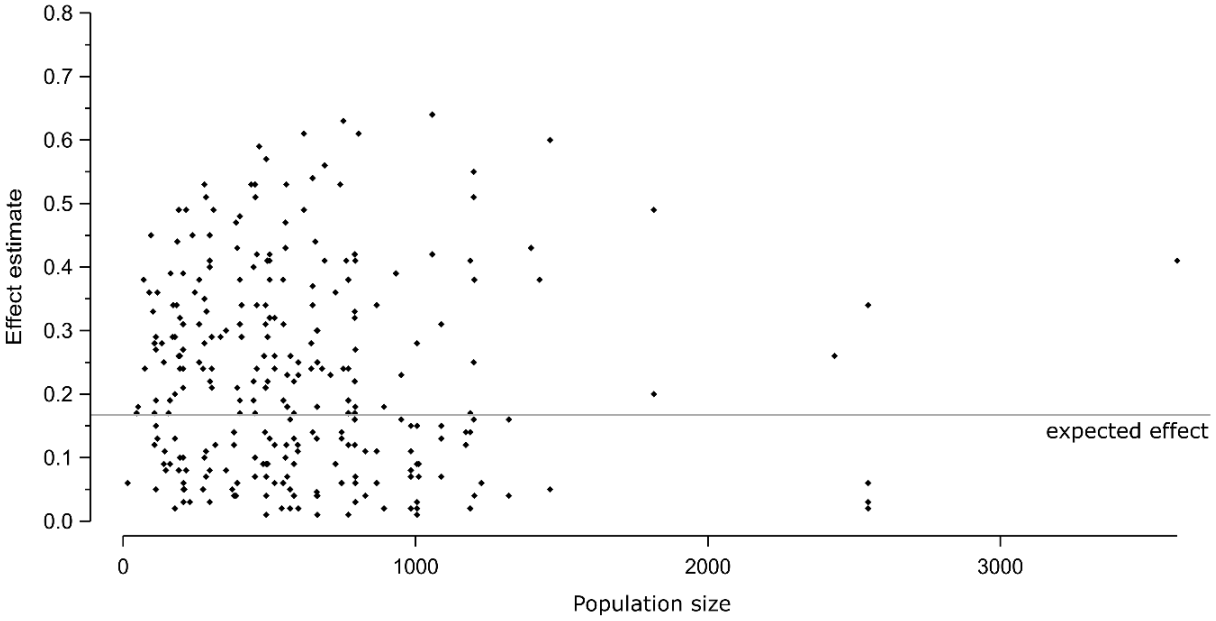
EAB: extent of added benefit.

Figure 7: Effect estimates for “serious/severe symptoms” outcomes with the EAB “major" - distribution by population size



EAB: extent of added benefit, HrQoL: health related quality of life.

Figure 8: Effect estimates for HrQoL outcomes with the EAB “major" - distribution by population size



