

Ausmaßmethodik Zusatznutzen – Empirie aus Dossierbewertungen

A horizontal bar composed of 18 rectangular segments of varying shades of blue and grey. The word 'ARBEITSPAPIER' is written in white capital letters on a dark blue segment that spans across the middle of the bar.

ARBEITSPAPIER

Projekt: GA23-01

Version: 1.1

Stand: 15.11.2024

IQWiG-Berichte – Nr. 1869

DOI: 10.60584/GA23-01_V1.1

Impressum

Herausgeber

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema

Ausmaßmethodik Zusatznutzen – Empirie aus Dossierbewertungen

Auftraggeber

Bearbeitung im Rahmen des Generalauftrags

Interne Projektnummer

GA23-01

DOI-URL

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

Anschrift des Herausgebers

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

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: berichte@iqwig.de

Internet: www.iqwig.de

ISSN: 1864-2500

Zitiervorschlag

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ausmaßmethodik Zusatznutzen – Empirie aus Dossierbewertungen; Arbeitspapier [online]. 2024 [Zugriff: TT.MM.JJJJ]. URL: https://doi.org/10.60584/GA23-01_V1.1.

Schlagwörter

Arzneimittelzulassung, Behandlungsergebnis, Staatliche Regulierung, Nutzenbewertung

Keywords

Drug Approval, Treatment Outcome, Government Regulation, Benefit Assessment

Dieses Arbeitspapier wurde ohne die Beteiligung externer Sachverständiger erstellt.

Für die Inhalte des Berichts ist allein das IQWiG verantwortlich.

Mitarbeiterinnen und Mitarbeiter des IQWiG

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

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	vi
Abbildungsverzeichnis	vii
Abkürzungsverzeichnis.....	x
Kurzfassung	xi
1 Hintergrund.....	1
2 Fragestellung.....	4
3 Projektverlauf.....	5
3.1 Spezifizierungen und Änderungen im Projektverlauf.....	5
4 Methoden	6
4.1 Informationsbeschaffung.....	6
4.2 Informationsbewertung.....	6
4.3 Informationssynthese und -analyse	6
5 Ergebnisse	7
5.1 Ergebnis der Informationsbeschaffung.....	7
5.2 Verteilung der tatsächlichen Effektschätzungen (alle Endpunkte).....	10
5.3 Detaillierte Betrachtung der Verteilung der tatsächlichen Effektschätzungen – nach Zielgrößen- und Ausmaßkategorie.....	14
5.3.1 Zielgrößenkategorie Mortalität, Ausmaßkategorie „erheblich“	14
5.3.2 Zielgrößenkategorie schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität, Ausmaßkategorie „erheblich“	16
5.3.3 Zielgrößenkategorie nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen, Ausmaßkategorie „beträchtlich“	19
6 Diskussion	22
7 Fazit	25
8 Literatur	26
Anhang A Verteilung der tatsächlichen Effektschätzungen – Sensitivitätsanalyse	27
Anhang B Verteilung der tatsächlichen Effektschätzungen nach Aussagesicherheit	28
B.1 Zielgrößenkategorie Mortalität, Ausmaßkategorie „erheblich“	28
B.2 Zielgrößenkategorie schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität, Ausmaßkategorie „erheblich“	30

B.2.1	Zielgröße schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen).....	30
B.2.2	Zielgröße gesundheitsbezogene Lebensqualität.....	32
B.2.3	Zielgröße schwerwiegende (bzw. schwere) Nebenwirkungen	33
B.3	Zielgrößenkategorie nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen, Ausmaßkategorie „beträchtlich“	35
B.3.1	Zielgröße nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen).....	35
B.3.2	Zielgröße nicht schwerwiegende (bzw. nicht schwere) Nebenwirkungen	37
Anhang C	Verteilung der tatsächlichen Effektschätzungen nach Themengebiet.....	39
Anhang D	Verteilung der tatsächlichen Effektschätzungen – Gesamtpopulation und Subgruppenpopulation	43

Tabellenverzeichnis

	Seite
Tabelle 1: Empirische Ergebnisse – Verteilung der Endpunkte mit Ausmaßbestimmung nach Evidenzart und Themengebiet.....	8
Tabelle 2: Empirische Ergebnisse – Verteilung der Endpunkte mit Ausmaßbestimmung nach Zielgrößenkategorien und Ausmaßkategorien	9
Tabelle 3: Empirische Ergebnisse – Verteilung der Endpunkte mit Ausmaßbestimmung nach Zielgrößen	9
Tabelle 4: Verteilung der vorliegenden tatsächlichen Effektschätzungen (alle Indikationen).....	12
Tabelle 5: Verteilung der vorliegenden tatsächlichen Effektschätzungen (Themengebiet Onkologie).....	39
Tabelle 6: Verteilung der vorliegenden tatsächlichen Effektschätzungen (Themengebiet nicht onkologische Bewertungen).....	41

Abbildungsverzeichnis

	Seite
Abbildung 1: Feststellung des Ausmaßes des Zusatznutzens – quantitative Operationalisierungen	2
Abbildung 2: Ergebnis der Informationsbeschaffung und -bewertung	7
Abbildung 3: Verteilung der vorliegenden tatsächlichen Effektschätzungen nach Zielgrößen- und Ausmaßkategorie – Auswertung aller Endpunkte	11
Abbildung 4: Effektschätzungen aller 45 Ausmaßbestimmungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ – Verteilung der Effektschätzungen und der zugehörigen Konfidenzintervalle (95 %-KI) in aufsteigender Reihenfolge	15
Abbildung 5: Effektschätzungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ – Verteilung nach Populationsgröße.....	16
Abbildung 6: Effektschätzungen aller 350 Ausmaßbestimmungen der Endpunkte der Zielgrößenkategorie schwerwiegend / schwer / HQoL mit Ausmaß „erheblich“ – Verteilung nach Effektschätzungen und der zugehörigen Konfidenzintervalle in aufsteigender Reihenfolge	17
Abbildung 7: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Symptome mit Ausmaß „erheblich“ – Verteilung nach Populationsgröße.....	18
Abbildung 8: Effektschätzungen der Endpunkte der Zielgröße HQoL mit Ausmaß „erheblich“ – Verteilung nach Populationsgröße.....	18
Abbildung 9: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende/schwere Nebenwirkungen mit Ausmaß „erheblich“ – Verteilung nach Populationsgröße	19
Abbildung 10: Effektschätzungen aller 581 Ausmaßbestimmungen der Endpunkte der Zielgrößenkategorie nicht schwerwiegend / nicht schwer mit Ausmaß „beträchtlich“ – Verteilung nach Effektschätzungen und der zugehörigen Konfidenzintervalle in aufsteigender Reihenfolge	20
Abbildung 11: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Symptome mit Ausmaß „beträchtlich“ – Verteilung nach Populationsgröße.....	21
Abbildung 12: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Nebenwirkungen mit Ausmaß „beträchtlich“ – Verteilung nach Populationsgröße.....	21
Abbildung 13: Sensitivitätsanalyse zur Verteilung der tatsächlichen Effektschätzungen – jeweils ein zufällig ausgewählter Endpunkt pro Dossierbewertung	27
Abbildung 14: Effektschätzungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße.....	28
Abbildung 15: Effektschätzungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße.....	28

Abbildung 16: Effektschätzungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße.....	29
Abbildung 17: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Symptome mit Ausmaß „erheblich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße	30
Abbildung 18: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Symptome mit Ausmaß „erheblich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße	31
Abbildung 19: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Symptome mit Ausmaß „erheblich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße	31
Abbildung 20: Effektschätzungen der Endpunkte der Zielgröße gesundheitsbezogene Lebensqualität mit Ausmaß „erheblich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße	32
Abbildung 21: Effektschätzungen der Endpunkte der Zielgröße gesundheitsbezogene Lebensqualität mit Ausmaß „erheblich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße	33
Abbildung 22: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Nebenwirkungen mit Ausmaß „erheblich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße	33
Abbildung 23: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Nebenwirkungen mit Ausmaß „erheblich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße	34
Abbildung 24: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Nebenwirkungen mit Ausmaß „erheblich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße	34
Abbildung 25: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Symptome mit Ausmaß „beträchtlich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße	35
Abbildung 26: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Symptome mit Ausmaß „beträchtlich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße	36
Abbildung 27: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Symptome mit Ausmaß „beträchtlich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße	36
Abbildung 28: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Nebenwirkungen mit Ausmaß „beträchtlich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße	37
Abbildung 29: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Nebenwirkungen mit Ausmaß „beträchtlich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße	37

Abbildung 30: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Nebenwirkungen mit Ausmaß „beträchtlich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße 38

Abbildung 31: Verteilung der tatsächlichen Effektschätzungen – Auswertung aller Endpunkte für das Themengebiet Onkologie..... 40

Abbildung 32: Verteilung der tatsächlichen Effektschätzungen – Auswertung aller Endpunkte für das Themengebiet nicht onkologische Bewertungen 42

Abbildung 33: Verteilung der tatsächlichen Effektschätzungen – Gesamtpopulation..... 43

Abbildung 34: Verteilung der tatsächlichen Effektschätzungen – Subgruppenpopulation..... 44

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AMNOG	Arzneimittelmarktneuordnungsgesetz
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
ESMO	European Society for Medical Oncology
G-BA	Gemeinsamer Bundesausschuss
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
HQoL	Health related Quality of Life (gesundheitsbezogene Lebensqualität)
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
SUE	schwerwiegendes unerwünschtes Ereignis
UE	unerwünschtes Ereignis

Kurzfassung

Im Rahmen des Generalauftrags wurde das Thema Ausmaßmethodik Zusatznutzen – Empirie aus Dossierbewertungen bearbeitet.

Hintergrund

Für die Bestimmung des Zusatznutzens in der Dossierbewertung ist gemäß Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) das Ausmaß des Zusatznutzens zu quantifizieren (Kategorien: erheblich, beträchtlich, gering sowie nicht quantifizierbar, kein Zusatznutzen belegt, geringerer Nutzen). Die Methodik zur Ausmaßbestimmung für relative Effektmaße wurde erstmals mit der Dossierbewertung zu Ticagrelor (A11-02) publiziert und nachfolgend in die Allgemeinen Methoden des Instituts übernommen.

Die Ausmaßmethodik basiert auf Annahmen von Effektgrößen für relative Effektmaße. Die Anforderungen an die Effektgrößen, die zur Einstufung eines Zusatznutzens als „erheblich“, „beträchtlich“ und „gering“ führen, variieren in Abhängigkeit von der Bedeutung der Zielgrößen. Die Zielgrößenkategorien lassen sich in 3 Kategorien gruppieren:

- 1) Mortalität
- 2) schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)
schwerwiegende (bzw. schwere) Nebenwirkungen
gesundheitsbezogene Lebensqualität (HQoL)
- 3) nicht schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)
nicht schwerwiegende (bzw. schwere) Nebenwirkungen

Für jede Zielgrößenkategorie wurde für die Einstufung eines Ergebnisses in die verschiedenen Ausmaße des Zusatznutzens (erheblich, beträchtlich, gering) jeweils ein gewünschter tatsächlicher Effekt zugrunde gelegt. Unter Verwendung einer hypothetischen Fallzahlschätzung wurde ausgehend von diesem gewünschten Effekt ein Schwellenwert berechnet. Im Sinne einer verschobenen Hypothesengrenze muss ein 95 %-Konfidenzintervall diesen Schwellenwert unterschreiten, damit das Ergebnis als erheblicher, beträchtlicher oder geringer Zusatznutzen eingestuft wird.

Seit Einführung des Arzneimittelmarktneuordnungsgesetzes (AMNOG) und der Dossierbewertung zu Ticagrelor (A11-02) wurden bis April 2023 insgesamt 679 Dossierbewertungen erstellt und – sofern geeignete Daten für die Bewertung vorlagen – das jeweilige Ausmaß auf Endpunktebene gemäß der genannten Methodik bestimmt. Auf Basis dieser Ergebnisse lässt sich beschreiben, ob und in welchem Umfang die tatsächlich vorliegenden Effektschätzungen in den verschiedenen Endpunktkategorien die gewünschten

Effektgrößen erreichen, die der Methodik zur Feststellung des Ausmaßes des Zusatznutzens zugrunde gelegt wurden.

Fragestellung

Ziel der vorliegenden Untersuchung ist es, auf Basis der Ergebnisse aus abgeschlossenen Dossierbewertungen für die verschiedenen Endpunktkategorien empirisch zu prüfen, inwieweit durch die Anwendung der im Jahr 2011 entwickelten Ausmaßmethodik die in der Entwicklung der Methodik beschriebenen gewünschten Effektgrößen tatsächlich erreicht werden. Zu diesem Zweck erfolgt ein Abgleich der tatsächlichen Effektschätzungen aus den Dossierbewertungen mit den für die Schwellenwertbestimmung festgelegten gewünschten Effektgrößen.

Die bei der Entwicklung der Methodik festgelegten gewünschten Effektgrößen werden im Folgenden beim Vergleich mit den tatsächlichen Effektschätzungen als erwartete Effektstärken bezeichnet.

Methoden

Für das vorliegende Projekt wurden alle Dossierbewertungen von 2011 bis einschließlich 2022 herangezogen und die relevanten Informationen extrahiert. Neben Basisinformationen (u. a. Auftragsnummer, Wirkstoff, Indikation und Fragestellungen) umfasst das insbesondere die Angaben zu den Effektschätzungen. Die Effektschätzungen der statistisch signifikanten Effekte, die in die Bestimmung des Ausmaßes des Zusatznutzens eingegangen sind, wurden extrahiert und ausgewertet. Dazu wurden die statistisch signifikanten Effekte aus der jeweiligen Tabelle zur Darstellung des Ausmaßes des Zusatznutzens auf Endpunktebene herangezogen, die in den Dossierbewertungen vorliegt. Als relevante Effektmaße wurden das relative Risiko (RR), das Hazard Ratio (HR), das Peto-Odds Ratio (Peto-OR) und das Rate Ratio berücksichtigt. Sofern eine Effektmodifikation durch Subgruppenmerkmale vorlag, wurden die Ergebnisse der entsprechenden Subgruppe(n) aus der Ausmaßtabelle extrahiert. Neben den Effekten wurden aus der Tabelle zum Ausmaß des Zusatznutzens auch Angaben zur Endpunktkategorie (z. B. gesundheitsbezogene Lebensqualität), zur Effektrichtung (positiv / negativ aus Sicht des bewerteten Arzneimittels), zum Ausmaß des Zusatznutzens sowie zur Wahrscheinlichkeit (Anhaltspunkt, Hinweis, Beleg) extrahiert. Zusätzlich relevante Informationen, u. a. die Größe der Auswertungspopulation, die Art der vorgelegten Evidenz (direkter, indirekter Vergleich oder Metaanalyse) und die Bezeichnung der für die Ableitung des Zusatznutzens eingeschlossenen Studien, wurden ebenfalls extrahiert.

Für die Bewertung der in die Untersuchung eingehenden Informationen wurde auf die Dossierbewertungen, ggf. einschließlich zugehöriger Addenda, zurückgegriffen. Im Falle von Addenda und Neubewertungen nach Fristablauf oder bei neuen wissenschaftlichen

Erkenntnissen wurden jeweils die aktuellsten für die Dossierbewertung relevanten Ergebnisse extrahiert und in die Auswertung einbezogen.

Die Informationen wurden einer Informationssynthese und -analyse unterzogen. Dazu wurde zunächst der Datenpool beschrieben, indem die eingeschlossenen Dossierbewertungen charakterisiert werden (z. B. Indikation, Wirkstoff). Die für die einzelnen Endpunktkategorien tatsächlich vorliegenden Effektschätzungen wurden dann mit den erwarteten Effektstärken, die in der Methodik zur Ermittlung des Ausmaßes des Zusatznutzens zur Ableitung der Schwellenwerte herangezogen wurden, abgeglichen.

Ergebnisse

Ergebnis der Informationsbeschaffung

Zur Beantwortung der vorliegenden Fragestellung lagen 667 Dossierbewertungen und 266 Addenda aus den Jahren 2011 bis einschließlich 2022 vor. In 239 Dossierbewertungen einschließlich Addenda lag für 277 Fragestellungen für mindestens 1 Endpunkt eine Ausmaßbestimmung vor. Insgesamt wurden 1747 Endpunkte extrahiert.

Von den insgesamt 1747 Ausmaßbestimmungen entstammen 89 % (n = 1557) einem direkten Vergleich einer Studie und 10 % (n = 174) Metaanalysen mit mehreren direkt vergleichenden Studien. Ein Großteil der Endpunkte mit Ausmaßbestimmungen (77,5 %) stammt aus Dossierbewertungen im Themengebiet Onkologie.

Für alle möglichen Kombinationen aus Zielgrößen und Ausmaßkategorie liegen jeweils Ausmaßbestimmungen auf Endpunktebene vor, sodass eine Betrachtung aller Kombinationen möglich ist. Mehr als die Hälfte aller Ausmaßbestimmungen (52 % [n = 911]) erfolgte bei Endpunkten der Zielgrößenkategorie schwerwiegend / schwer / HQoL, wobei der überwiegende Teil aus Endpunkten zu Nebenwirkungen stammt.

Verteilung der tatsächlichen Effektschätzungen (alle Endpunkte)

In der Gesamtbetrachtung zeigen sich je nach Zielgrößen- und Ausmaßkategorie sehr unterschiedliche Verteilungen der tatsächlichen Effektschätzungen.

Die Werte der tatsächlichen Effektschätzungen in den jeweiligen höchsten Ausmaßkategorien („erheblich“ für Mortalität und schwerwiegend / schwer / HQoL; „beträchtlich“ für nicht schwerwiegend / nicht schwer) befinden sich im Median leicht oberhalb der erwarteten Effektstärken: mediane Effektschätzung bei 0,54 (erwarteter Effekt: 0,50) bzw. 0,24 (erwarteter Effekt: 0,17) für das Ausmaß „erheblich“ und bei 0,35 (erwarteter Effekt: 0,33) für das Ausmaß „beträchtlich“. Damit haben mehr als die Hälfte der Endpunkte, für die das höchstmögliche Ausmaß abgeleitet wurde, einen kleineren Effekt als bei der Entwicklung der Methodik für das jeweilige Ausmaß festgelegt.

In den jeweiligen zweithöchsten Ausmaßkategorien („beträchtlich“ für Zielgrößenkategorien Mortalität sowie schwerwiegend / schwer / HQoL; „gering“ für die Zielgrößenkategorie nicht schwerwiegend / nicht schwer) liegen die tatsächlichen Effektschätzungen fast vollständig bzw. größtenteils unterhalb der erwarteten Effektstärken: mediane Effektschätzung bei 0,70 (erwarteter Effekt: 0,83) bzw. 0,49 (erwarteter Effekt: 0,67) für das Ausmaß „beträchtlich“ und bei 0,58 (erwarteter Effekt: 0,67) für das Ausmaß „gering“. Damit werden die erwarteten Effektstärken in der ganz überwiegenden Anzahl der Fälle für diese Ausmaßkategorien erreicht.

Insgesamt zeigt sich gerade in der jeweils höchsten Ausmaßkategorie („erheblich“ für Mortalität sowie schwerwiegend / schwer / HQoL; „beträchtlich“ für nicht schwerwiegend / nicht schwer), dass die Effekte in mehr als der Hälfte der Ausmaßbestimmungen kleiner sind als bei der Entwicklung der Methodik für das jeweilige Ausmaß festgelegt.

Verteilung der tatsächlichen Effektschätzungen – nach Zielgrößen- und Ausmaßkategorie

Zielgrößenkategorie Mortalität, Ausmaßkategorie „erheblich“

Etwa 62 % der Werte der tatsächlichen Effektschätzungen liegen oberhalb der erwarteten Effektstärke (0,50), zeigen also kleinere Effekte als bei der Entwicklung der Methodik für diese Ausmaßkategorie festgelegt. Zudem befindet sich für ca. 29 % dieser Endpunkte zusätzlich auch die untere Konfidenzintervallgrenze oberhalb der erwarteten Effektstärke. Dies bedeutet, dass der erwartete Effekt von 0,50 in diesen Fällen nicht nur größer ist, sondern auch außerhalb des 95 %-Konfidenzintervalls der tatsächlichen Effektschätzung liegt.

Mit zunehmender Populationsgröße (Anzahl ausgewerteter Patientinnen und Patienten) steigen die Werte der tatsächlichen Effektschätzungen tendenziell: Je größer die Anzahl ausgewerteter Patientinnen und Patienten, desto kleiner wird die tatsächliche Effektschätzung in dieser Ausmaßkategorie. Dies erscheint plausibel, da die Populationsgröße Einfluss auf die Präzision der Effektschätzung einer Studie hat. Deshalb unterschreitet in größeren Populationen die obere Grenze des 95 %-Konfidenzintervalls bei kleineren Effekten häufiger den Schwellenwert für eine Ausmaßkategorie. Dieser potenzielle Zusammenhang zwischen Populationsgröße und Effektstärke ist auch bei der Verteilung der tatsächlichen Effektschätzungen getrennt nach Subgruppenpopulation (Ausmaßbestimmung nach Subgruppenmerkmal wegen Subgruppeneffekt) und Gesamtpopulationen erkennbar, jedoch nur für diese Zielgrößenkategorie.

Zielgrößenkategorie schwerwiegende Symptome bzw. Nebenwirkungen und gesundheitsbezogene Lebensqualität, Ausmaßkategorie „erheblich“

Etwa 61 % der Werte der tatsächlichen Effektschätzungen dieser Zielgrößenkategorie liegen oberhalb der erwarteten Effektstärke (0,17), zeigen also kleinere Effekte als bei der Entwicklung der Methodik für diese Ausmaßkategorie festgelegt. Ebenso ist für mehr als die

Hälfte dieser Endpunkte zusätzlich auch die untere Grenze des Konfidenzintervalls oberhalb der erwarteten Effektstärke. Auch für die Zielgrößen schwerwiegende / schwere Symptome und gesundheitsbezogene Lebensqualität steigen mit zunehmender Populationsgröße (Anzahl ausgewerteter Patientinnen und Patienten) die Werte der tatsächlichen Effektschätzungen tendenziell. Eine Abhängigkeit der Effekte von der Populationsgröße für die Zielgröße schwerwiegende / schwere Nebenwirkungen ist nicht erkennbar.

Zielgrößenkategorie nicht schwerwiegende Symptome bzw. Nebenwirkungen,
Ausmaßkategorie „beträchtlich“

Etwa 53 % der Werte der tatsächlichen Effektschätzungen liegen oberhalb der erwarteten Effektstärke (0,33), zeigen also kleinere Effekte als bei der Entwicklung der Methodik für diese Ausmaßkategorie festgelegt. Ebenso ist für 44 % dieser Endpunkte zusätzlich auch die untere Grenze des Konfidenzintervalls oberhalb der erwarteten Effektstärke. Im Gegensatz zu der Zielgrößenkategorie Mortalität und den Zielgrößen schwerwiegende / schwere Symptome und gesundheitsbezogene Lebensqualität ist in dieser Kategorie nicht zu erkennen, dass der Effekt mit zunehmender Populationsgröße kleiner wird.

Fazit

Das Ziel der vorliegenden Untersuchung war es, auf Basis der Ergebnisse aus abgeschlossenen Dossierbewertungen für die verschiedenen Endpunktkategorien empirisch zu prüfen, inwieweit durch die Anwendung der im Jahr 2011 entwickelten Ausmaßmethodik die in der Entwicklung der Methodik festgelegten gewünschten Effektgrößen tatsächlich erreicht werden. Zu diesem Zweck erfolgt ein Abgleich der tatsächlichen Effektschätzungen aus den Dossierbewertungen mit den für die Schwellenwertbestimmung festgelegten gewünschten Effektgrößen.

Mit insgesamt 1747 Ausmaßbestimmungen auf Endpunktebene aus 667 Dossierbewertungen und 266 Addenda der Jahre 2011 bis 2022 liegt eine umfangreiche Datengrundlage vor.

Die empirisch festgestellten tatsächlichen Effektschätzungen erreichen nur teilweise die gemäß der Entwicklung der Methodik erwarteten Effektstärken:

- Für die höchsten Ausmaßkategorien („erheblich“ für Zielgrößenkategorien Mortalität und schwerwiegend / schwer / HQoL, „beträchtlich“ für Zielgrößenkategorie nicht schwerwiegend / nicht schwer) sind mehr als die Hälfte der tatsächlichen Effektschätzungen kleiner als die erwarteten Effektstärken.
- Für die übrigen Ausmaßkategorien („beträchtlich“ und „gering“ für die Zielgrößenkategorien Mortalität und schwerwiegend / schwer / HQoL, „gering“ für die Zielgrößenkategorie nicht schwerwiegend / nicht schwer) liegen die Effektschätzungen fast vollständig bzw. größtenteils unterhalb der erwarteten Effektstärken.

- Mit Ausnahme der Populationsgröße konnten keine weiteren Faktoren identifiziert werden, für die sich zumindest teilweise unterschiedliche Verteilungen der tatsächlichen Effektschätzungen zeigen.
- Eine mögliche Anpassung der Methodik zur Ausmaßbestimmung sollte sich auf die höchsten Ausmaßkategorien fokussieren.

1 Hintergrund

Bestimmung des Zusatznutzens gemäß Arzneimittel-Nutzenbewertungsverordnung

Für die Bestimmung des Zusatznutzens in der Dossierbewertung ist gemäß Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) das Ausmaß des Zusatznutzens zu quantifizieren (Kategorien: erheblich, beträchtlich, gering sowie nicht quantifizierbar, kein Zusatznutzen belegt, geringerer Nutzen). Die Methodik zur Ausmaßbestimmung für relative Effektmaße wurde erstmals mit der Dossierbewertung zu Ticagrelor (A11-02) publiziert und nachfolgend in die Allgemeinen Methoden des Instituts übernommen [1,2].

Ausmaßmethodik

Die Ausmaßmethodik basiert auf Annahmen von Effektgrößen für relative Effektmaße. Die Anforderungen an die Effektgrößen, die zur Einstufung eines Zusatznutzens als „erheblich“, „beträchtlich“ und „gering“ führen, variieren in Abhängigkeit von der Bedeutung der Zielgrößen. Dabei werden die Zielgrößen gemäß ihrer Bedeutung wie folgt gruppiert:

- 1) Mortalität
- 2) schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)
schwerwiegende (bzw. schwere) Nebenwirkungen
gesundheitsbezogene Lebensqualität (HQoL)
- 3) nicht schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)
nicht schwerwiegende (bzw. schwere) Nebenwirkungen

Zur einfacheren Darstellung und besseren Lesbarkeit werden für die Zielgrößenkategorien in dem vorliegenden Arbeitspapier die folgenden Bezeichnungen verwendet:

- 1) Mortalität
- 2) schwerwiegend / schwer / HQoL
- 3) nicht schwerwiegend / nicht schwer

Für jede Zielgrößenkategorie wurde für die Einstufung eines Ergebnisses in die verschiedenen Ausmaße des Zusatznutzens (erheblich, beträchtlich, gering) jeweils ein gewünschter tatsächlicher Effekt zugrunde gelegt. Unter Verwendung einer hypothetischen Fallzahlschätzung wurde ausgehend von diesem gewünschten Effekt ein Schwellenwert berechnet. Im Sinne einer verschobenen Hypothesengrenze muss ein 95 %-Konfidenzintervall diesen Schwellenwert unterschreiten, damit das Ergebnis als erheblicher, beträchtlicher oder geringer Zusatznutzen eingestuft wird.

Für die Zielgröße Mortalität wird eine Verlängerung der Überlebensdauer als „erheblich“ bewertet, wenn der aus einem gewünschten Effekt von 0,50 (relatives Risiko) abgeleitete Schwellenwert von 0,85 durch die obere Grenze des 95 %-Konfidenzintervalls unterschritten wird. Ein relatives Risiko von 0,50 wurde von Djulbegovic et al. 2008 als Anforderung für einen „Durchbruch“ postuliert [3]. Davon ausgehend wurden für die verschiedenen o. g. Zielgrößenkategorien zunächst die gewünschten Effektgrößen und davon abgeleitet zugehörige Schwellenwerte der jeweiligen 95%-Konfidenzintervalle für die Ableitung der unterschiedlichen Ausmaße festgelegt (siehe Abbildung 1 und Dossierbewertung A11-02) [1,2,4].

		Zielgrößenkategorie			
		Überlebenszeit (Mortalität)	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Lebensqualität	Nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
Zusatznutzen	Erheblich nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapielevanten Nutzens	Erhebliche Verlängerung der Überlebensdauer KI ₅ : 0,85 (RR ₁ = 0,50)	Langfristige Freiheit bzw. weitgehende Vermeidung KI ₅ : 0,75 (RR ₁ = 0,17) und Risiko $\geq 5\%^2$	Erhebliche Verbesserung ¹ KI ₅ : 0,75 (RR ₁ = 0,17) und Risiko $\geq 5\%^2$	Nicht besetzt
	Beträchtlich gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapielevanten Nutzens	Moderate Verlängerung der Überlebensdauer KI ₅ : 0,95 (RR ₁ = 0,83)	Abschwächung bzw. relevante Vermeidung KI ₅ : 0,90 (RR ₁ = 0,67)	Bedeutsame Verbesserung ¹ KI ₅ : 0,90 (RR ₁ = 0,67)	Bedeutsame Vermeidung KI ₅ : 0,80 (RR ₁ = 0,33)
	Gering gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapielevanten Nutzens	Jegliche (statistisch signifikante) Verlängerung der Überlebensdauer KI ₅ : 1,00	Jegliche (statistisch signifikante) Verringerung KI ₅ : 1,00	Relevante Verbesserung ¹ KI ₅ : 1,00	Relevante Vermeidung KI ₅ : 0,90 (RR ₁ = 0,67)
Ergänzungen gegenüber AM-NutzenV <i>kursiv</i> gesetzt 1: Voraussetzung ist die Verwendung eines validierten Instruments sowie eines validierten Responsekriteriums. Werte gelten für Non-Response. 2: für mindestens eine der beiden zu vergleichenden Gruppen. AM-NutzenV: Arzneimittel-Nutzenbewertungsverordnung, KI ₅ : Schwellenwert für obere Grenze des 95%-Konfidenzintervalls, RR ₁ : tatsächliches Relatives Risiko					

Abbildung 1: Feststellung des Ausmaßes des Zusatznutzens – quantitative Operationalisierungen¹ [1]

Dossierbewertungen seit Einführung der Ausmaßmethodik

Seit Einführung des Arzneimittelmarktneuordnungsgesetzes (AMNOG) und der Dossierbewertung zu Ticagrelor (A11-02) wurden bis April 2023 insgesamt 679 Dossierbewertungen erstellt und – sofern geeignete Daten für die Bewertung vorlagen – das jeweilige Ausmaß auf Endpunktebene gemäß der genannten Methodik bestimmt.

¹ Die bei der Entwicklung der Methodik für das jeweilige Ausmaß gewünschten Effektgrößen wurden ursprünglich auf Basis des relativen Risikos festgelegt. Die in der Abbildung definierte Abkürzung „RR₁: tatsächliches relatives Risiko“ entspricht der im Arbeitspapier beschriebenen erwarteten Effektstärken.

Auf Basis dieser Ergebnisse lässt sich beschreiben, ob und in welchem Umfang die tatsächlich vorliegenden Effektschätzungen in den verschiedenen Endpunktkategorien die gewünschten Effektgrößen erreichen, die der Methodik zur Feststellung des Ausmaßes des Zusatznutzens zugrunde gelegt wurden [2,4]. Der Abgleich zwischen gewünschten und tatsächlichen Effekten kann zudem die Ausgangsbasis für eine Diskussion zur Weiterentwicklung der Ausmaßmethodik darstellen.

2 Fragestellung

Ziel der vorliegenden Untersuchung ist es, auf Basis der Ergebnisse aus abgeschlossenen Dossierbewertungen für die verschiedenen Endpunktkategorien empirisch zu prüfen, inwieweit durch die Anwendung der im Jahr 2011 entwickelten Ausmaßmethodik die in der Entwicklung der Methodik festgelegten gewünschten Effektgrößen tatsächlich erreicht werden. Zu diesem Zweck erfolgt ein Abgleich der tatsächlichen Effektschätzungen aus den Dossierbewertungen mit den für die Schwellenwertbestimmung festgelegten gewünschten Effektgrößen.

Die bei der Entwicklung der Methodik festgelegten gewünschten Effektgrößen werden im Folgenden beim Vergleich mit den tatsächlichen Effektschätzungen als erwartete Effektstärken bezeichnet.

3 Projektverlauf

Im Rahmen des Generalauftrags wurde das Thema Ausmaßmethodik Zusatznutzen – Empirie aus Dossierbewertungen bearbeitet. Bearbeitungsbeginn war der 08.05.2023.

Auf Basis einer internen Projektskizze wurde ein Arbeitspapier erstellt. Dieser Bericht wurde am 21.10.2024 an den G-BA übermittelt und 4 Wochen später auf der Website des IQWiG veröffentlicht.

3.1 Spezifizierungen und Änderungen im Projektverlauf

Änderungen im Vergleich zur Version 1.0

Die vorliegende Version 1.1 vom 15.11.2024 ersetzt Version 1.0 des Arbeitspapiers vom 21.10.2024. Folgende Änderungen sind in Version 1.1 im Vergleich zu Version 1.0 enthalten:

- Die Referenzen zu den Allgemeinen Methoden des IQWiG wurden auf die aktuelle Version 7.0 aktualisiert. Die Referenz zu Abbildung 1 wurde korrigiert.

Die Änderungen haben keine Auswirkungen auf das Fazit des Arbeitspapiers.

4 Methoden

4.1 Informationsbeschaffung

Für das vorliegende Projekt wurden alle Dossierbewertungen von 2011 bis einschließlich 2022 herangezogen und die relevanten Informationen extrahiert. Neben Basisinformationen (u. a. Auftragsnummer, Wirkstoff, Indikation und Fragestellungen) umfasst das insbesondere die Angaben zu den Effektschätzungen. Die Effektschätzungen der statistisch signifikanten Effekte, die in die Bestimmung des Ausmaßes des Zusatznutzens eingegangen sind, wurden extrahiert und ausgewertet. Dazu wurden die statistisch signifikanten Effekte aus der jeweiligen Tabelle zur Darstellung des Ausmaßes des Zusatznutzens auf Endpunktebene herangezogen, die in den Dossierbewertungen vorliegt. Als relevante Effektmaße wurden das relative Risiko (RR), das Hazard Ratio (HR), das Peto-Odds Ratio (Peto-OR) und das Rate Ratio berücksichtigt. Sofern eine Effektmodifikation durch Subgruppenmerkmale vorlag, wurden die Ergebnisse der entsprechenden Subgruppe(n) aus der Ausmaßtabelle extrahiert. Neben den Effekten wurden aus der Tabelle zum Ausmaß des Zusatznutzens auch Angaben zur Endpunktkategorie (z. B. gesundheitsbezogene Lebensqualität), zur Effektrichtung (positiv / negativ aus Sicht des bewerteten Arzneimittels), zum Ausmaß des Zusatznutzens sowie zur Wahrscheinlichkeit (Anhaltspunkt, Hinweis, Beleg) extrahiert. Zusätzlich relevante Informationen, u. a. die Größe der Auswertungspopulation, die Art der vorgelegten Evidenz (direkter, indirekter Vergleich oder Metaanalyse) und die Bezeichnung der für die Ableitung des Zusatznutzens eingeschlossenen Studien, wurden ebenfalls extrahiert.

Extrahiert und somit in die Analyse einbezogen wurden die Dossierbewertungen von 2011 bis einschließlich 2022 von A11-02 bis A22-137 (einschließlich der dazugehörigen Addenda).

4.2 Informationsbewertung

Für die Bewertung der in die Untersuchung eingehenden Informationen wurde auf die Dossierbewertungen, ggf. einschließlich zugehöriger Addenda, zurückgegriffen. In diesen Produkten wurde eine standardisierte Datenextraktion durchgeführt. Im Falle von Addenda und Neubewertungen nach Fristablauf oder bei neuen wissenschaftlichen Erkenntnissen wurden jeweils ausschließlich die aktuellsten für die Dossierbewertung relevanten Ergebnisse extrahiert und in die Auswertung einbezogen.

4.3 Informationssynthese und -analyse

Die Informationen wurden einer Informationssynthese und -analyse unterzogen. Dazu wurde zunächst der Datenpool beschrieben, indem die eingeschlossenen Dossierbewertungen charakterisiert werden (z. B. Indikation, Wirkstoff). Die für die einzelnen Endpunktkategorien tatsächlich vorliegenden Effektschätzungen wurden dann mit den erwarteten Effektstärken, die in der Methodik zur Ermittlung des Ausmaßes des Zusatznutzens zur Ableitung der Schwellenwerte herangezogen wurden, abgeglichen [2,4].

5 Ergebnisse

5.1 Ergebnis der Informationsbeschaffung

Zur Beantwortung der vorliegenden Fragestellung lagen 667 Dossierbewertungen und 266 Addenda aus den Jahren 2011 bis einschließlich 2022 vor. In 239 Dossierbewertungen einschließlich Addenda lag für 277 Fragestellungen für mindestens 1 Endpunkt eine Ausmaßbestimmung vor. Insgesamt wurden 1747 Endpunkte extrahiert. Abbildung 2 zeigt das Ergebnis der Informationsbeschaffung und -bewertung.

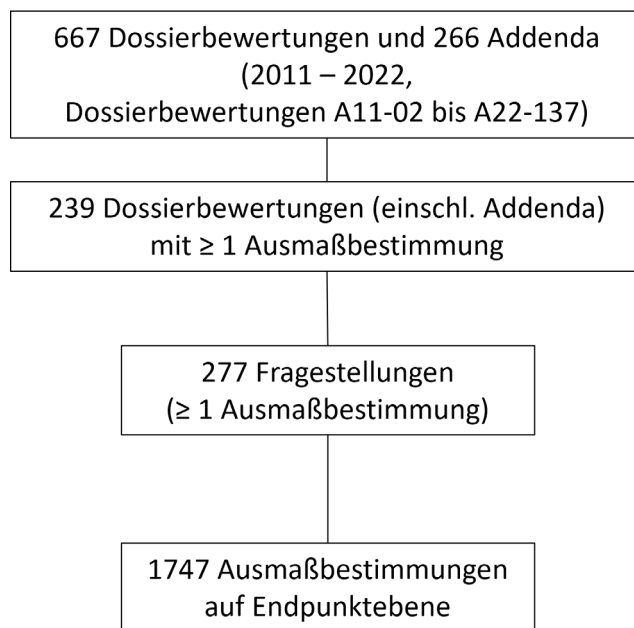


Abbildung 2: Ergebnis der Informationsbeschaffung und -bewertung

Charakterisierung der Endpunkte mit Ausmaßbestimmung

Tabelle 1 zeigt die Charakterisierung der Endpunkte mit Ausmaßbestimmung nach Evidenzart und Themengebiet.

Tabelle 1: Empirische Ergebnisse – Verteilung der Endpunkte mit Ausmaßbestimmung nach Evidenzart und Themengebiet

Charakteristikum Kategorie	Endpunkte mit Ausmaßbestimmung N = 1747
Evidenzart, n (%)	
direkter Vergleich	1557 (89,1)
Metaanalyse	174 (10,0)
indirekter Vergleich \geq 2 Studien ^a	16 (0,9)
Themengebiet, n (%)	
Diabetes mellitus	42 (2,4)
Herz-Kreislauf / Lunge	65 (3,7)
Infektiologie	77 (4,4)
Onkologie	1353 (77,5)
Psychiatrie / Neurologie	48 (2,8)
Sonstige ^b	162 (9,3)
a. einschließlich Metaanalysen	
b. Indikationen, die nicht in die übrigen genannten Kategorien eingeordnet werden können.	
n: Anzahl Endpunkte	

Von den insgesamt 1747 Ausmaßbestimmungen entstammen 89 % (n = 1557) einem direkten Vergleich einer Studie und 10 % (n = 174) Metaanalysen mit mehreren direkt vergleichenden Studien. Ein Großteil der Endpunkte mit Ausmaßbestimmungen (77,5 %) stammt aus Dossierbewertungen im Themengebiet Onkologie (siehe Tabelle 1). Die übrigen Endpunkte mit Ausmaßbestimmungen stammen aus Dossierbewertungen der Themengebiete Diabetes mellitus (2,4 %), Herz-Kreislauf / Lunge (3,7 %), Infektiologie (4,4 %), Psychiatrie / Neurologie (2,8 %) und Sonstige (9,3 %).

In Tabelle 2 und Tabelle 3 ist die Aufteilung der Endpunkte mit Ausmaßbestimmung auf die verschiedenen Zielgrößen- und Ausmaßkategorien dargestellt.