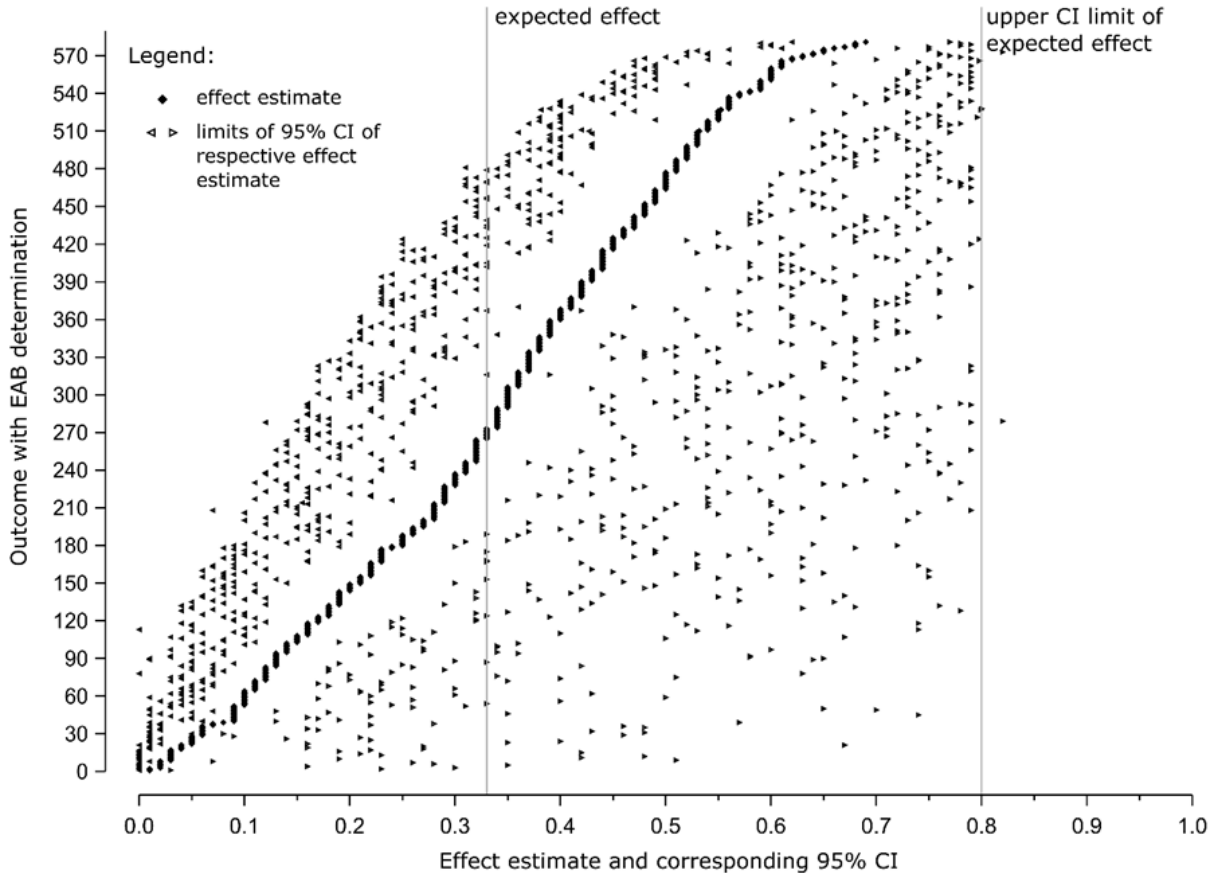
EAB: extent of added benefit.

Figure 9: Effect estimates for “serious/severe adverse effects” outcomes with the EAB “major" - distribution by population size

In Figure 7 and Figure 8 it can be seen that for the outcomes “serious/severe symptoms” and “HrQoL”, the values of the actual effect estimates also tend to increase with increasing population size (number of patients analysed): The larger the number of patients analysed, the smaller the actual effect estimate in this EAB category. However, effect dependence on the population size is not shown for the outcome “serious/severe adverse effects” (see Figure 9). This tendency is also not shown when evaluating the results of the subgroup populations (EAB determination by subgroup characteristic due to subgroup effect) and total populations in the distribution of the actual effect estimates (see Appendix D of the full report).

5.3.3 Outcome category “non-serious/non-severe”, EAB category "considerable"

In the outcome category “non-serious/non-severe” with the EAB "considerable", with a median of 0.35 the values of the actual effect estimates are above the expected effect size of 0.33 in more than half of the cases. Figure 10 shows the effect estimates and related CIs for all outcomes in this outcome category.