Tabelle 2: Empirische Ergebnisse – Verteilung der Endpunkte mit Ausmaßbestimmung nach Zielgrößenkategorien und Ausmaßkategorien

Zielgrößenkategorie	Ausmaßkategorie			Gesamt n (% ^a)
	Endpunkt mit Ausmaß „erheblich“ n (% ^a)	Endpunkt mit Ausmaß „beträchtlich“ n (% ^a)	Endpunkt mit Ausmaß „gering“ n (% ^a)	
Mortalität	45 (2,6)	42 (2,4)	19 (1,1)	106 (6,1)
Schwerwiegend / schwer / HQoL	350 (20,0)	282 (16,1)	279 (16,0)	911 (52,1)
nicht schwerwiegend / nicht schwer	– ^b	581 (33,3)	149 (8,5)	730 (41,8)

a. Prozentangaben beziehen sich auf die Grundgesamtheit aller extrahierten Endpunkte (N = 1747).
b. Gemäß den Allgemeinen Methoden des Instituts [2] wird für die Kategorie nicht schwerwiegend / nicht schwer kein Ausmaß „erheblich“ bestimmt.

n: Anzahl Endpunkte; HQoL: gesundheitsbezogene Lebensqualität

Tabelle 3: Empirische Ergebnisse – Verteilung der Endpunkte mit Ausmaßbestimmung nach Zielgrößen

Endpunktkategorie Zielgröße	Ausmaßkategorie		
	Endpunkt mit Ausmaß „erheblich“ n (% ^a)	Endpunkt mit Ausmaß „beträchtlich“ n (% ^a)	Endpunkt mit Ausmaß „gering“ n (% ^a)
Mortalität	45 (2,6)	42 (2,4)	19 (1,1)
Morbidität			
schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)	25 (1,4)	55 (3,1)	33 (1,9)
nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen)	– ^b	137 (7,8)	75 (4,3)
gesundheitsbezogene Lebensqualität	40 (2,3)	52 (3,0)	75 (4,3)
Nebenwirkungen			
schwerwiegende (bzw. schwere) Nebenwirkungen	285 (16,3)	175 (10,0)	171 (9,8)
nicht schwerwiegende (bzw. nicht schwere) Nebenwirkungen	– ^b	444 (25,4)	74 (4,2)

a. Prozentangaben beziehen sich auf die Grundgesamtheit aller extrahierten Endpunkte (N = 1747).
b. Gemäß den Allgemeinen Methoden des Instituts [2] wird für die Kategorie nicht schwerwiegend / nicht schwer kein Ausmaß „erheblich“ bestimmt.

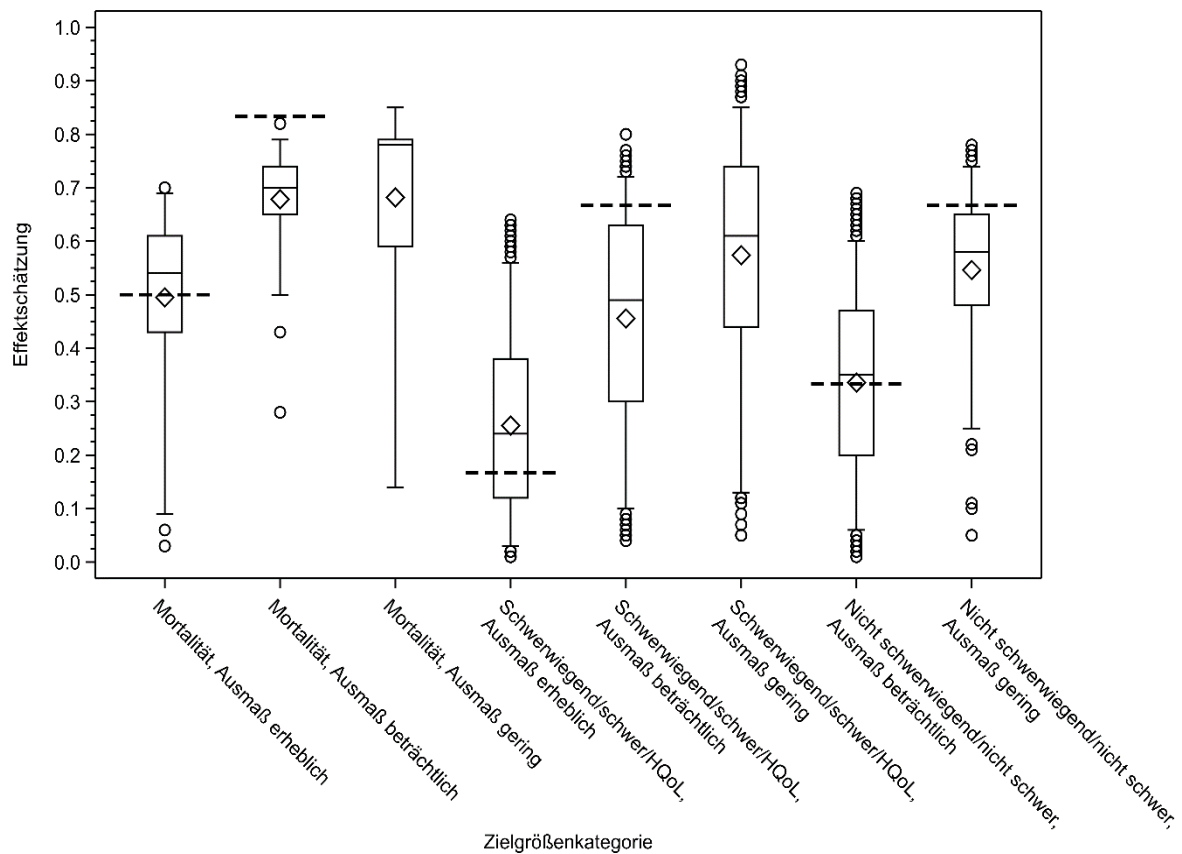
n: Anzahl Endpunkte

Aus den Tabellen lässt sich entnehmen, dass für alle möglichen Kombinationen aus Zielgrößen und Ausmaßkategorie jeweils Ausmaßbestimmungen auf Endpunktebene vorliegen, sodass eine Betrachtung aller Kombinationen möglich ist. Mehr als die Hälfte aller Ausmaßbestimmungen (52 % [n = 911]) erfolgte bei Endpunkten der Zielgrößenkategorie schwerwiegend / schwer / HQoL, wobei der überwiegende Teil aus Endpunkten zu Nebenwirkungen stammt.

5.2 Verteilung der tatsächlichen Effektschätzungen (alle Endpunkte)

Wie in Kapitel 1 beschrieben, wurden zur Ableitung der Ausmaße für die verschiedenen Zielgrößenkategorien unterschiedliche Schwellenwerte (obere Grenze des 95 %-Konfidenzintervalls) basierend auf den erwarteten Effektstärken festgelegt. Für die nachfolgenden Auswertungen erfolgte ein Abgleich der vorliegenden tatsächlichen Effektschätzungen aus den Dossierbewertungen mit den für die Schwellenwertbestimmung erwarteten Effektstärken.

Die Verteilung der tatsächlichen, aus den Nutzenbewertungen extrahierten Effektschätzungen aufgeteilt nach Zielgrößen- und Ausmaßkategorie ist in Abbildung 3 und Tabelle 4 dargestellt.



Erläuterung Boxplot:

Die Box entspricht dem Bereich des 25 %- bis 75 %-Quantils, die durchgezogene horizontale Linie innerhalb der Box dem Median, die Raute innerhalb der Box dem Mittelwert. Die Linien außerhalb der Box entsprechen dem 5 %- und 95 %-Quantil. Werte außerhalb dieser Quantile sind als Kreise dargestellt.

Die gestrichelten horizontalen Linien entsprechen den jeweils erwarteten Effektstärken.

Abbildung 3: Verteilung der vorliegenden tatsächlichen Effektschätzungen nach Zielgrößen- und Ausmaßkategorie – Auswertung aller Endpunkte

Tabelle 4: Verteilung der vorliegenden tatsächlichen Effektschätzungen (alle Indikationen)

Zielgrößenkategorie	Ausmaßkategorie		
	Endpunkt mit Ausmaß „erheblich“	Endpunkt mit Ausmaß „beträchtlich“	Endpunkt mit Ausmaß „gering“
Mortalität	n = 45	n = 42	n = 19
erwartete Effektstärke (RR) ^a	0,50	0,83	– ^b
tatsächliche Effektschätzung ^c			
Median [Q1; Q3]	0,54 (0,43; 0,61)	0,70 (0,65; 0,74)	0,78 (0,59; 0,79)
Mittelwert	0,49	0,68	0,68
Schwerwiegend / schwer / HQoL	n = 350	n = 282	n = 279
erwartete Effektstärke (RR) ^a	0,17	0,67	– ^b
tatsächliche Effektschätzung ^c			
Median [Q1; Q3]	0,24 (0,12; 0,38)	0,49 (0,30; 0,63)	0,61 (0,44; 0,74)
Mittelwert	0,26	0,46	0,57
nicht schwerwiegend / nicht schwer		n = 581	n = 149
erwartete Effektstärke (RR) ^a	– ^d	0,33	0,67
tatsächliche Effektschätzung ^c			
Median [Q1; Q3]	k. A.	0,35 (0,20; 0,47)	0,58 (0,48; 0,65)
Mittelwert	k. A.	0,34	0,55
<p>a. ursprünglich entwickelt auf Basis des relativen Risikos, jedoch bezogen auf alle relativen Effektmaße (HR, RR, Peto-OR und Rate Ratio)</p> <p>b. Gemäß den Allgemeinen Methoden des Instituts [2] wird jeglicher zum üblichen Irrtumsniveau 5 % statistisch signifikante Effekt zumindest als geringer Zusatznutzen eingestuft.</p> <p>c. umfasst die Effektschätzungen HR, RR, Peto-OR und Rate Ratio</p> <p>d. Gemäß den Allgemeinen Methoden des Instituts [2] wird für die Kategorie nicht schwerwiegend / nicht schwer kein Ausmaß „erheblich“ bestimmt.</p> <p>HQoL: gesundheitsbezogene Lebensqualität; HR: Hazard Ratio; k. A.: keine Angaben; n: Anzahl Endpunkte; OR: Odds Ratio; Q1: 1. Quartil; Q3: 3. Quartil; RR: relatives Risiko</p>			

In der Gesamtbetrachtung zeigen sich je nach Zielgrößen- und Ausmaßkategorie sehr unterschiedliche Verteilungen der tatsächlichen Effektschätzungen (siehe Abbildung 3).

Vergleich der verschiedenen Ausmaßkategorien: Je höher die Kategorie, desto größer der Effekt

Beim Vergleich der verschiedenen Ausmaßkategorien zeigt sich, wie auch erwartet, dass die tatsächliche Effektschätzung mit steigender Ausmaßkategorie größer wird. Dies drückt sich auch jeweils mit einem niedrigeren Wert der medianen bzw. mittleren Effektschätzung in den höheren Ausmaßkategorien aus (siehe Tabelle 4).

Effekte in der höchsten Ausmaßkategorie häufig kleiner als bei der Entwicklung der Methodik festgelegt

Die Werte der tatsächlichen Effektschätzungen in den jeweiligen höchsten Ausmaßkategorien („erheblich“ für die Zielgrößenkategorien Mortalität sowie schwerwiegend / schwer / HQoL; „beträchtlich“ für die Zielgrößenkategorie nicht schwerwiegend / nicht schwer) befinden sich im Median leicht oberhalb der erwarteten Effektstärken: Für die Zielgrößenkategorien Mortalität und schwerwiegend / schwer / HQoL liegen die Werte der jeweiligen tatsächlichen medianen Effektschätzung bei 0,54 (erwarteter Effekt: 0,50) bzw. 0,24 (erwarteter Effekt: 0,17) für das Ausmaß „erheblich“. Für die Zielgrößenkategorie nicht schwerwiegend / nicht schwer liegen die Werte der tatsächlichen medianen Effektschätzung bei 0,35 (erwarteter Effekt: 0,33) für das Ausmaß „beträchtlich“ (siehe Abbildung 3 und Tabelle 4). Damit haben mehr als die Hälfte der Endpunkte, für die das höchstmögliche Ausmaß abgeleitet wurde, einen kleineren Effekt als bei der Entwicklung der Methodik für das jeweilige Ausmaß festgelegt.

Es zeigt sich zudem, dass die Werte der tatsächlichen Effektschätzungen in den jeweiligen zweithöchsten Ausmaßkategorien („beträchtlich“ für die Zielgrößenkategorien Mortalität sowie schwerwiegend / schwer / HQoL; „gering“ für die Zielgrößenkategorie nicht schwerwiegend / nicht schwer) fast vollständig bzw. größtenteils unterhalb der erwarteten Effektstärken liegen: Für die Zielgrößenkategorien Mortalität und schwerwiegend / schwer / HQoL liegen die Werte der jeweiligen medianen Effektschätzung bei 0,70 (erwarteter Effekt: 0,83) bzw. 0,49 (erwarteter Effekt: 0,67) für das Ausmaß „beträchtlich“. Für die Zielgrößenkategorie nicht schwerwiegend / nicht schwer liegen die Werte der medianen Effektschätzung bei 0,58 (erwarteter Effekt: 0,67) für das Ausmaß „gering“ (siehe Abbildung 3 und Tabelle 4). Damit werden die erwarteten Effektstärken in der ganz überwiegenden Anzahl der Endpunkte für diese Ausmaßkategorien erreicht.

Für die beiden Zielgrößenkategorien Mortalität und schwerwiegend / schwer / HQoL liegt zum Erreichen eines geringen Zusatznutzens keine Annahme im Sinne einer verschobenen Hypothesengrenze vor, vielmehr wird jeglicher statistisch signifikante Effekt zum üblichen Irrtumsniveau 5 % als „gering“ eingestuft.

Insgesamt zeigt sich gerade in der jeweils höchsten Ausmaßkategorie („erheblich“ für Mortalität sowie schwerwiegend / schwer / HQoL; „beträchtlich“ für nicht schwerwiegend / nicht schwer), dass die tatsächlichen Effektschätzungen in mehr als der Hälfte der Ausmaßbestimmungen kleiner sind als bei der Entwicklung der Methodik für das Ausmaß festgelegt.

Untersuchung von Clustereffekten

In den vorliegenden Auswertungen wurden alle Effektschätzungen, die zu einer Ausmaßbestimmung geführt haben, gleichermaßen berücksichtigt. Dabei wird außer Acht gelassen, dass Effektschätzungen von Endpunkten aus der gleichen Studie, möglicherweise korreliert sind. Das könnte zu Verzerrungen der Ergebnisse führen. Um zu untersuchen, ob der sogenannte Clustereffekt eine relevante Rolle spielt, wurden Sensitivitätsanalysen durchgeführt. Da der überwiegende Anteil an Dossierbewertungen auf Ergebnissen aus nur einer Studie beruht, wurde für diese Analysen jeweils nur ein Endpunkt pro Dossierbewertung zufällig ausgewählt und ausgewertet. Der Abgleich zwischen den Haupt- und Sensitivitätsanalysen zeigte keine nennenswerten Unterschiede. Zum Vergleich ist die Sensitivitätsanalyse zu Abbildung 3 in Anhang A dargestellt.

5.3 Detaillierte Betrachtung der Verteilung der tatsächlichen Effektschätzungen – nach Zielgrößen- und Ausmaßkategorie

Auf Basis der Ergebnisse zur Verteilung der tatsächlichen Effektschätzungen in Abbildung 3 und Tabelle 4 wurden die Ergebnisse zu den Zielgrößen- und Ausmaßkategorien Mortalität und schwerwiegend / schwer / HQoL mit dem Ausmaß „erheblich“ und nicht schwerwiegend / nicht schwer mit dem Ausmaß „beträchtlich“ detaillierter betrachtet, da hier die größten Abweichungen im Sinne kleinerer Effektschätzungen von der jeweils erwarteten Effektstärke vorliegen. Dafür wurde für diese Zielgrößen- und Ausmaßkategorien in einem weiteren Schritt auch betrachtet, ob sich die Ergebnisse bei einer differenzierten Betrachtung nach Populationsgröße, Aussagesicherheit und Themengebieten unterscheiden. Bei der differenzierten Betrachtung nach Populationsgröße zeigen sich jeweils unterschiedliche Ergebnisse in den Zielgrößen, die in den nachfolgenden Abschnitten 5.3.1, 5.3.2 und 5.3.3 beschrieben sind. Bei der differenzierten Betrachtung nach Aussagesicherheit und Themengebieten zeigen sich keine relevanten Unterschiede in der Verteilung der tatsächlichen Effektschätzungen (siehe Anhang B und Anhang C). Dabei ist auch zu berücksichtigen, dass teilweise nur relativ wenige Ausmaßbestimmungen vorliegen wie beispielsweise für die Zielgröße Mortalität in nicht onkologischen Themengebieten oder auch für verschiedene Betrachtungen differenziert nach Aussagesicherheit.

5.3.1 Zielgrößenkategorie Mortalität, Ausmaßkategorie „erheblich“

Mit einem Median von 0,54 liegen die Werte der tatsächlichen Effektschätzungen in der Zielgrößenkategorie Mortalität mit dem Ausmaß „erheblich“ in mehr als der Hälfte der Fälle oberhalb der erwarteten Effektstärke von 0,50. Dies spiegelt sich auch in Abbildung 4 wider, in der die Effektschätzungen und dazugehörigen Konfidenzintervalle aller Endpunkte dieser Zielgrößenkategorie dargestellt sind.

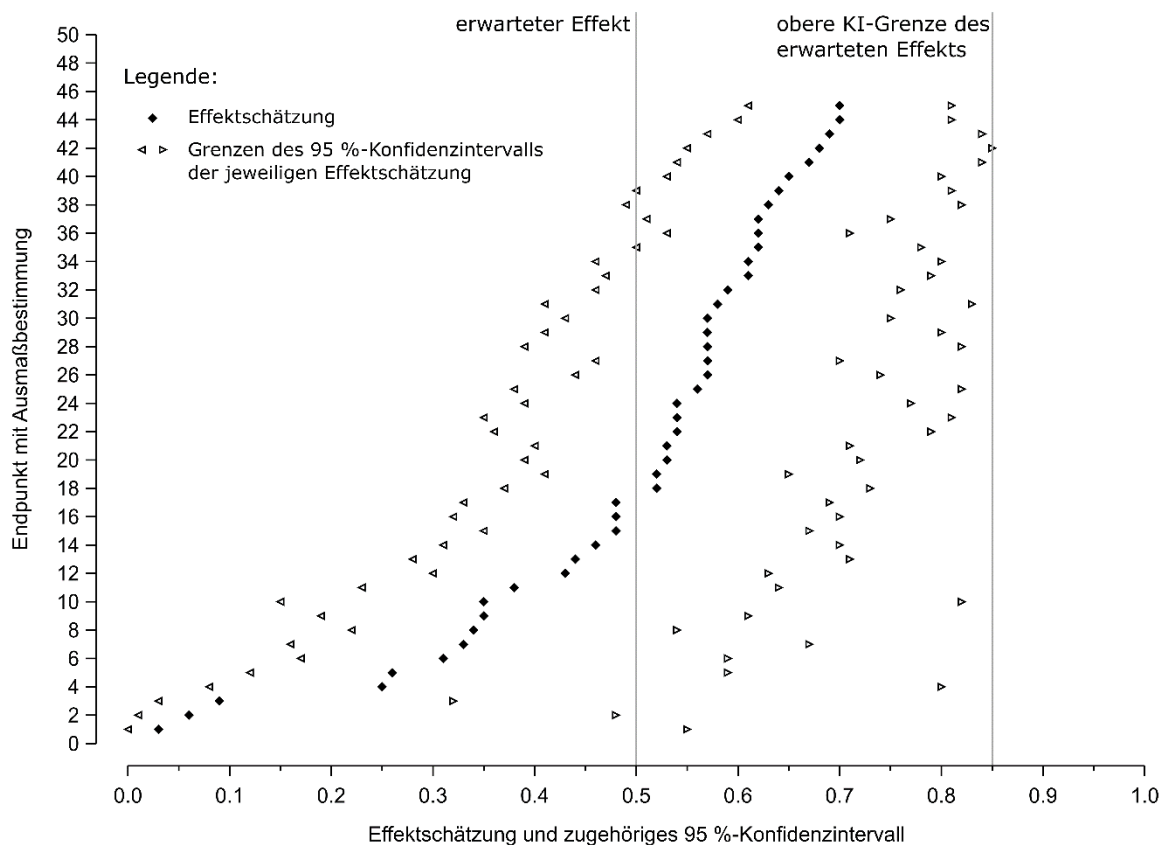


Abbildung 4: Effektschätzungen aller 45 Ausmaßbestimmungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ – Verteilung der Effektschätzungen und der zugehörigen Konfidenzintervalle (95 %-KI) in aufsteigender Reihenfolge

Etwa 62 % der Werte der tatsächlichen Effektschätzung liegen oberhalb der erwarteten Effektstärke (0,50), zeigen also kleinere Effekte als bei der Entwicklung der Methodik für diese Ausmaßkategorie festgelegt. Zudem befindet sich für ca. 29 % dieser Endpunkte zusätzlich auch die untere Konfidenzintervallgrenze oberhalb der erwarteten Effektstärke. Dies bedeutet, dass der erwartete Effekt von 0,50 in diesen Fällen nicht nur größer ist, sondern auch außerhalb des 95 %-Konfidenzintervalls der tatsächlichen Effektschätzung liegt.

In Abbildung 5 ist die Verteilung der Effektschätzungen in Abhängigkeit von der Populationsgröße dargestellt.

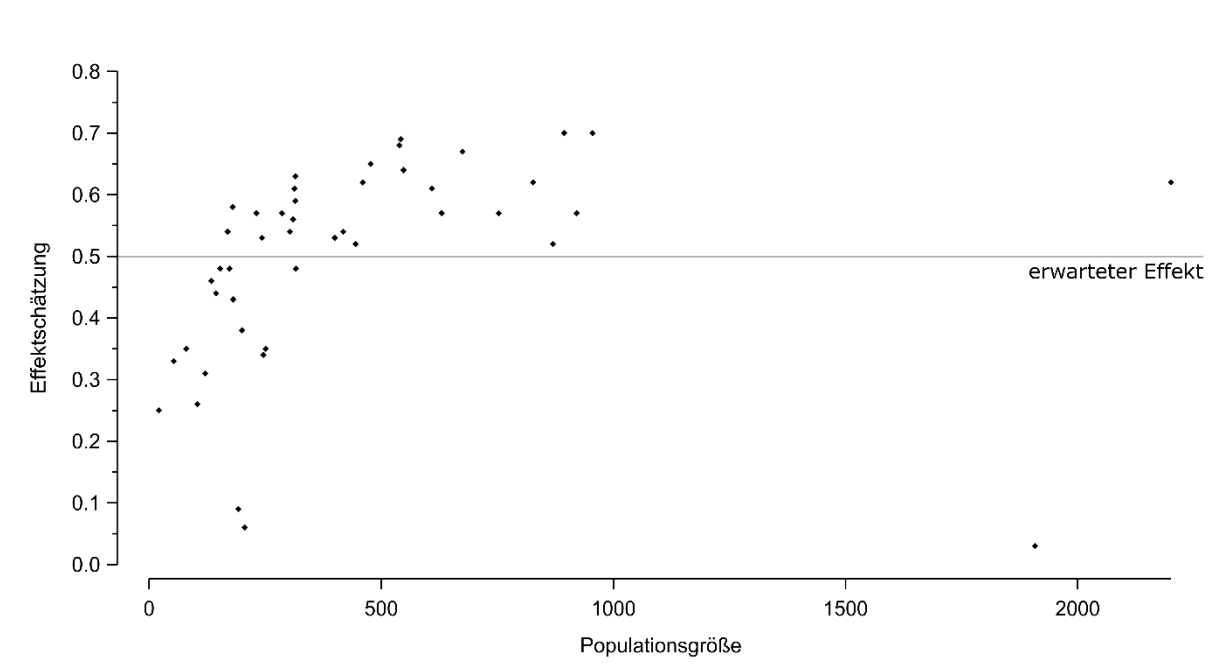


Abbildung 5: Effektschätzungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ – Verteilung nach Populationsgröße

In Abbildung 5 ist erkennbar, dass mit zunehmender Populationsgröße (Anzahl ausgewerteter Patientinnen und Patienten) die Werte der tatsächlichen Effektschätzungen tendenziell steigen: Je größer die Anzahl ausgewerteter Patientinnen und Patienten, desto kleiner wird die tatsächliche Effektschätzung in dieser Ausmaßkategorie. Dies erscheint plausibel, da die Populationsgröße Einfluss auf die Präzision der Effektschätzung einer Studie hat. Deshalb unterschreitet in größeren Populationen die obere Grenze des 95 %-Konfidenzintervalls bei kleineren Effekten häufiger den Schwellenwert für eine Ausmaßkategorie. Dieser potenzielle Zusammenhang zwischen Populationsgröße und Effektschätzung ist auch in den Ergebnissen der Subgruppenpopulationen (Ausmaßbestimmung nach Subgruppenmerkmal wegen Subgruppeneffekt) und Gesamtpopulationen bei der Verteilung der tatsächlichen Effektschätzungen erkennbar (siehe Anhang D), jedoch nur für diese Zielgrößenkategorie.

5.3.2 Zielgrößenkategorie schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität, Ausmaßkategorie „erheblich“

In der Zielgrößenkategorie schwerwiegend / schwer / HQoL mit dem Ausmaß „erheblich“ liegen die Werte der tatsächlichen Effektschätzungen mit einem Median von 0,24 in mehr als der Hälfte der Fälle oberhalb der erwarteten Effektstärke von 0,17. In Abbildung 6 sind die Effektschätzungen und dazugehörigen Konfidenzintervalle aller Endpunkte dieser Zielgrößenkategorie dargestellt.

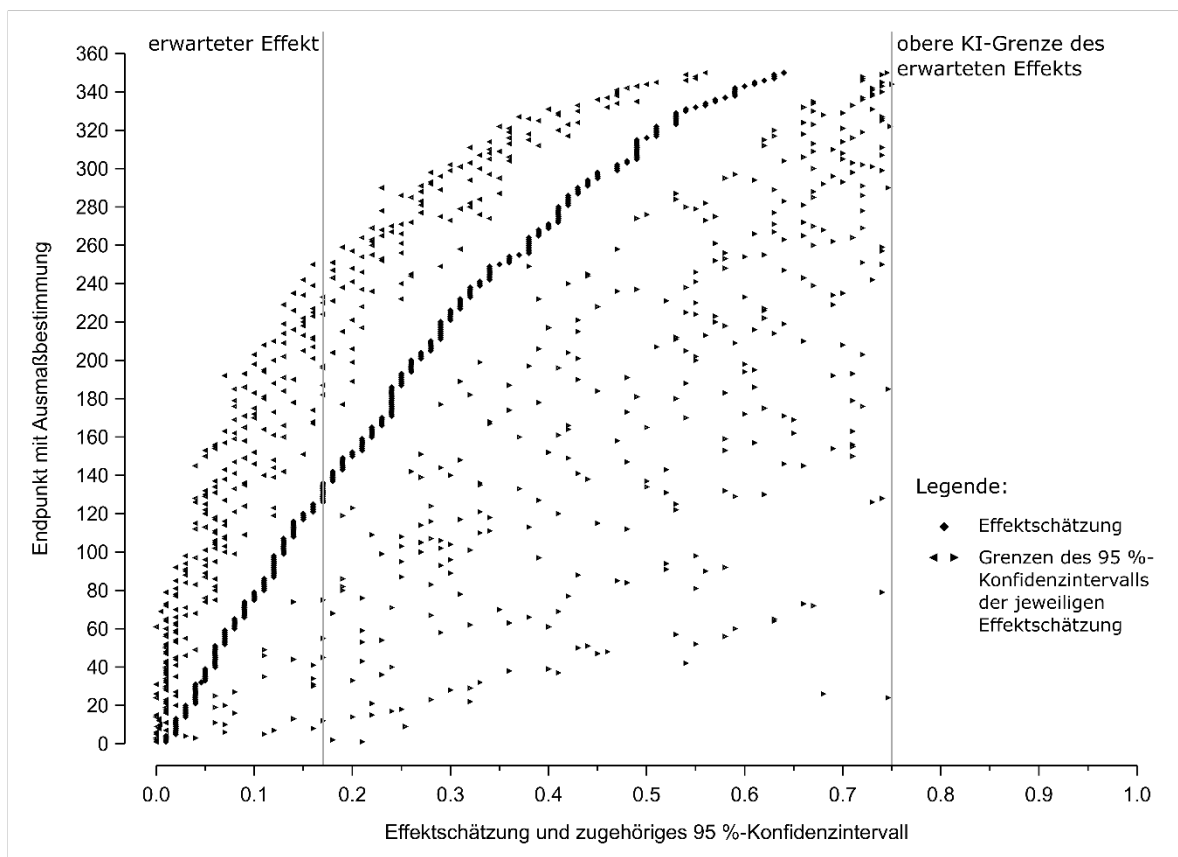


Abbildung 6: Effektschätzungen aller 350 Ausmaßbestimmungen der Endpunkte der Zielgrößenkategorie schwerwiegend / schwer / HQoL mit Ausmaß „erheblich“ – Verteilung nach Effektschätzungen und der zugehörigen Konfidenzintervalle in aufsteigender Reihenfolge

Etwa 61 % der Werte der tatsächlichen Effektschätzungen dieser Zielgrößenkategorie liegen oberhalb der erwarteten Effektstärke (0,17), zeigen also kleinere Effekte als bei der Entwicklung der Methodik für diese Ausmaßkategorie festgelegt. Ebenso liegt für mehr als die Hälfte dieser Endpunkte zusätzlich auch die untere Grenze des Konfidenzintervalls oberhalb der erwarteten Effektstärke.

In Abbildung 7, Abbildung 8 und Abbildung 9 ist die Verteilung der Effektschätzungen je Zielgröße in Abhängigkeit von der Populationsgröße dargestellt.

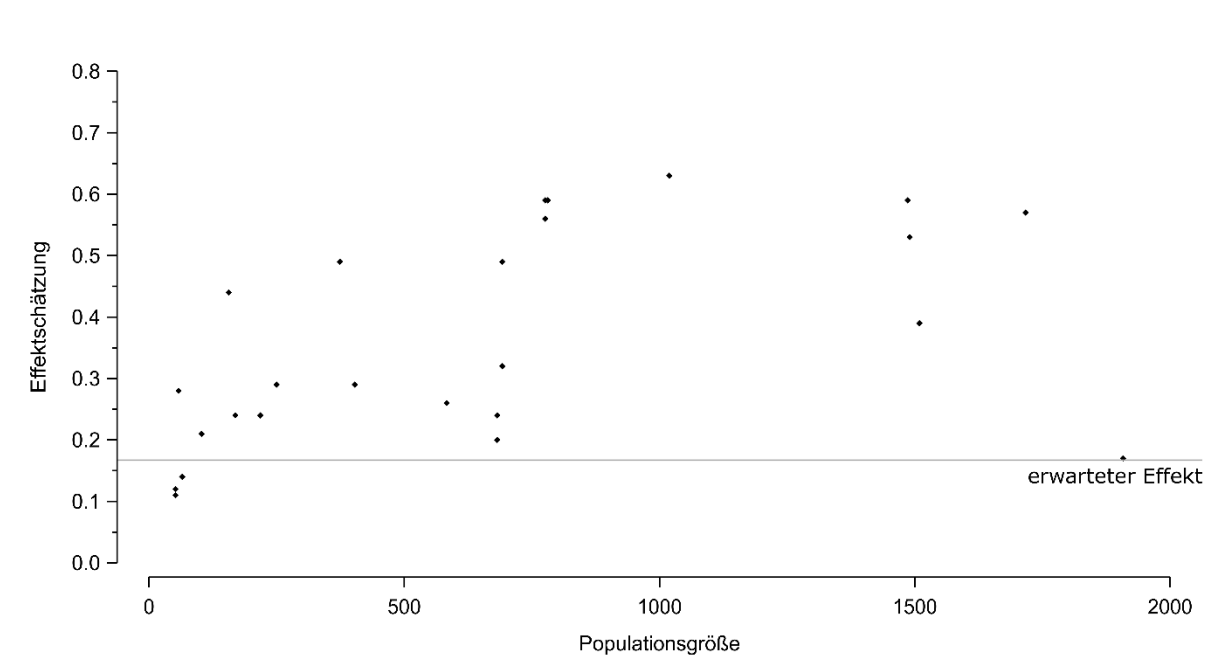


Abbildung 7: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Symptome mit Ausmaß „erheblich“ – Verteilung nach Populationsgröße

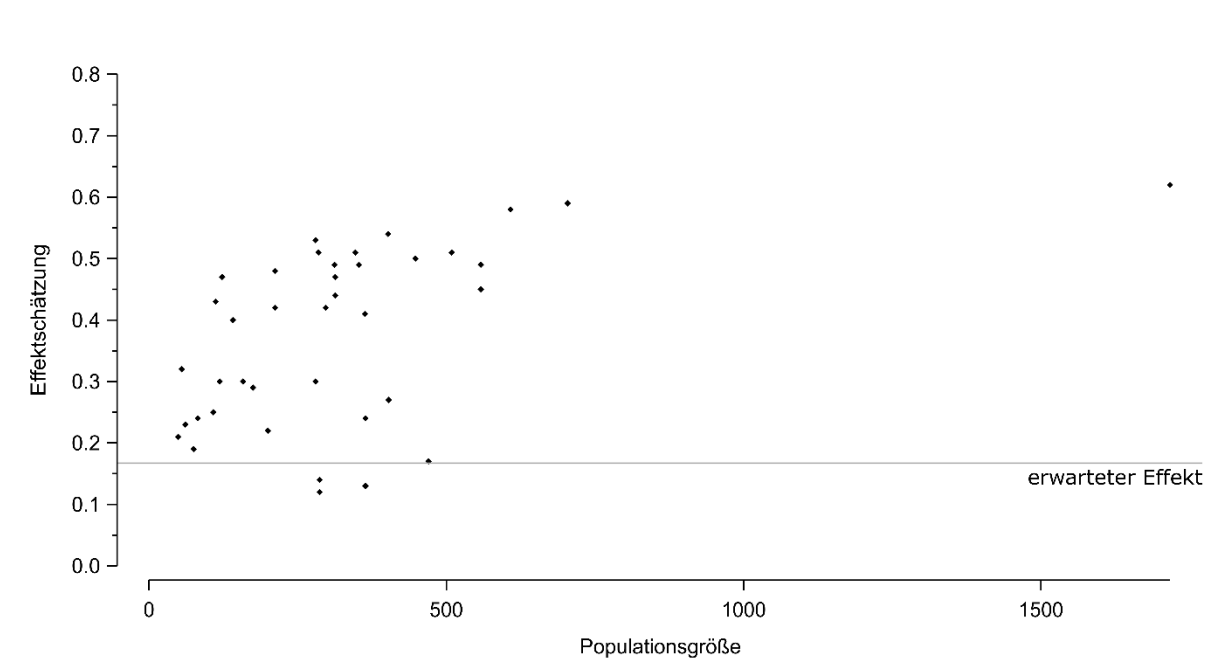


Abbildung 8: Effektschätzungen der Endpunkte der Zielgröße HQoL mit Ausmaß „erheblich“ – Verteilung nach Populationsgröße

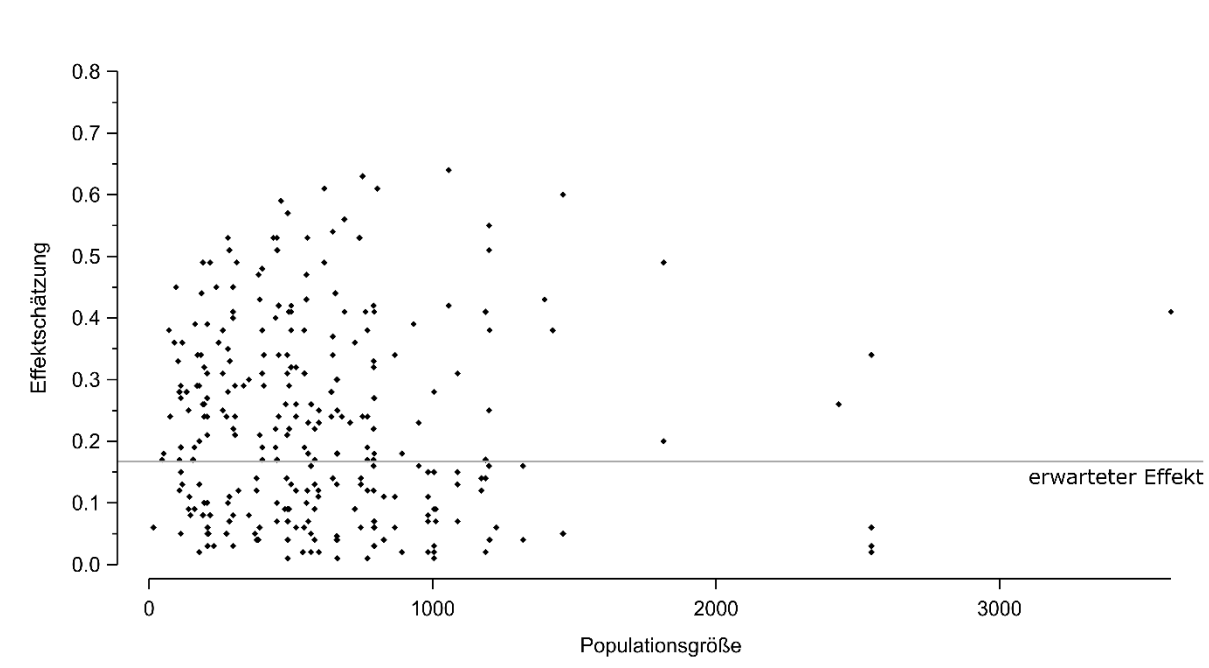


Abbildung 9: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende/schwere Nebenwirkungen mit Ausmaß „erheblich“ – Verteilung nach Populationsgröße

In Abbildung 7 und Abbildung 8 ist erkennbar, dass auch für die Zielgrößen schwerwiegende / schwere Symptome und gesundheitsbezogene Lebensqualität mit zunehmender Populationsgröße (Anzahl ausgewerteter Patientinnen und Patienten) die Werte der tatsächlichen Effektschätzungen tendenziell steigen: Je größer die Anzahl ausgewerteter Patientinnen und Patienten, desto kleiner wird die tatsächliche Effektschätzung in dieser Ausmaßkategorie. Eine Abhängigkeit der Effekte von der Populationsgröße ist jedoch für die Zielgröße schwerwiegende / schwere Nebenwirkungen nicht erkennbar (siehe Abbildung 9). Diese Tendenz ist ebenfalls nicht bei der Betrachtung der Ergebnisse der Subgruppenpopulationen (Ausmaßbestimmung nach Subgruppenmerkmal wegen Subgruppeneffekt) und Gesamtpopulationen bei der Verteilung der tatsächlichen Effektschätzungen erkennbar (siehe Anhang D).

5.3.3 Zielgrößenkategorie nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen, Ausmaßkategorie „beträchtlich“

In der Zielgrößenkategorie nicht schwerwiegend / nicht schwer mit dem Ausmaß „beträchtlich“ liegen die Werte der tatsächlichen Effektschätzungen mit einem Median von 0,35 in mehr als der Hälfte der Fälle oberhalb der erwarteten Effektstärke von 0,33. In Abbildung 10 sind die Effektschätzungen und dazugehörigen Konfidenzintervalle aller Endpunkte dieser Zielgrößenkategorie dargestellt.

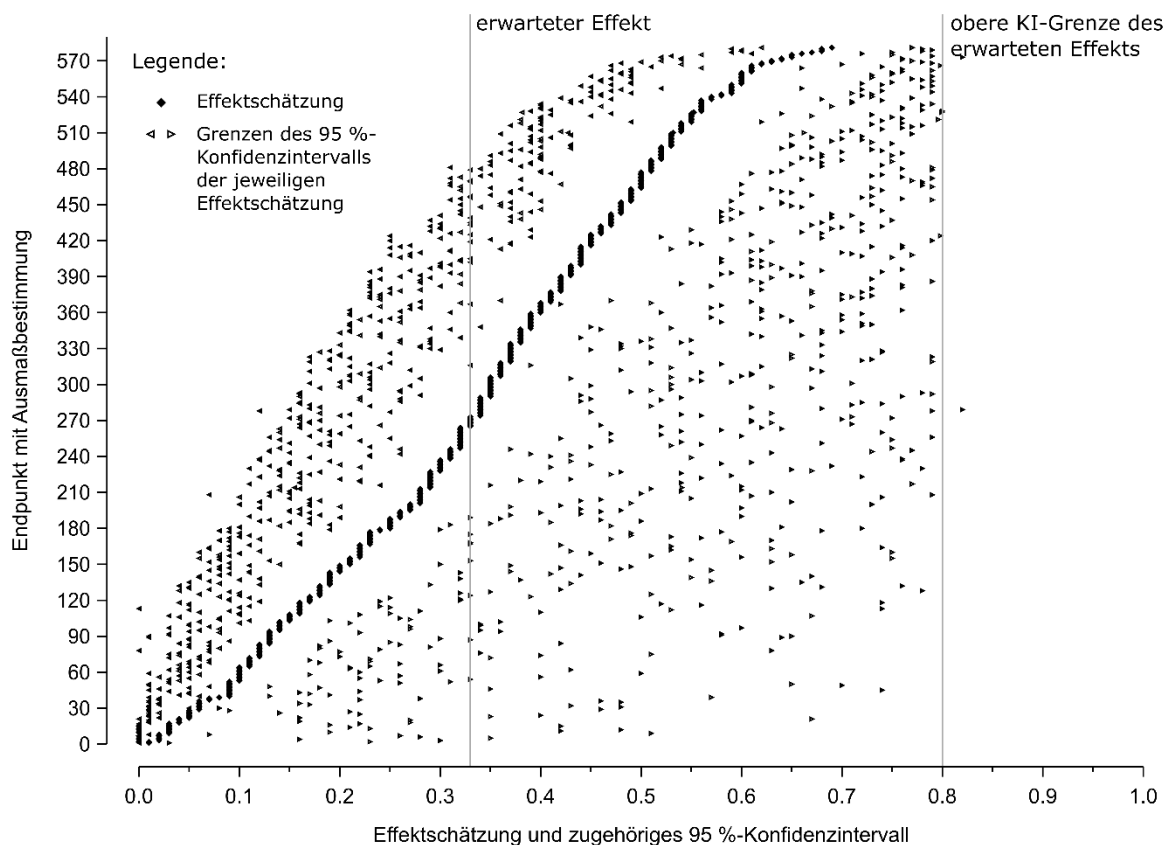


Abbildung 10: Effektschätzungen aller 581² Ausmaßbestimmungen der Endpunkte der Zielgrößenkategorie nicht schwerwiegend / nicht schwer mit Ausmaß „beträchtlich“ – Verteilung nach Effektschätzungen und der zugehörigen Konfidenzintervalle in aufsteigender Reihenfolge

Etwa 53 % der Werte der tatsächlichen Effektschätzungen liegen oberhalb der erwarteten Effektstärke (0,33), zeigen also kleinere Effekte als bei der Entwicklung der Methodik für diese Ausmaßkategorie festgelegt. Ebenso ist für 44 % dieser Endpunkte zusätzlich auch die untere Grenze des Konfidenzintervalls oberhalb der erwarteten Effektstärke.