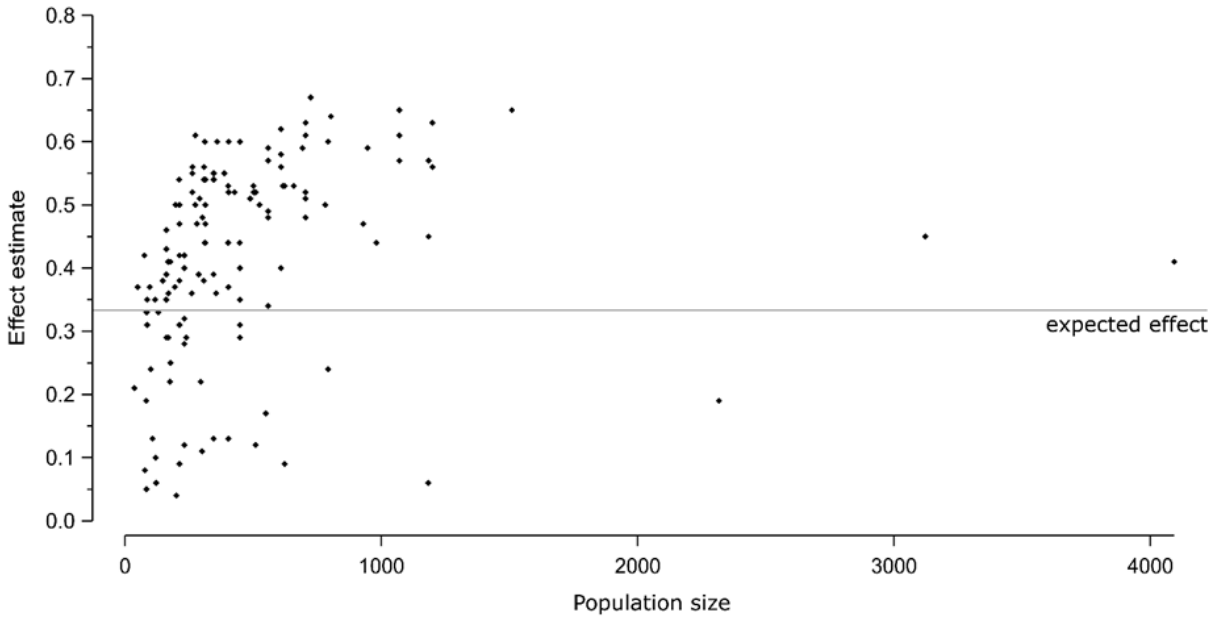
CI: confidence interval; EAB: extent of added benefit.

Figure 10: Effect estimates of all 581 EAB determinations for the outcomes of the outcome category “non-serious/non-severe” with the EAB “considerable" - distribution according to effect estimates and related CIs in ascending order[6]

Approximately 53% of the values of the actual effect estimates are above the expected effect size (0.33), i.e. they show smaller effects than specified in the development of the EAB methods for this EAB category. Likewise, the lower limit of the CI is also above the expected effect size for 44% of these outcomes.

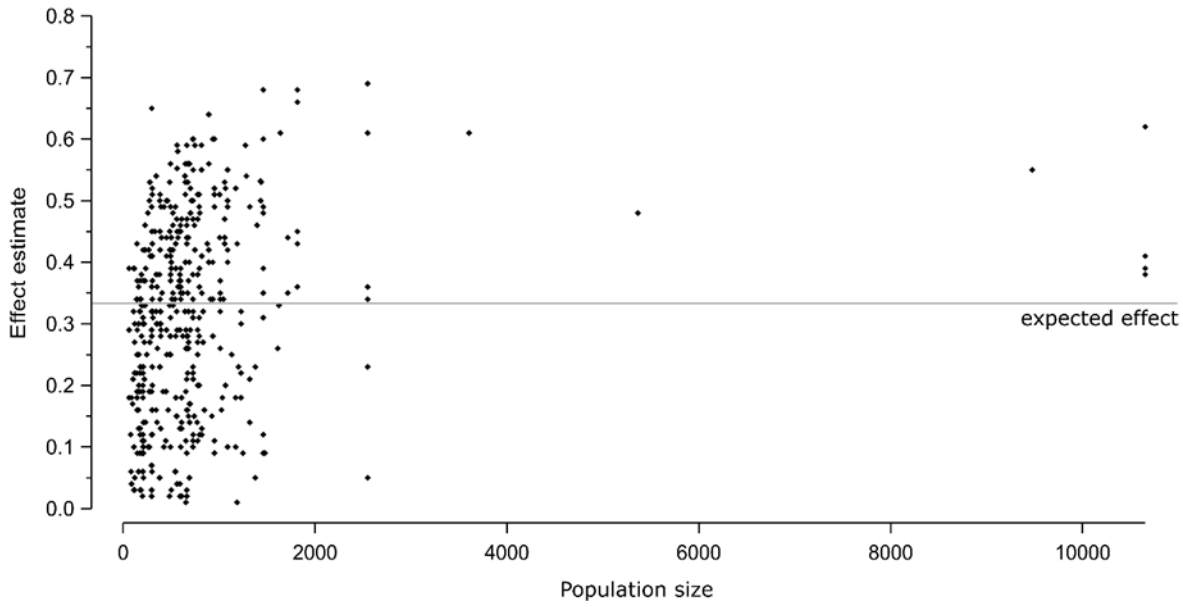
Figure 11 and Figure 12 show the distribution of effect estimates per outcome depending on population size.

[6] It should be noted that the upper limit of the 95% CI was exceeded for 2 outcomes with an EAB "considerable", as can be seen in the figure. This is due to the fact that in these two cases, further content-related aspects were taken into account when determining the EAB.



EAB: extent of added benefit.

Figure 11: Effect estimates for "non-serious/non-serious symptoms” outcomes with the EAB “considerable" - distribution by population size



EAB: extent of added benefit.

Figure 12: Effect estimates for "non-serious/non-serious adverse effects” outcomes with the EAB “considerable" - distribution by population size

In contrast to the outcome category “mortality” and the outcomes “serious/severe symptoms” and “HrQoL”, it is not as clear in this category that the effect decreases with increasing population size (see Figure 12).

6 Discussion

Evidence base

More than 13 years after the introduction of AMNOG, with 667 completed dossier assessments and 266 related addenda up to and including 2022, a comprehensive evidence base for an empirical analysis of EAB derivations at the outcome level is available. This makes it possible to analyse all combinations of outcome and EAB categories.

Results

The analysis of the actual effect estimates shows very different distributions of the actual effect estimates in relation to the expected effect size, depending on the outcome and EAB category. It is particularly striking here that in the highest EAB categories, in more than half of the cases the values of the actual effect estimates in all outcome categories are above the expected effect sizes, which in these cases means that the actual effect estimate is smaller than specified in the development of the EAB methods for the respective EAB category. This means that for these cases, the expected effect size was not achieved with an EAB determination based on the upper CI limit. In contrast, the values of the effect estimates in the respective second-highest EAB categories ("considerable" and "minor") are almost completely or largely below the expected effect sizes, which means that the expected effect size is achieved or exceeded on the basis of the EAB methods applied.

In addition to the effect size and the derived EAB, various other data were evaluated in order to identify possible additional factors for which the distribution of the effect estimates differs within the outcome categories and across outcomes. For this purpose, the topics, the certainty of conclusions about the outcome (hint, indication, proof) and the population size were analysed in more detail.

The distributions of the effect estimates do not differ among the topics of the dossier assessments from which the outcomes were extracted. However, it should be noted that more than 75% of all outcomes originate from dossier assessments of oncological indications, which therefore dominate the overall analysis. The evidence base for non-oncological indications is therefore markedly smaller. The evaluation of the certainty of conclusions also shows no differences in the distribution of effect estimates.

From the factors analysed, within the outcome categories a possible correlation with the distribution of the actual effect sizes was only shown for the population size. In the assessments evaluated, this ranged from 15 to 18,140 included patients. It is apparent that the effect estimate becomes smaller with increasing population size, from which the effect estimates are derived, while at the same time the effects become more precise. This is plausible insofar as the size of the population is a determining factor for the precision of a study. For the outcome category “mortality” with the EAB “major", this is shown in Figure 4

by the narrower CIs for smaller effects and in Figure 5 by the distribution of the effect estimate depending on the population size. A similar correlation can be seen for the outcomes “serious/ severe symptoms” and “HrQoL”, but the correlation in these outcomes is markedly less pronounced than for mortality (Figure 6, Figure 7 and Figure 8). This correlation also remains visible when these are considered separately according to subgroup results and results from total populations (see Appendix D of the full report). However, this effect cannot be observed for the outcome “serious/severe adverse effects” (see Figure 9).