In Abbildung 11 und Abbildung 12 ist die Verteilung der Effektschätzungen je Zielgröße in Abhängigkeit von der Populationsgröße dargestellt.

² Es ist anzumerken, dass für 2 Endpunkte mit Ausmaß „beträchtlich“ die obere Grenze des 95 %-Konfidenzintervalls überschritten wurde, wie aus der Abbildung erkennbar ist. Dies ist darin begründet, dass in diesen beiden Fällen bei der Ausmaßbestimmung weitere inhaltliche Aspekte berücksichtigt wurden.

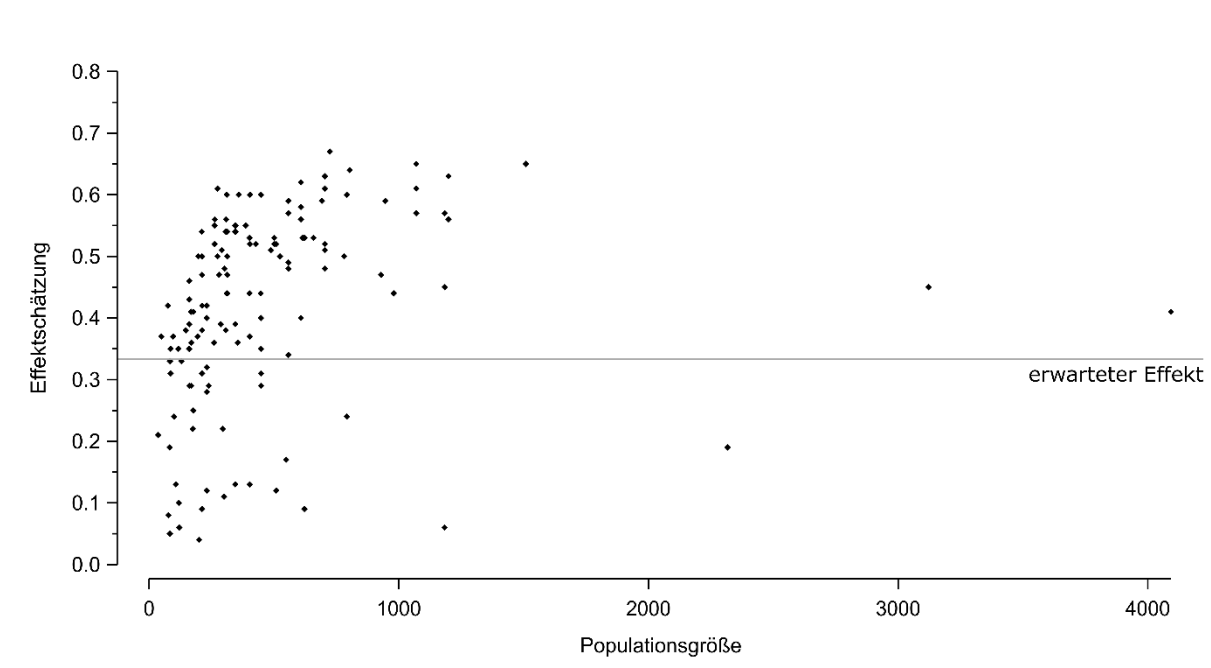


Abbildung 11: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Symptome mit Ausmaß „beträchtlich“ – Verteilung nach Populationsgröße

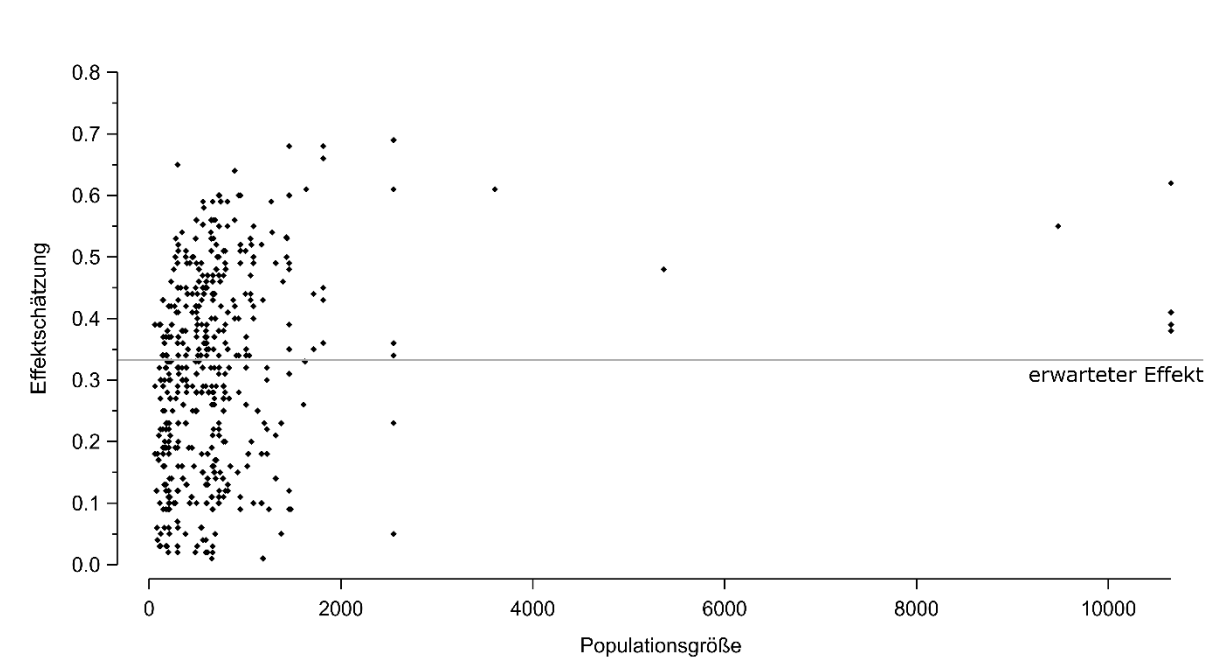


Abbildung 12: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Nebenwirkungen mit Ausmaß „beträchtlich“ – Verteilung nach Populationsgröße

Anders als bei der Zielgrößenkategorie Mortalität und den Zielgrößen schwerwiegende / schwere Symptome und gesundheitsbezogene Lebensqualität ist in dieser Kategorie nicht so klar zu erkennen, dass der Effekt mit zunehmender Populationsgröße kleiner wird (siehe Abbildung 12).

6 Diskussion

Datenbasis

Mit 667 abgeschlossenen Dossierbewertungen und 266 zugehörigen Addenda bis 2022 liegt mehr als 13 Jahre nach Einführung des AMNOG eine umfassende Datengrundlage für eine empirische Auswertung der Ausmaßableitungen auf Endpunktebene vor. Diese ermöglicht die Betrachtung aller Kombinationen aus Zielgrößen und Ausmaßkategorien.

Ergebnisse

Für die Auswertung der tatsächlichen Effektschätzungen zeigen sich je nach Zielgrößen- und Ausmaßkategorie sehr unterschiedliche Verteilungen der tatsächlichen Effektschätzungen im Verhältnis zur erwarteten Effektstärke. Besonders auffällig ist hier, dass jeweils in den höchsten Ausmaßkategorien die Werte der tatsächlichen Effektschätzungen in allen Zielgrößenkategorien in mehr als der Hälfte der Fälle oberhalb der erwarteten Effektstärken liegen, was in diesen Fällen bedeutet, dass die tatsächliche Effektschätzung kleiner ist als bei der Entwicklung der Methodik für die jeweilige Ausmaßkategorie festgelegt. Damit wurde für diese Fälle mit einer Ausmaßbestimmung auf Basis der oberen Konfidenzintervallgrenze die erwartete Effektstärke nicht erreicht. Die Werte der Effektschätzungen in den jeweiligen zweithöchsten Ausmaßkategorien („beträchtlich“ bzw. „gering“) liegen hingegen fast vollständig bzw. größtenteils unterhalb der erwarteten Effektstärken, was bedeutet, dass die erwartete Effektstärke auf Basis der angewandten Methodik erreicht oder übertroffen wird.

Neben der Effektstärke und dem abgeleiteten Ausmaß wurden verschiedene weitere Daten betrachtet, um mögliche weitere Faktoren zu identifizieren, für die sich die Verteilung der Effektschätzungen innerhalb der Zielgrößenkategorien sowie zielgrößenübergreifend unterscheidet. Hierfür wurden die Themengebiete, die Aussagesicherheit des Endpunkts (Anhaltspunkt, Hinweis, Beleg) und die Populationsgröße genauer betrachtet.

Die Verteilungen der Effektschätzungen unterscheiden sich nicht zwischen den Themengebieten der Dossierbewertungen, aus denen die Endpunkte extrahiert wurden. Dabei ist allerdings zu berücksichtigen, dass mehr als 75 % aller Endpunkte aus Dossierbewertungen onkologischer Indikationen stammen, die damit die Gesamtauswertung dominieren. Die Datenbasis zu nicht onkologischen Indikationen ist damit deutlich kleiner. Auch die Betrachtung der Aussagesicherheit zeigt keine Unterschiede in der Verteilung der Effektschätzungen.

Aus den untersuchten Faktoren hat sich innerhalb der Zielgrößenkategorien ausschließlich für die Populationsgröße ein möglicher Zusammenhang mit der Verteilung der tatsächlichen Effektgrößen gezeigt. Diese hat in den betrachteten Bewertungen eine Spannweite zwischen 15 und 18 140 eingeschlossenen Patientinnen und Patienten. Dabei zeichnet sich ab, dass mit steigender Populationsgröße, aus der die Effektschätzungen abgeleitet werden, die

Effektschätzung kleiner wird, während die Effekte gleichzeitig präziser werden. Dies ist insofern plausibel, als die Größe der Population einen bestimmenden Faktor für die Präzision einer Studie darstellt. Für die Zielgrößenkategorie Mortalität mit dem Ausmaß „erheblich“ ist dies in Abbildung 4 an den schmaleren Konfidenzintervallen bei kleineren Effekten und in Abbildung 5 an der Verteilung der Effektschätzung in Abhängigkeit von der Populationsgröße zu erkennen. Ein ähnlicher Zusammenhang ist für die Zielgrößen schwerwiegende / schwere Symptome und gesundheitsbezogene Lebensqualität zu sehen, jedoch ist der Zusammenhang in diesen Zielgrößen deutlich geringer ausgeprägt als für die Mortalität (Abbildung 6, Abbildung 7 und Abbildung 8). Dieser Zusammenhang bleibt auch sichtbar, wenn diese getrennt nach Subgruppenergebnissen und Ergebnissen aus Gesamtpopulationen betrachtet werden (siehe Anhang D). Für die Zielgröße schwerwiegende / schwere Nebenwirkungen ist dieser Effekt allerdings nicht zu beobachten (siehe Abbildung 9).

In diesem Zusammenhang ist zu beachten, dass sich die Zielgrößenkategorien in ihrer Zusammensetzung in Zielgrößen und Endpunkten stark unterscheiden. In die Zielgrößenkategorie Mortalität fällt ausschließlich der Endpunkt Gesamtüberleben bzw. Gesamtmortalität, während in die beiden anderen Zielgrößenkategorien alle Endpunkte der Endpunktkategorien Morbidität, gesundheitsbezogene Lebensqualität und Nebenwirkungen gezählt werden. Die Aufteilung auf diese beiden Zielgrößenkategorien ist nur durch den Schweregrad definiert. Während die Ergebnisse der Kategorie Mortalität somit eine Homogenität in Bezug auf die Art des Endpunkts aufweisen, ergibt sich für die beiden anderen Zielgrößenkategorien, definitionsgemäß, eine heterogene Zusammensetzung der Endpunkte. Aus diesem Grund wurden für die Zielgrößenkategorie schwerwiegende / schwere Symptome bzw. Nebenwirkungen und gesundheitsbezogene Lebensqualität separate Auswertungen nach Zielgrößen durchgeführt (vgl. Abbildung 7, Abbildung 8 und Abbildung 9). Diese zeigen, dass insbesondere die Zielgröße Nebenwirkungen sehr unterschiedliche Verteilungen der tatsächlichen Effektschätzungen auch im Verhältnis zur erwarteten Effektstärke aufweist, die die Heterogenität von Nebenwirkungen widerspiegeln.

Zusammenfassend zeigen sich innerhalb der Zielgrößen- und Ausmaßkategorien sehr unterschiedliche Verteilungen der tatsächlichen Effektschätzungen im Verhältnis zur erwarteten Effektstärke, wobei vor allem in den höchstmöglichen Ausmaßkategorien die Werte der tatsächlichen Effektschätzungen in vielen Fällen deutlich oberhalb der erwarteten Effektstärken liegen (Abbildung 3). Die Ausmaßbestimmung auf Basis von oberen Konfidenzintervallgrenzen als Schwellenwerten führt damit in der jeweils höchsten Ausmaßkategorie dazu, dass der erwartete Effekt in mehr als der Hälfte der Fälle nicht erreicht wird. Es ist daher mit Blick auf eine potenzielle Weiterentwicklung der Methodik zu diskutieren, ob weitere oder andere Kriterien herangezogen werden sollten. So zieht beispielsweise die European Society for Medical Oncology in ihrer Magnitude of Clinical Benefit Scale (ESMO-MCBS) bezogen auf onkologische Studien andere Kriterien heran und

berücksichtigt neben absoluten Kriterien die untere Konfidenzintervallgrenze als Schwelle [5,6]. Für einen hohen Anteil der in Abschnitt 5.3 beschriebenen und in Abbildung 4, Abbildung 6 und Abbildung 10 dargestellten Endpunkte würde beispielsweise nicht das höchstmögliche Ausmaß abgeleitet werden, wenn zusätzlich die Anforderung bestünde, dass mindestens die untere KI-Grenze unterhalb des erwarteten Effekts liegt.

7 Fazit

Das Ziel der vorliegenden Untersuchung war es, auf Basis der Ergebnisse aus abgeschlossenen Dossierbewertungen für die verschiedenen Endpunktkategorien empirisch zu prüfen, inwieweit durch die Anwendung der im Jahr 2011 entwickelten Ausmaßmethodik die in der Entwicklung der Methodik festgelegten gewünschten Effektgrößen tatsächlich erreicht werden. Zu diesem Zweck erfolgt ein Abgleich der tatsächlichen Effektschätzungen aus den Dossierbewertungen mit den für die Schwellenwertbestimmung festgelegten gewünschten Effektgrößen.

Mit insgesamt 1747 Ausmaßbestimmungen auf Endpunktebene aus 667 Dossierbewertungen und 266 Addenda der Jahre 2011 bis 2022 liegt eine umfangreiche Datengrundlage vor.

Die empirisch festgestellten tatsächlichen Effektschätzungen erreichen nur teilweise die gemäß der Entwicklung der Methodik erwarteten Effektstärken:

- Für die höchsten Ausmaßkategorien („erheblich“ für Zielgrößenkategorien Mortalität und schwerwiegend / schwer / HQoL, „beträchtlich“ für Zielgrößenkategorie nicht schwerwiegend / nicht schwer) sind mehr als die Hälfte der tatsächlichen Effektschätzungen kleiner als die erwarteten Effektstärken.
- Für die übrigen Ausmaßkategorien („beträchtlich“ und „gering“ für die Zielgrößenkategorien Mortalität und schwerwiegend / schwer / HQoL, „gering“ für die Zielgrößenkategorie nicht schwerwiegend / nicht schwer) liegen die Effektschätzungen fast vollständig bzw. größtenteils unterhalb der erwarteten Effektstärken.
- Mit Ausnahme der Populationsgröße konnten keine weiteren Faktoren identifiziert werden, für die sich zumindest teilweise unterschiedliche Verteilungen der tatsächlichen Effektschätzungen zeigen.
- Eine mögliche Anpassung der Methodik zur Ausmaßbestimmung sollte sich auf die höchsten Ausmaßkategorien fokussieren.

8 Literatur

1. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ticagrelor – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung [online]. 2011 [Zugriff: 11.07.2023]. URL: https://www.iqwig.de/download/a11-02_ticagrelor_nutzenbewertung_35a_sgb_v_.pdf.
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 7.0 [online]. 2023 [Zugriff: 06.10.2023]. URL: https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.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>.

Anhang A Verteilung der tatsächlichen Effektschätzungen – Sensitivitätsanalyse

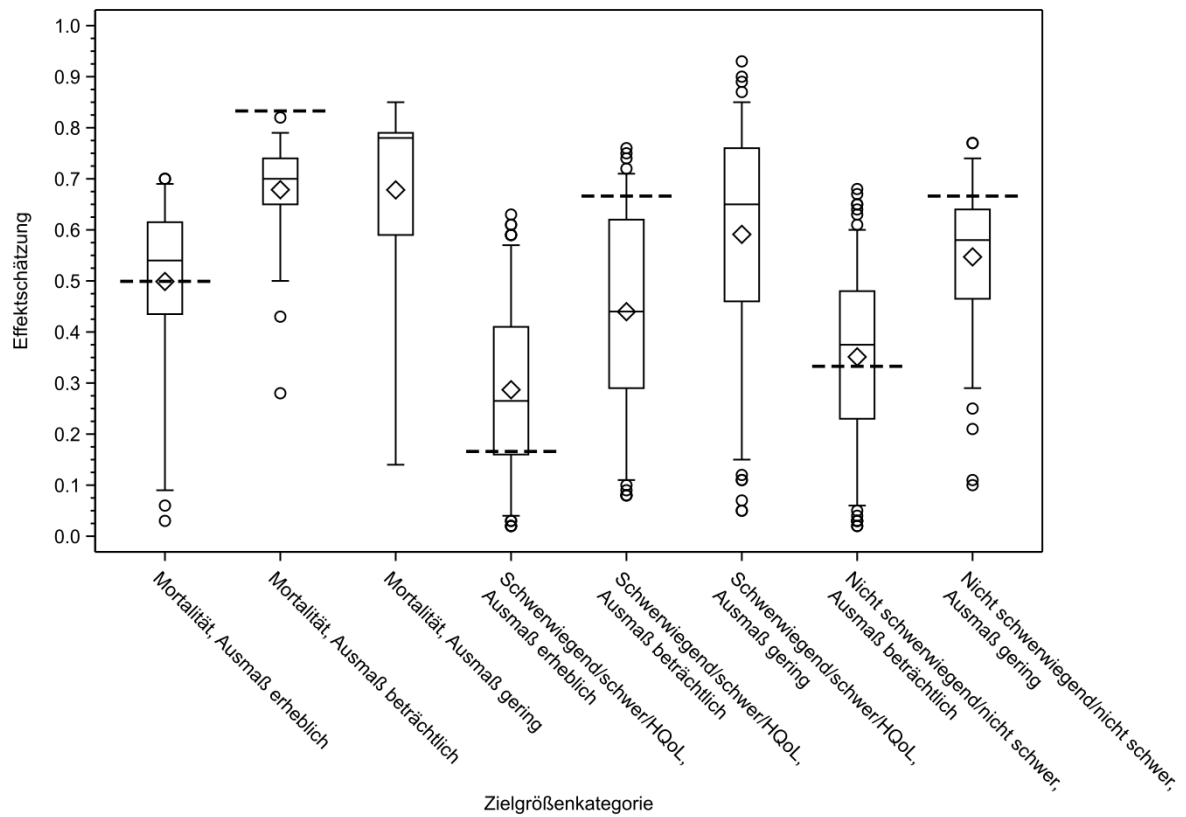


Abbildung 13: Sensitivitätsanalyse zur Verteilung der tatsächlichen Effektschätzungen – jeweils ein zufällig ausgewählter Endpunkt pro Dossierbewertung