In this context, it should be noted that the outcome categories differ greatly in their composition of outcomes. The outcome category “mortality” only includes the outcome “overall survival” or “all-cause mortality”, while the other two outcome categories include all outcomes in the outcome categories “morbidity”, “HrQoL” and “adverse effects”. The allocation to these two outcome categories is only defined by severity. While the results of the outcome category “mortality” thus show homogeneity with regard to the type of outcome, the other two outcome categories have, by definition, a heterogeneous composition of outcomes. For this reason, separate analyses by outcome were conducted for the outcome category “serious/severe/HrQoL” (see Figure 7, Figure 8, Figure 9). These show that in particular the outcome “adverse effects” exhibits very different distributions of the actual effect estimates, also in relation to the expected effect size, which reflect the heterogeneity of adverse effects.

In summary, there are very different distributions of actual effect estimates in relation to the expected effect size within the outcome and EAB categories, with the values of the actual effect estimates being markedly higher than the expected effect sizes in many cases, particularly in the highest possible EAB categories (Figure 3). In the highest possible EAB category, EAB determination on the basis of upper CI limits as thresholds thus leads to the expected effect not being achieved in more than half of the cases. With a view to a potential further development of the EAB methods, it should therefore be discussed whether further or other criteria should be used. For example, for oncological studies the European Society for Medical Oncology uses other criteria in its Magnitude of Clinical Benefit Scale (ESMO-MCBS) and besides absolute criteria considers the lower CI limit as a threshold [5,6]. For example, for a high proportion of the outcomes described in Section 5.3 and shown in Figure 4, Figure 6 and Figure 10, the maximum possible EAB would not be derived if there were an additional requirement that at least the lower CI limit is below the expected effect.

7 Conclusion

Based on the results of completed dossier assessments, the aim of the present analysis was to empirically examine, for the various outcome categories, how far the desired effect sizes specified in the development of the EAB methods in 2011 are actually achieved when these methods are applied. For this purpose, the actual effect estimates from the dossier assessments are compared with the desired effect sizes specified for threshold determination.

With a total of 1747 EAB determinations at the outcome level from 667 dossier assessments and 266 addenda from 2011 up to and including 2022, a comprehensive evidence base is available.

The empirically specified actual effect estimates only partially reach the effect sizes expected according to the development of the EAB methods:

- For the highest EAB categories ("major" for the outcome categories "mortality" and "severe/severe/HrQoL", "considerable" for the outcome category "non-serious/non-severe"), more than half of the actual effect estimates are smaller than the expected effect sizes.
- For the other EAB categories ("considerable" and "minor" for the outcome categories "mortality" and "serious/severe/HrQoL", "minor" for the outcome category "non-serious/non-severe"), the effect estimates are almost completely or mostly below the expected effect sizes.
- With the exception of population size, no other factors could be identified for which at least partially different distributions of the actual effect estimates were found.
- A possible adaptation of the EAB methods should focus on the highest EAB categories.

References for English extract

Please see full working paper for full reference list.

1. Institute for Quality and Efficiency in Health Care. General Methods; Version 7.0 [online]. 2023 [Accessed: 07.11.2024]. URL: https://www.iqwig.de/en/about-us/methods/methods-paper/.

2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ticagrelor; Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung; Auftrag A11-02 [online]. 2011 [Accessed: 11.03.2013]. URL: https://www.iqwig.de/download/A11-02_Ticagrelor_Nutzenbewertung_35a_SGB_V_.pdf.

3. Djulbegovic B, Kumar A, Soares HP et al. Treatment success in cancer; new cancer treatment successes identified in phase 3 randomized controlled trials conducted by the National Cancer Institute-sponsored cooperative oncology groups, 1955 to 2006. Arch Intern Med 2008; 168(6): 632-642. https://doi.org/10.1001/archinte.168.6.632.

4. Skipka G, Wieseler B, Kaiser T 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. https://doi.org/10.1002/bimj.201300274.

5. Cherny NI, Sullivan R, Dafni U et al. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies; the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). Ann Oncol 2015; 26(8): 1547-1573. https://doi.org/10.1093/annonc/mdv249.

6. Kiesewetter B, Dafni U, de Vries EGE et al. ESMO-Magnitude of Clinical Benefit Scale for haematological malignancies (ESMO-MCBS:H); version 1.0. Annals of Oncology 2023; 34(9): 734-771. https://doi.org/10.1016/j.annonc.2023.06.002.

The full report (German version) is published under
https://www.iqwig.de/en/projects/ga23-01.html