Anhang B Verteilung der tatsächlichen Effektschätzungen nach Aussagesicherheit

B.1 Zielgrößenkategorie Mortalität, Ausmaßkategorie „erheblich“

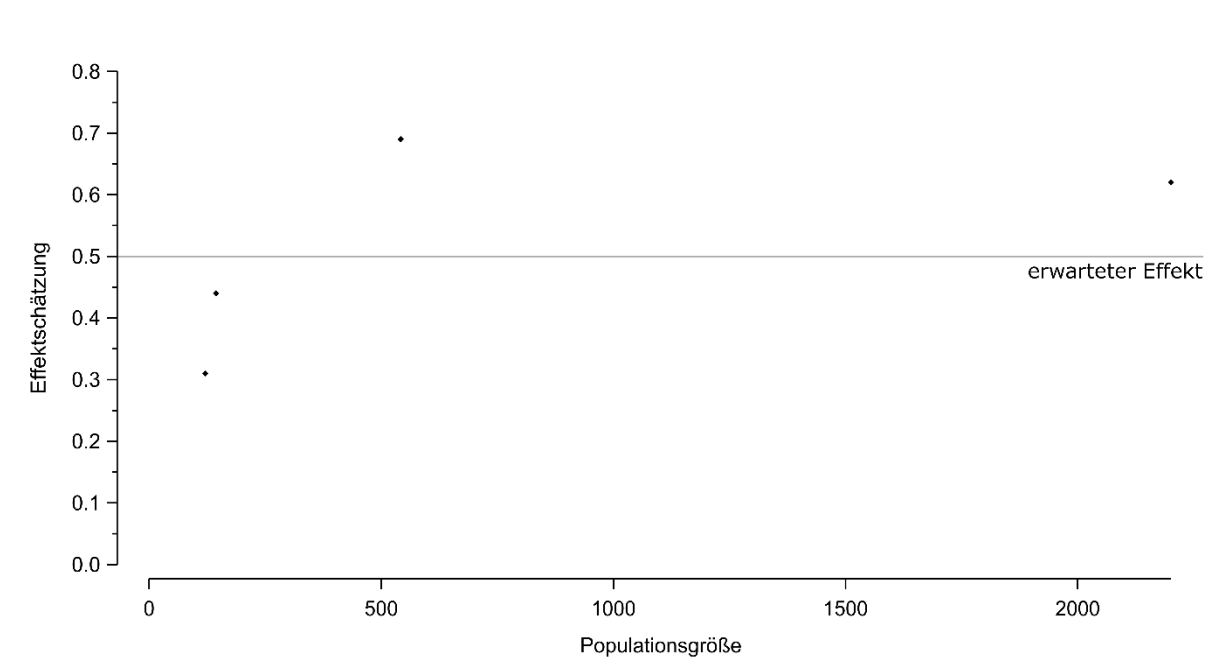


Abbildung 14: Effektschätzungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße

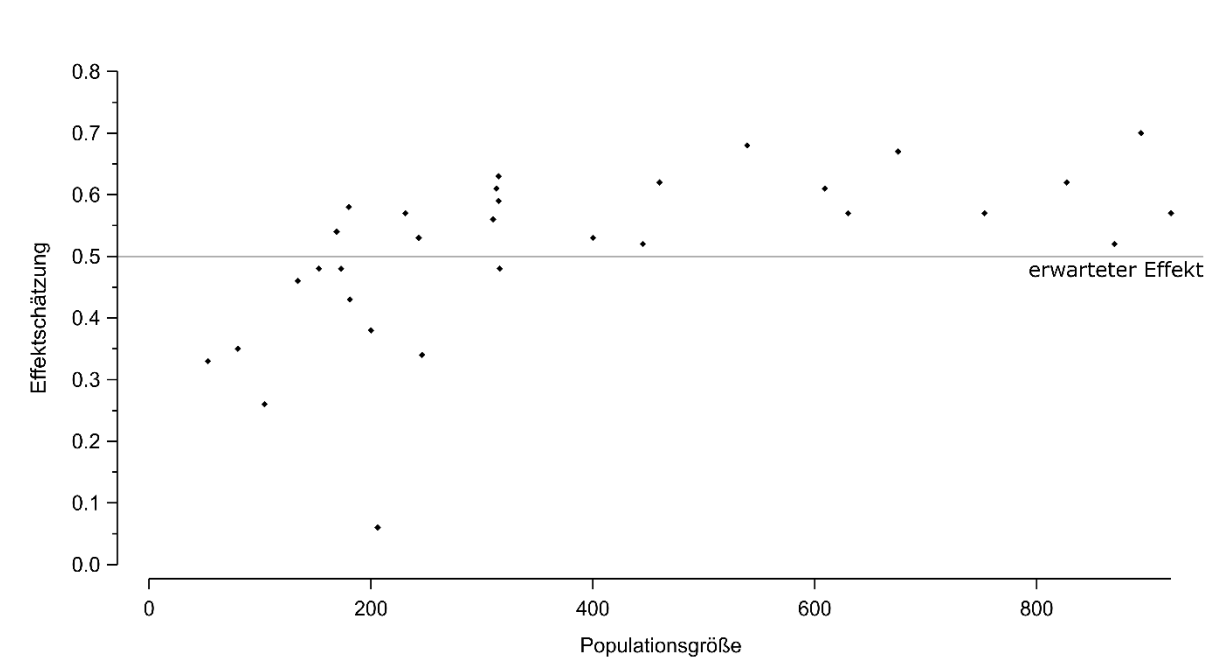


Abbildung 15: Effektschätzungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße

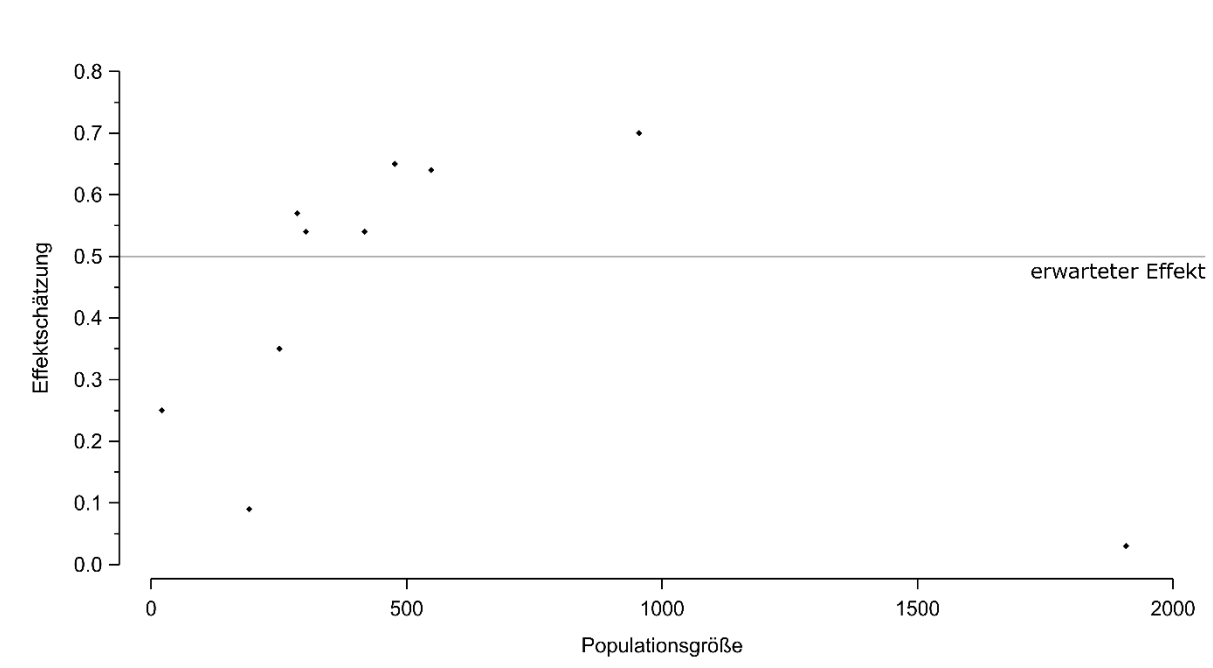


Abbildung 16: Effektschätzungen der Endpunkte der Zielgrößenkategorie Mortalität mit Ausmaß „erheblich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße

B.2 Zielgrößenkategorie schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität, Ausmaßkategorie „erheblich“

B.2.1 Zielgröße schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen)

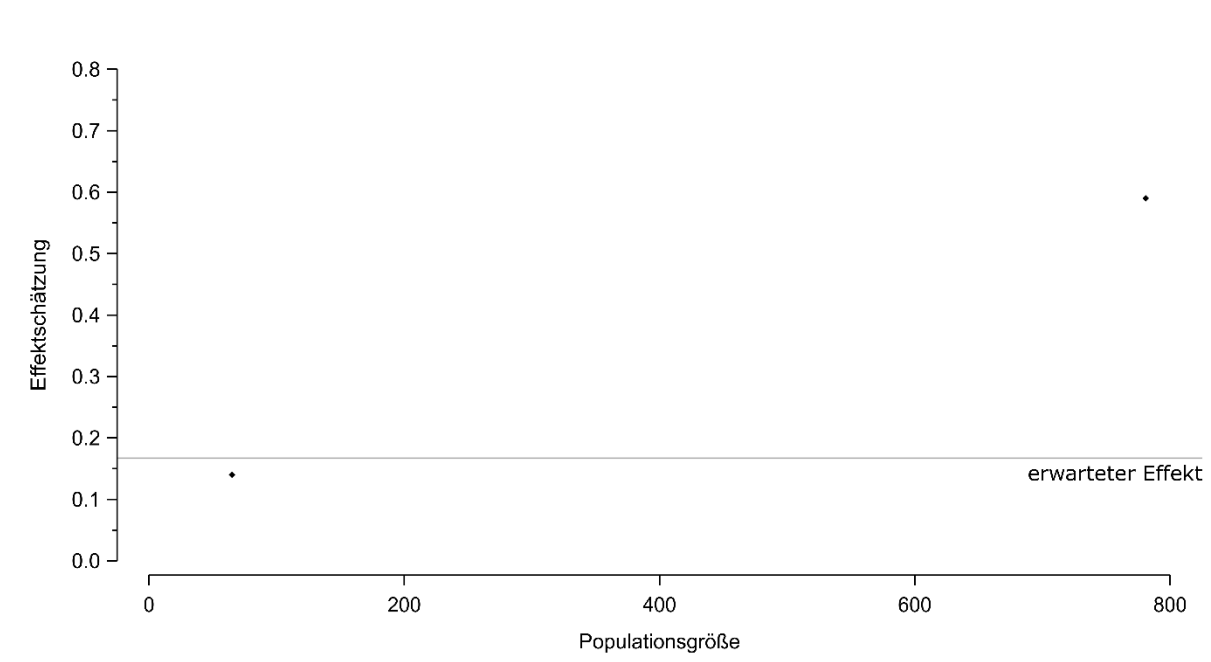


Abbildung 17: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Symptome mit Ausmaß „erheblich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße

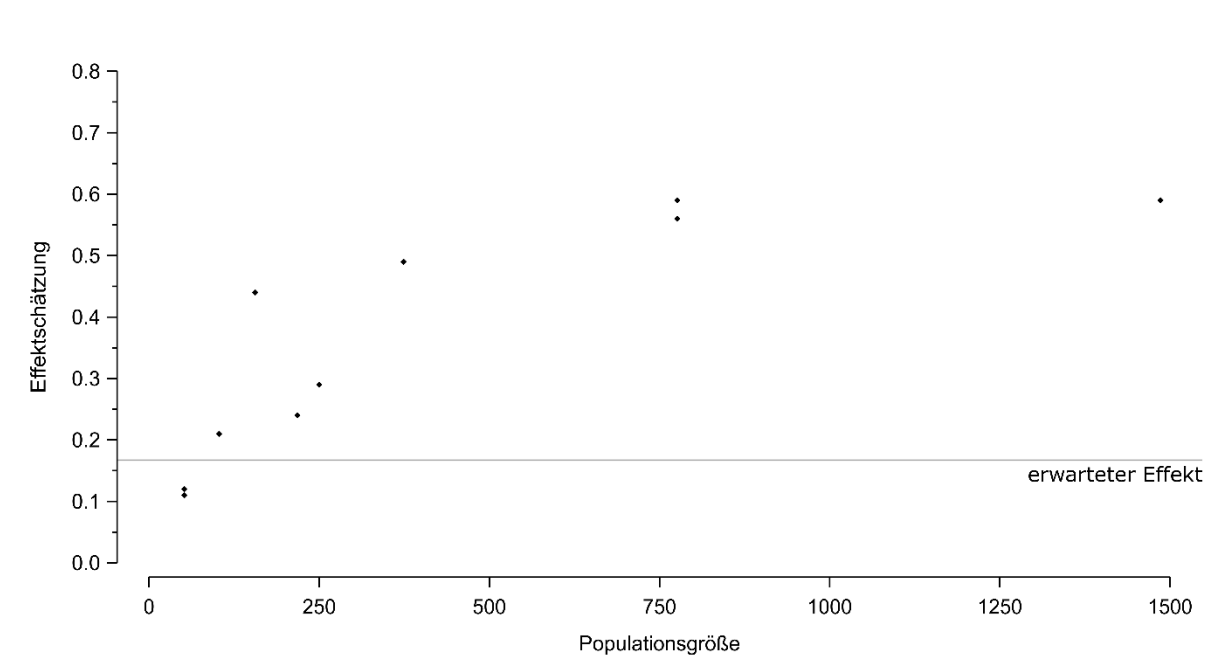


Abbildung 18: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Symptome mit Ausmaß „erheblich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße

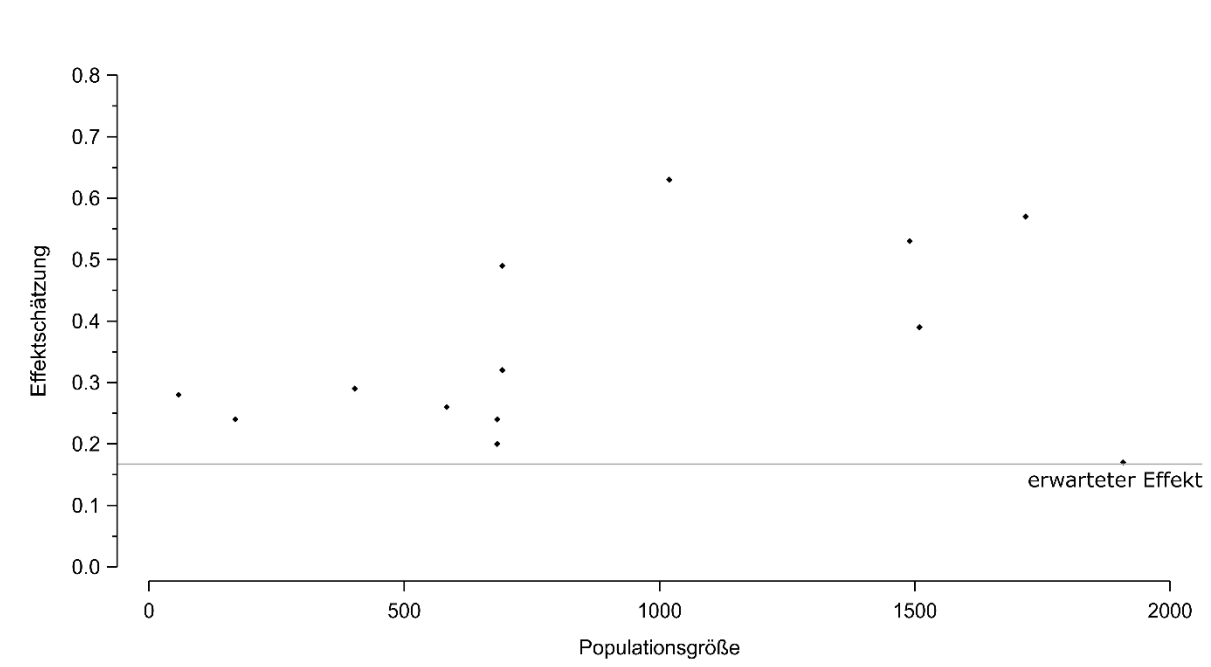


Abbildung 19: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Symptome mit Ausmaß „erheblich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße

B.2.2 Zielgröße gesundheitsbezogene Lebensqualität

Für die Zielgröße gesundheitsbezogene Lebensqualität lagen keine Endpunkte mit der Aussagesicherheit Beleg vor.

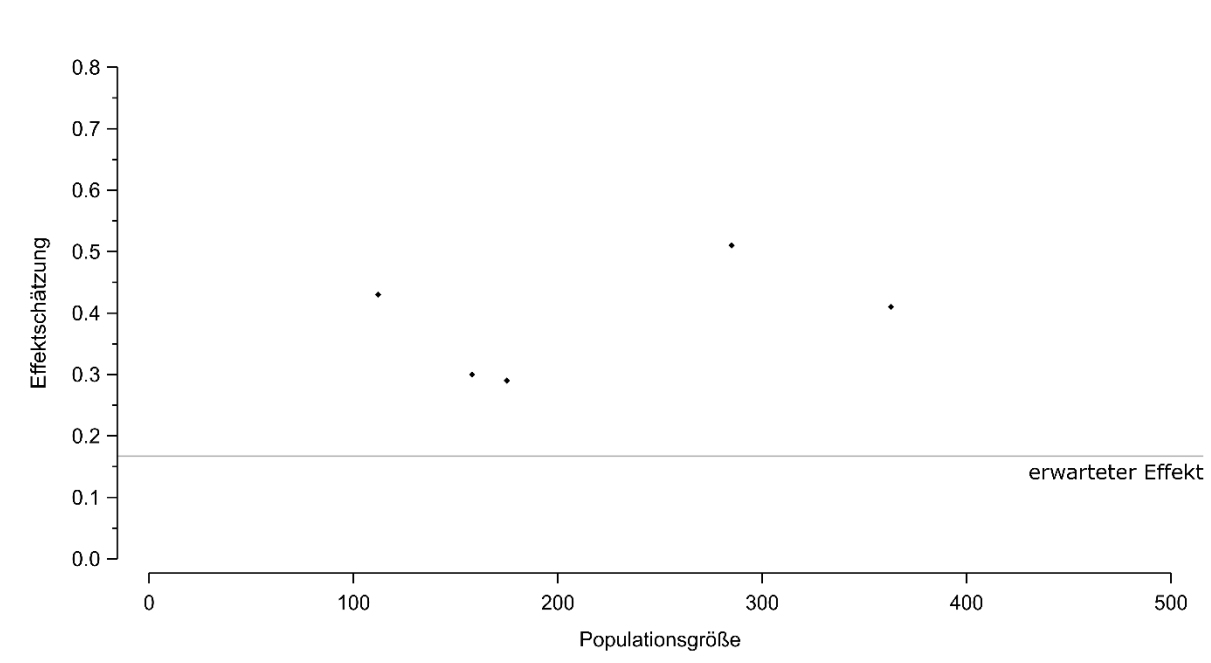


Abbildung 20: Effektschätzungen der Endpunkte der Zielgröße gesundheitsbezogene Lebensqualität mit Ausmaß „erheblich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße

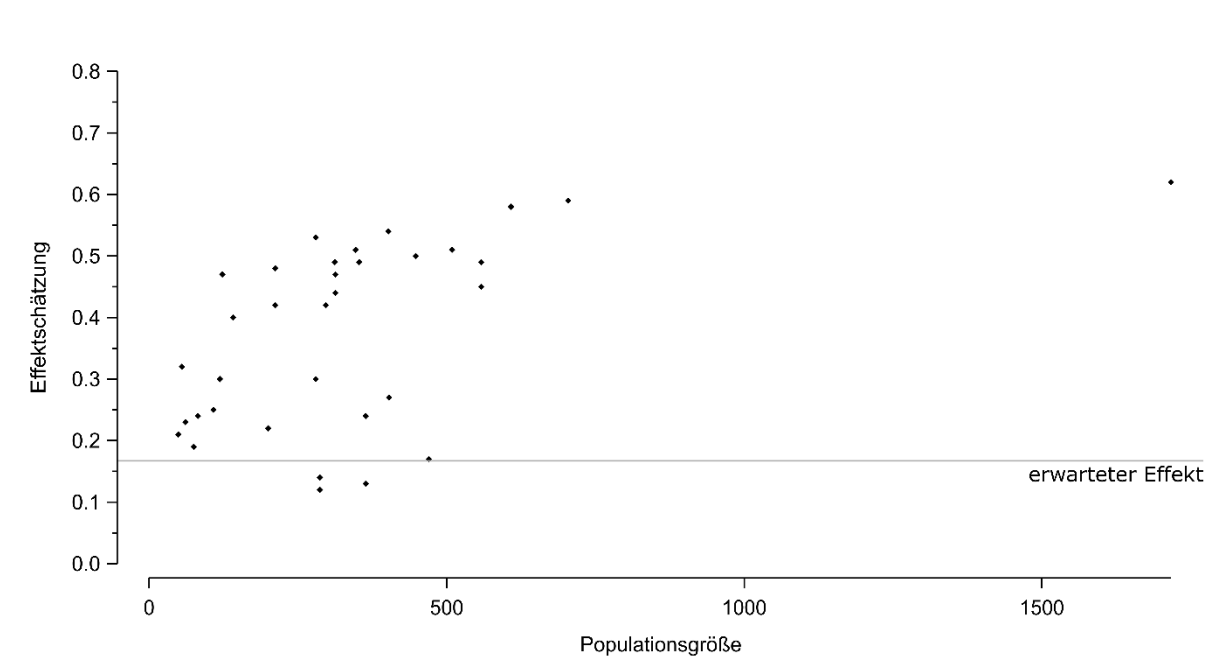


Abbildung 21: Effektschätzungen der Endpunkte der Zielgröße gesundheitsbezogene Lebensqualität mit Ausmaß „erheblich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße

B.2.3 Zielgröße schwerwiegende (bzw. schwere) Nebenwirkungen

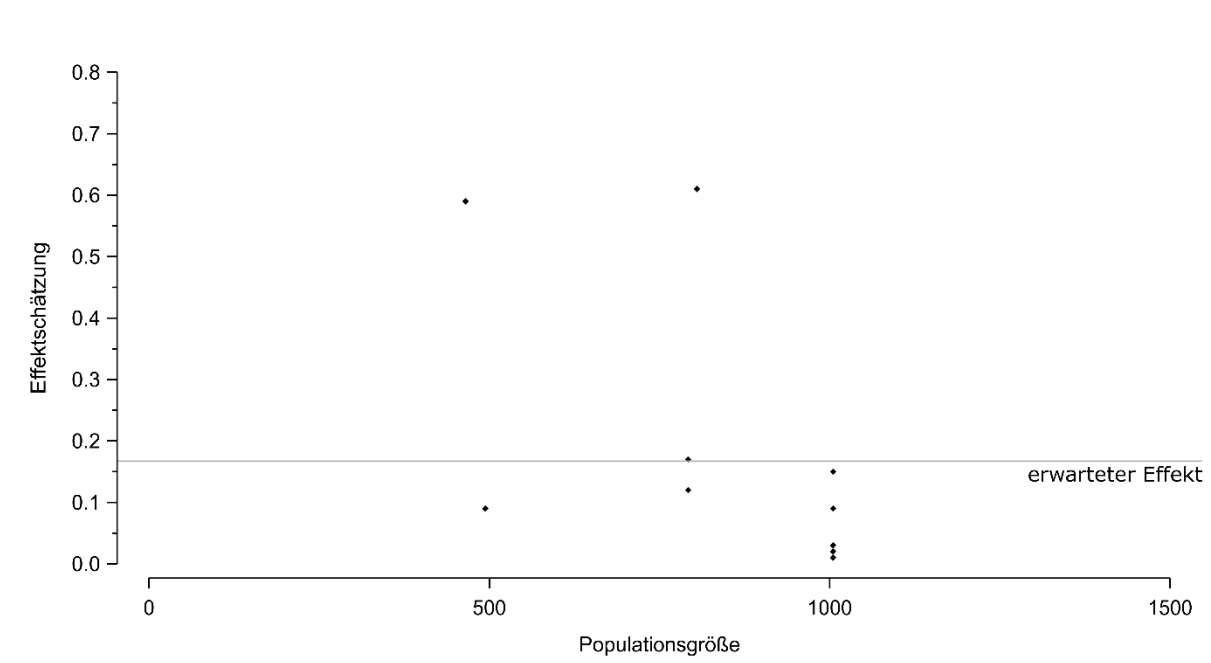


Abbildung 22: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Nebenwirkungen mit Ausmaß „erheblich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße

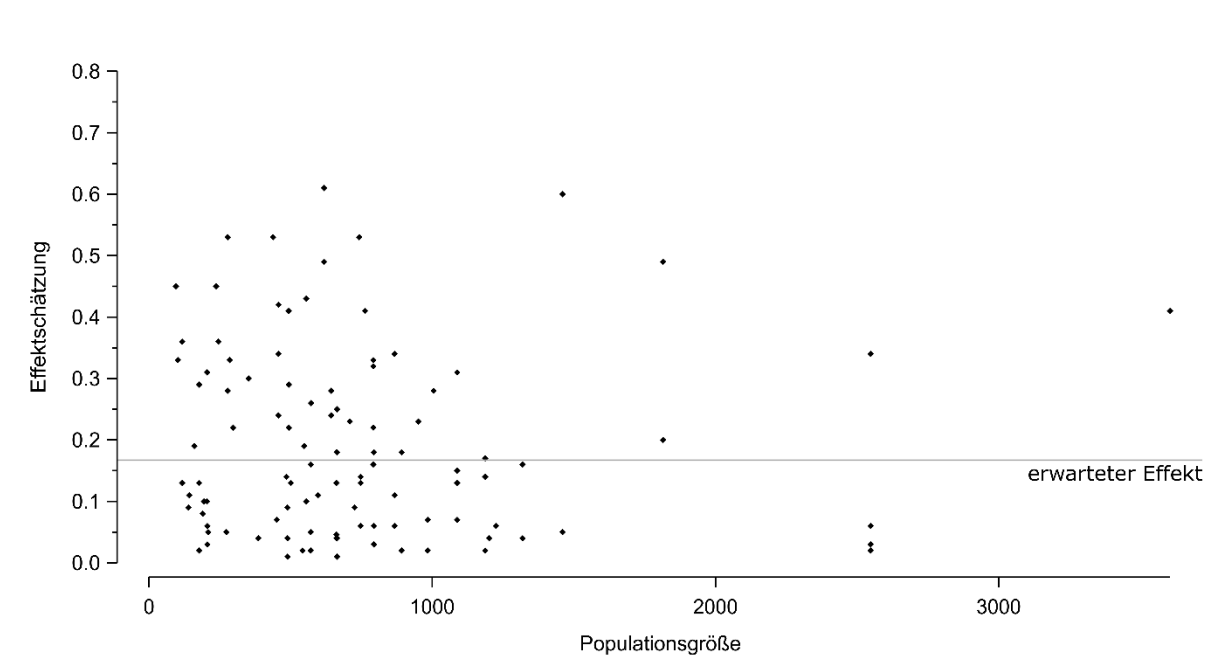


Abbildung 23: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Nebenwirkungen mit Ausmaß „erheblich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße

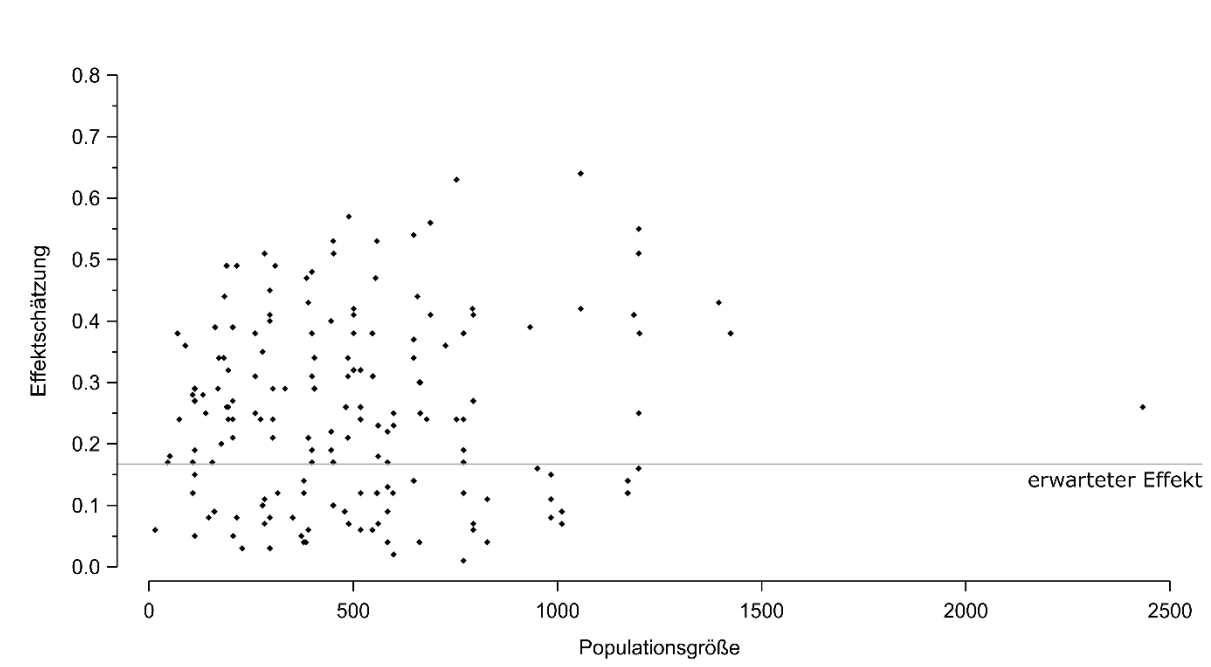


Abbildung 24: Effektschätzungen der Endpunkte der Zielgröße schwerwiegende / schwere Nebenwirkungen mit Ausmaß „erheblich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße

B.3 Zielgrößenkategorie nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen, Ausmaßkategorie „beträchtlich“

B.3.1 Zielgröße nicht schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen)

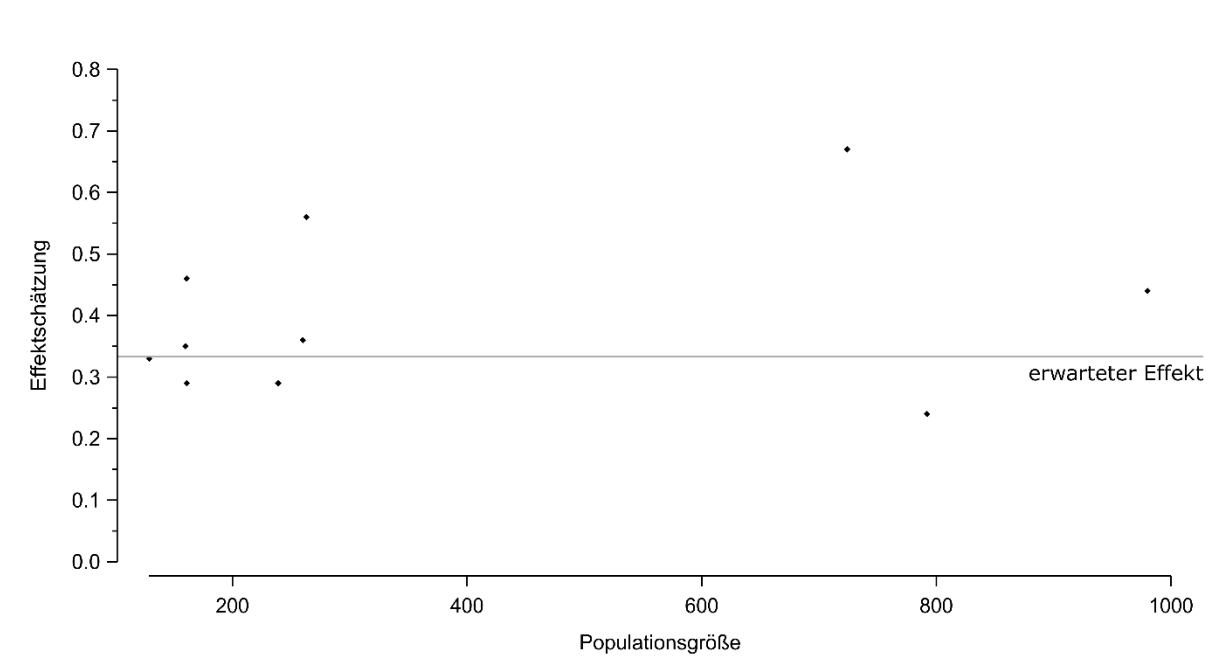


Abbildung 25: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Symptome mit Ausmaß „beträchtlich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße

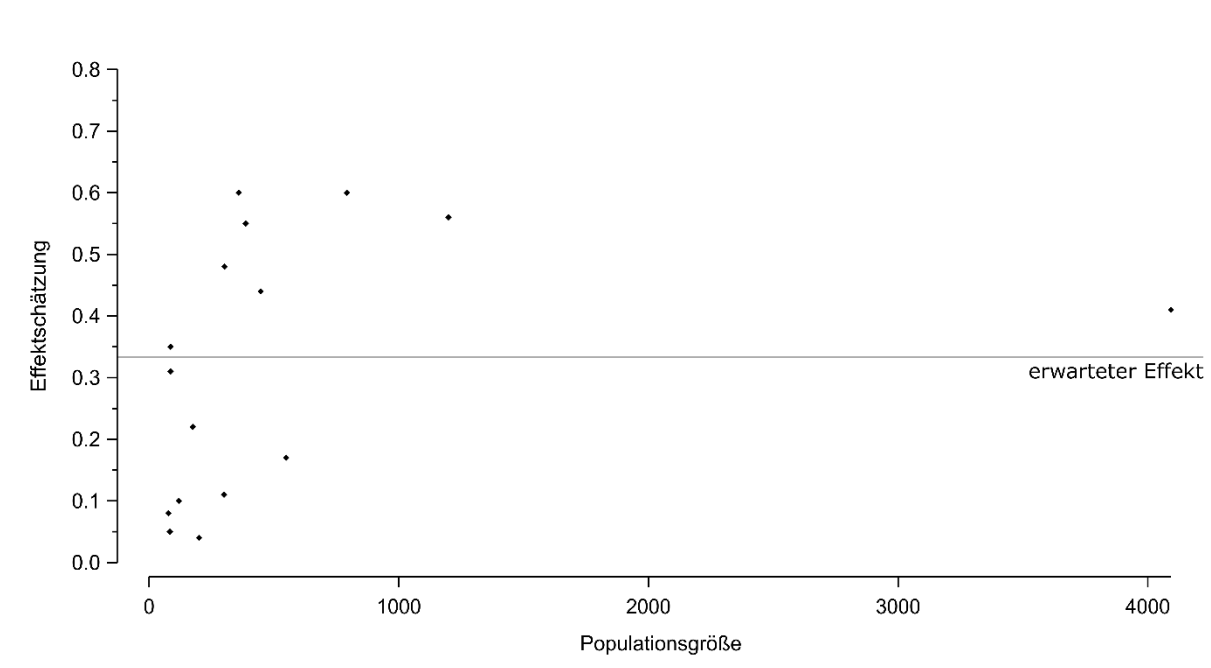


Abbildung 26: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Symptome mit Ausmaß „beträchtlich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße

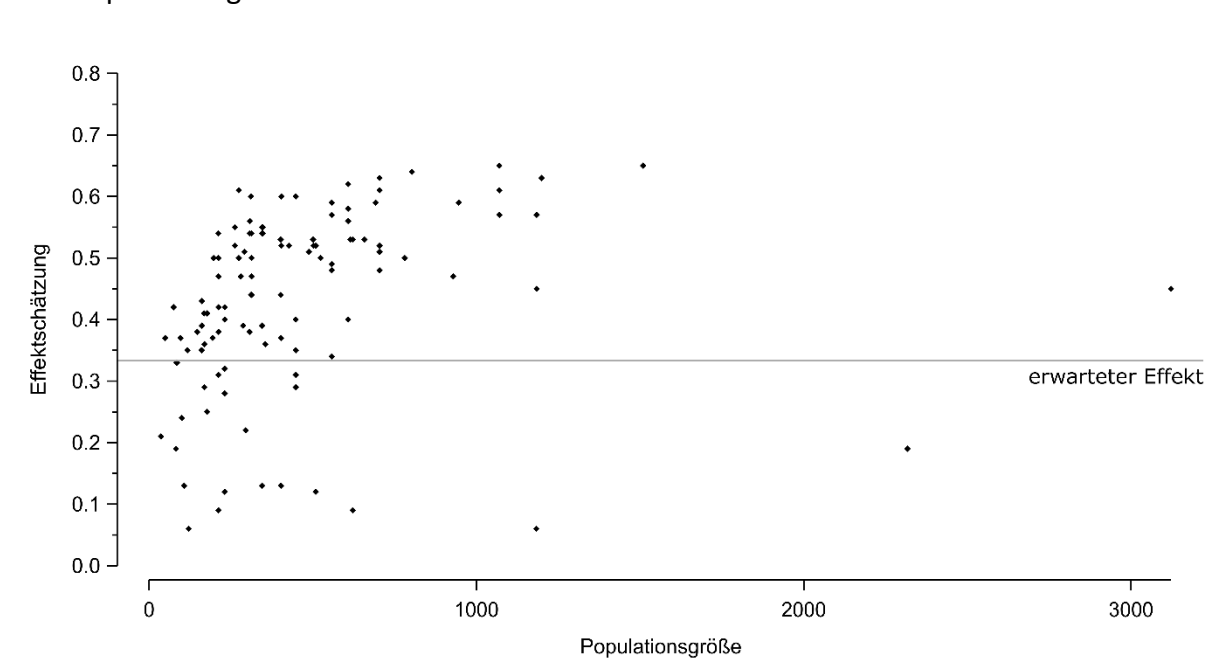


Abbildung 27: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Symptome mit Ausmaß „beträchtlich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße

B.3.2 Zielgröße nicht schwerwiegende (bzw. nicht schwere) Nebenwirkungen

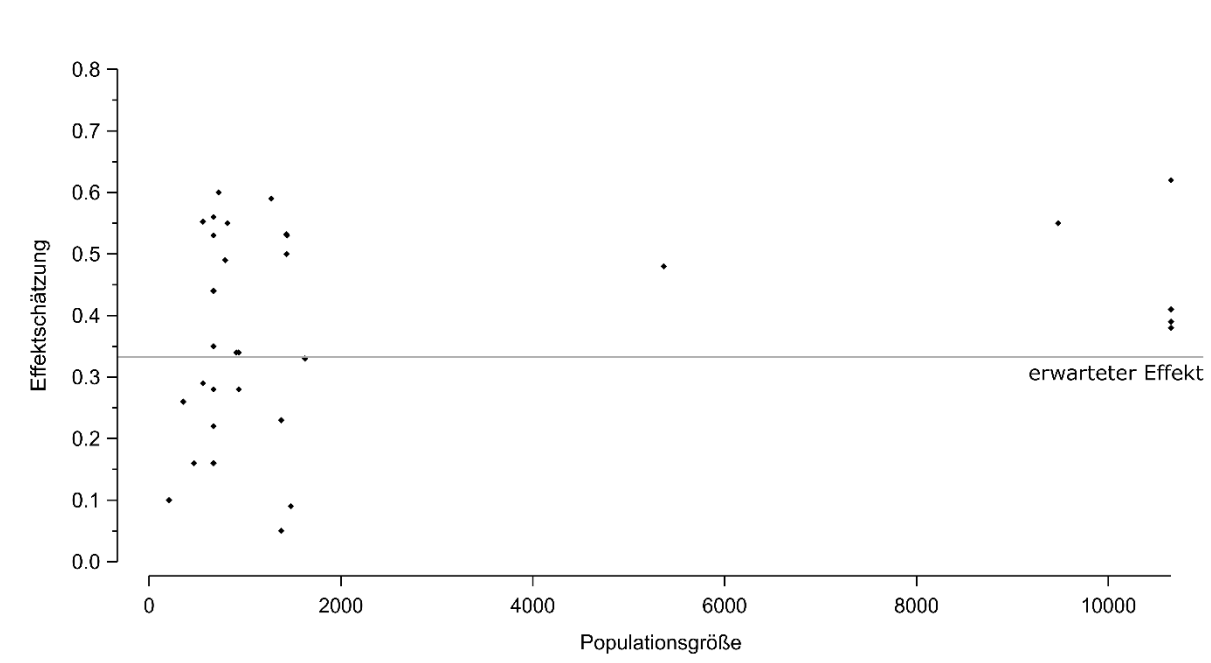


Abbildung 28: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Nebenwirkungen mit Ausmaß „beträchtlich“ und Aussagesicherheit Beleg – Verteilung nach Populationsgröße

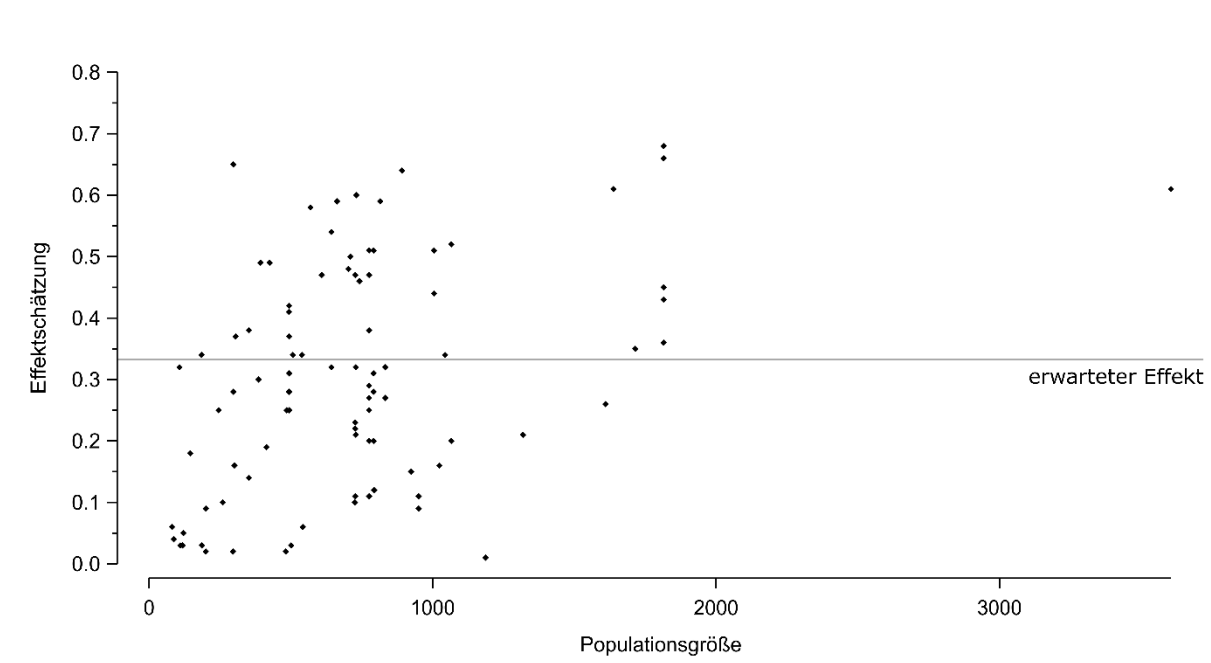


Abbildung 29: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Nebenwirkungen mit Ausmaß „beträchtlich“ und Aussagesicherheit Hinweis – Verteilung nach Populationsgröße

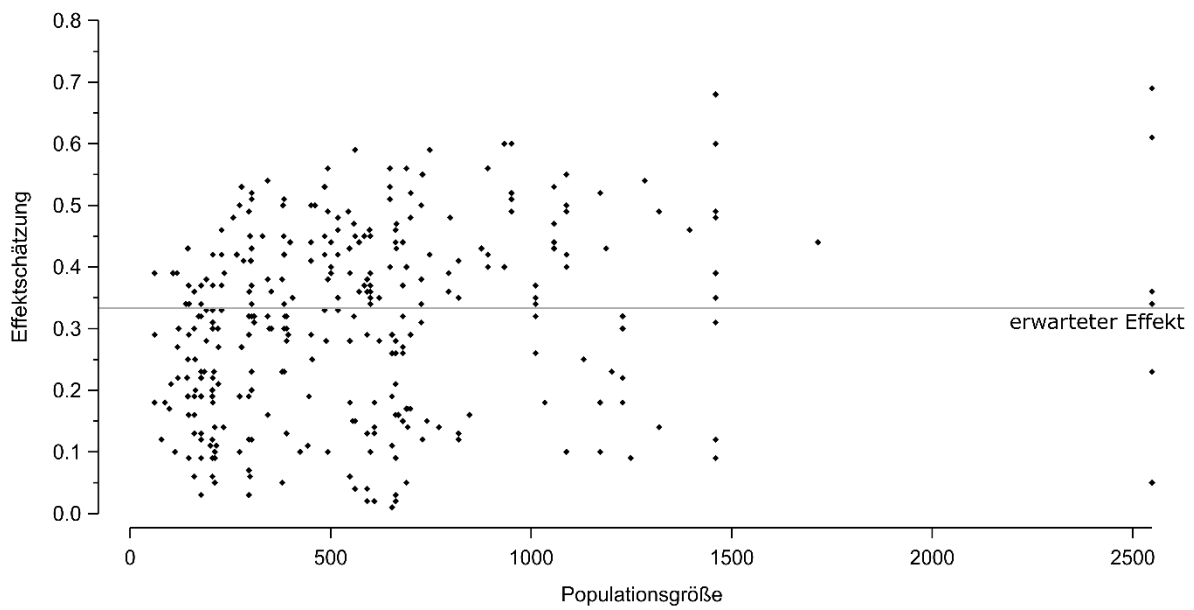
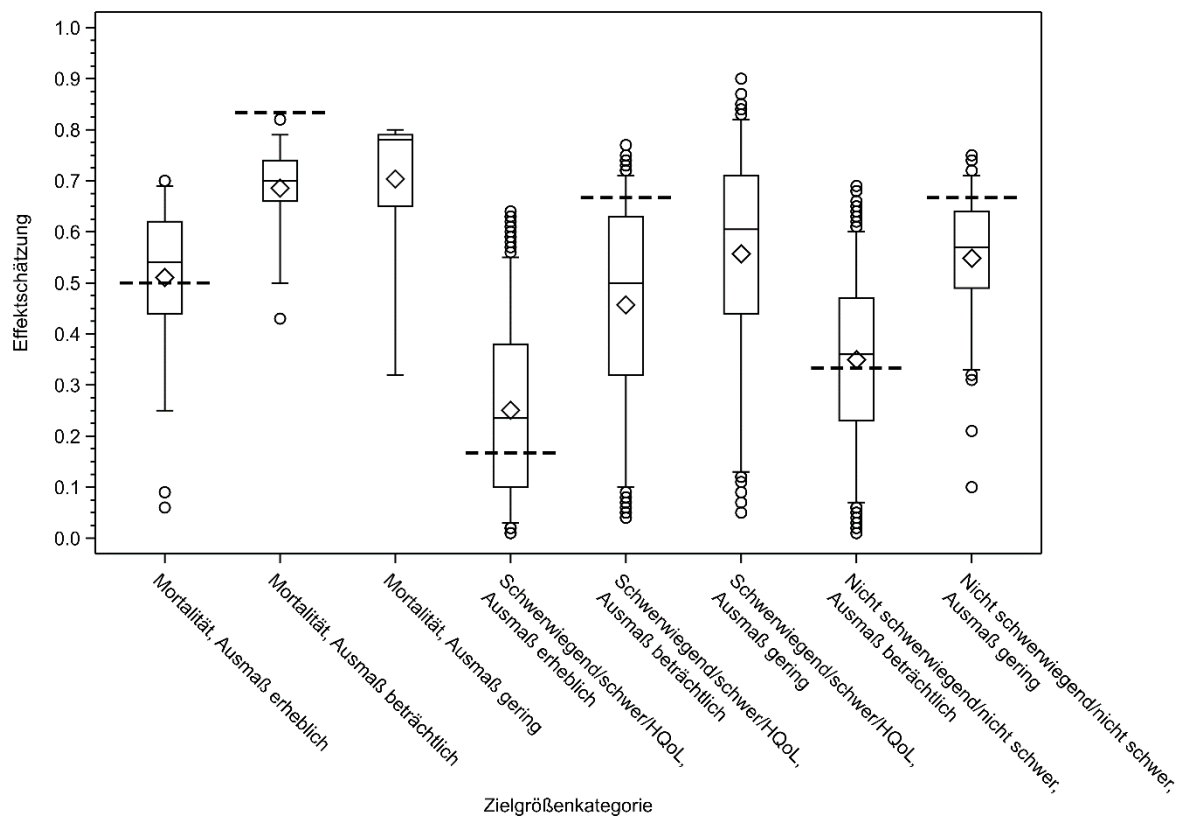


Abbildung 30: Effektschätzungen der Endpunkte der Zielgröße nicht schwerwiegende / nicht schwere Nebenwirkungen mit Ausmaß „beträchtlich“ und Aussagesicherheit Anhaltspunkt – Verteilung nach Populationsgröße

Anhang C Verteilung der tatsächlichen Effektschätzungen nach Themengebiet

Tabelle 5: Verteilung der vorliegenden tatsächlichen Effektschätzungen (Themengebiet Onkologie)

Zielgrößenkategorie	Ausmaßkategorie		
	Endpunkt mit Ausmaß „erheblich“	Endpunkt mit Ausmaß „beträchtlich“	Endpunkt mit Ausmaß „gering“
Mortalität	n = 43	n = 37	n = 14
erwartete Effektstärke (RR) ^a	0,50	0,83	– ^b
tatsächliche Effektschätzung ^c			
Median [Q1; Q3]	0,54 (0,44; 0,62)	0,70 (0,66; 0,74)	0,78 (0,65; 0,79)
Mittelwert	0,51	0,69	0,70
Schwerwiegend / schwer / HQoL	n = 310	n = 225	n = 218
erwartete Effektstärke (RR) ^a	0,17	0,67	– ^b
tatsächliche Effektschätzung ^c			
Median [Q1; Q3]	0,24 (0,10; 0,38)	0,50 (0,32; 0,63)	0,61 (0,44; 0,71)
Mittelwert	0,25	0,46	0,56
nicht schwerwiegend / nicht schwer		n = 411	n = 95
erwartete Effektstärke (RR) ^a	– ^d	0,33	0,67
tatsächliche Effektschätzung ^c			
Median [Q1; Q3]	k. A.	0,36 (0,23; 0,47)	0,57 (0,49; 0,64)
Mittelwert	k. A.	0,35	0,55
<p>a. ursprünglich entwickelt auf Basis des relativen Risikos, jedoch bezogen auf alle relativen Effektmaße (HR, RR, Peto-OR und Rate Ratio)</p> <p>b. Gemäß den Allgemeinen Methoden des Instituts [2] wird jeglicher zum üblichen Irrtumsniveau 5 % statistisch signifikante Effekt zumindest als geringer Zusatznutzen eingestuft.</p> <p>c. umfasst die Effektschätzungen HR, RR, Peto-OR und Rate Ratio</p> <p>d. Gemäß den Allgemeinen Methoden des Instituts [2] wird für die Kategorie nicht schwerwiegend / nicht schwer kein Ausmaß „erheblich“ bestimmt.</p> <p>HQoL: gesundheitsbezogene Lebensqualität; HR: Hazard Ratio; k. A.: keine Angaben; n: Anzahl Endpunkte mit Ausmaßbestimmung; OR: Odds Ratio; Q1: 1. Quartil; Q3: 3. Quartil; RR: relatives Risiko</p>			



Erläuterung Boxplot:

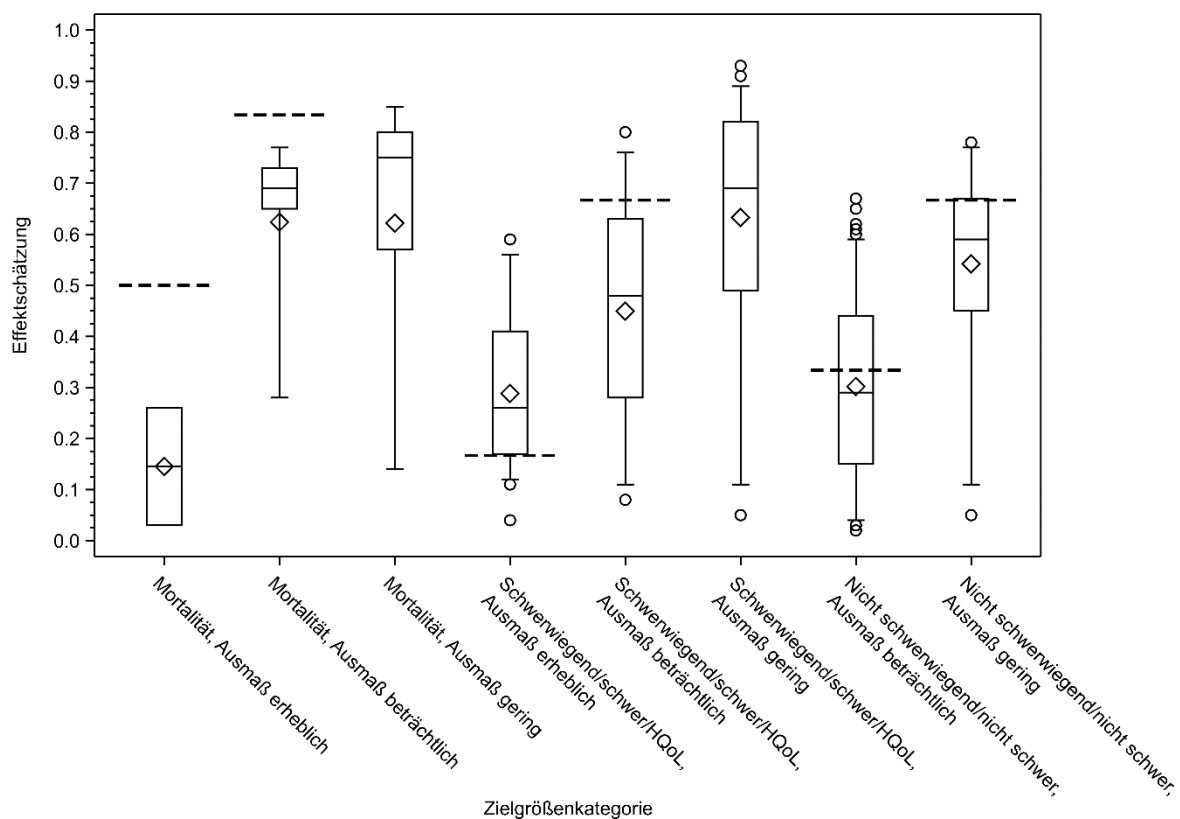
Die Box entspricht dem Bereich des 25 %- bis 75 %-Quantils, die durchgezogene horizontale Linie innerhalb der Box dem Median, die Raute innerhalb der Box dem Mittelwert. Die Linien außerhalb der Box entsprechen dem 5 %- und 95 %-Quantil. Werte außerhalb dieser Quantile sind als Kreise dargestellt.

Die gestrichelten horizontalen Linien entsprechen den jeweils erwarteten Effektstärken.

Abbildung 31: Verteilung der tatsächlichen Effektschätzungen – Auswertung aller Endpunkte für das Themengebiet Onkologie

Tabelle 6: Verteilung der vorliegenden tatsächlichen Effektschätzungen (Themengebiet nicht onkologische Bewertungen^a)

Zielgrößenkategorie	Ausmaßkategorie		
	Endpunkt mit Ausmaß „erheblich“	Endpunkt mit Ausmaß „beträchtlich“	Endpunkt mit Ausmaß „gering“
Mortalität	n = 2	n = 5	n = 5
erwartete Effektstärke (RR) ^b	0,50	0,83	– ^c
tatsächliche Effektschätzung ^d			
Median [Q1; Q3]	0,15 (0,03; 0,26)	0,69 (0,65; 0,73)	0,75 (0,57; 0,80)
Mittelwert	0,15	0,62	0,62
Schwerwiegend / schwer / HQoL	n = 40	n = 57	n = 61
erwartete Effektstärke (RR) ^b	0,17	0,67	– ^c
tatsächliche Effektschätzung ^d			
Median [Q1; Q3]	0,26 (0,17; 0,41)	0,48 (0,28; 0,63)	0,69 (0,49; 0,82)
Mittelwert	0,29	0,45	0,63
nicht schwerwiegend / nicht schwer		n = 170	n = 54
erwartete Effektstärke (RR) ^b	– ^e	0,33	0,67
tatsächliche Effektschätzung ^d			
Median [Q1; Q3]	k. A.	0,29 (0,15; 0,44)	0,59 (0,45; 0,67)
Mittelwert	k. A.	0,30	0,54
<p>a. beinhaltet die Ergebnisse aller Ausmaßbestimmungen aus den Themengebieten Diabetes mellitus, Herz-Kreislauf / Lunge, Infektiologie, Psychiatrie / Neurologie und Sonstige.</p> <p>b. ursprünglich entwickelt auf Basis des relativen Risikos, jedoch bezogen auf alle relativen Effektmaße (HR, RR, Peto-OR und Rate Ratio)</p> <p>c. Gemäß den Allgemeinen Methoden des Instituts [2] wird jeglicher zum üblichen Irrtumsniveau 5 % statistisch signifikante Effekt zumindest als geringer Zusatznutzen eingestuft.</p> <p>d. umfasst die Effektschätzungen HR, RR, Peto-OR und Rate Ratio</p> <p>e. Gemäß den Allgemeinen Methoden des Instituts [2] wird für die Kategorie nicht schwerwiegend / nicht schwer kein Ausmaß „erheblich“ bestimmt.</p> <p>HQoL: gesundheitsbezogene Lebensqualität; HR: Hazard Ratio; k. A.: keine Angaben; n: Anzahl Endpunkte mit Ausmaßbestimmung; OR: Odds Ratio; Q1: 1. Quartil; Q3: 3. Quartil; RR: relatives Risiko</p>			



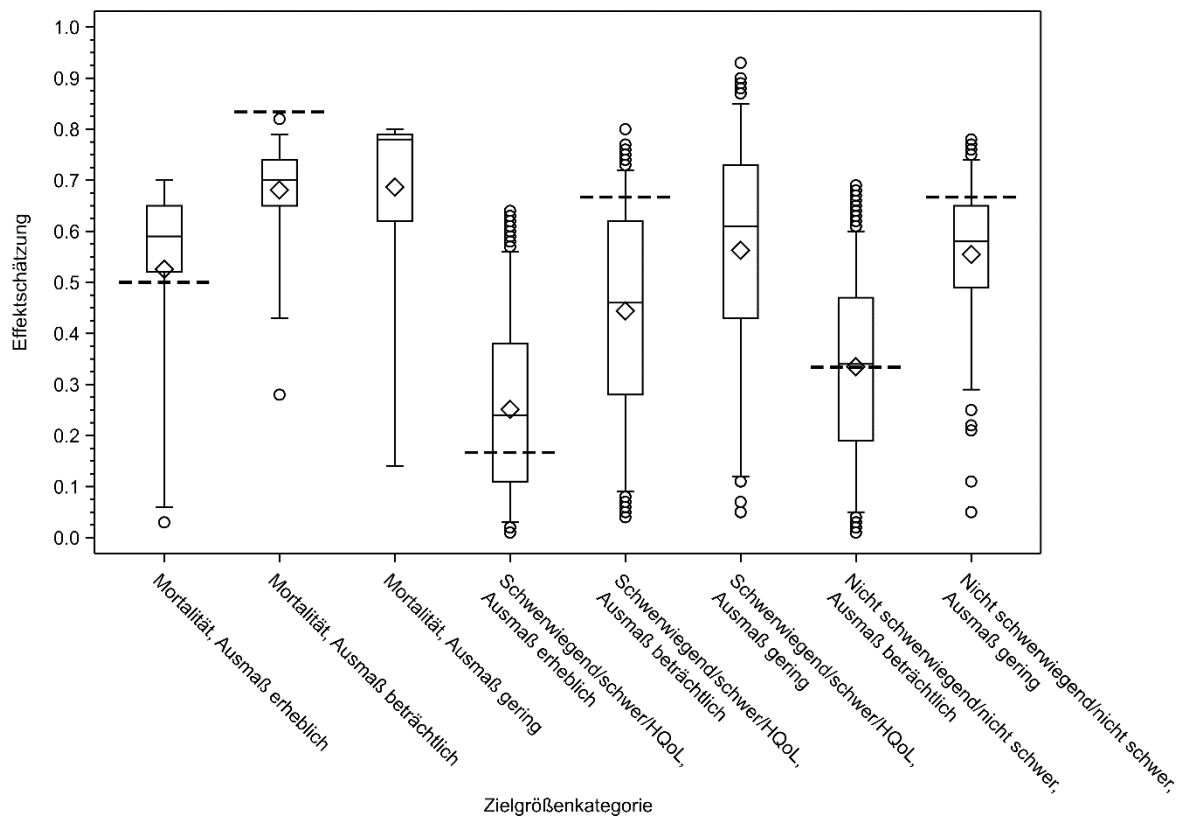
Erläuterung Boxplot:

Die Box entspricht dem Bereich des 25 %- bis 75 %-Quantils, die durchgezogene horizontale Linie innerhalb der Box dem Median, die Raute innerhalb der Box dem Mittelwert. Die Linien außerhalb der Box entsprechen dem 5 %- und 95 %-Quantil. Werte außerhalb dieser Quantile sind als Kreise dargestellt.

Die gestrichelten horizontalen Linien entsprechen den jeweils erwarteten Effektstärken.

Abbildung 32: Verteilung der tatsächlichen Effektschätzungen – Auswertung aller Endpunkte für das Themengebiet nicht onkologische Bewertungen

Anhang D Verteilung der tatsächlichen Effektschätzungen – Gesamtpopulation und Subgruppenpopulation

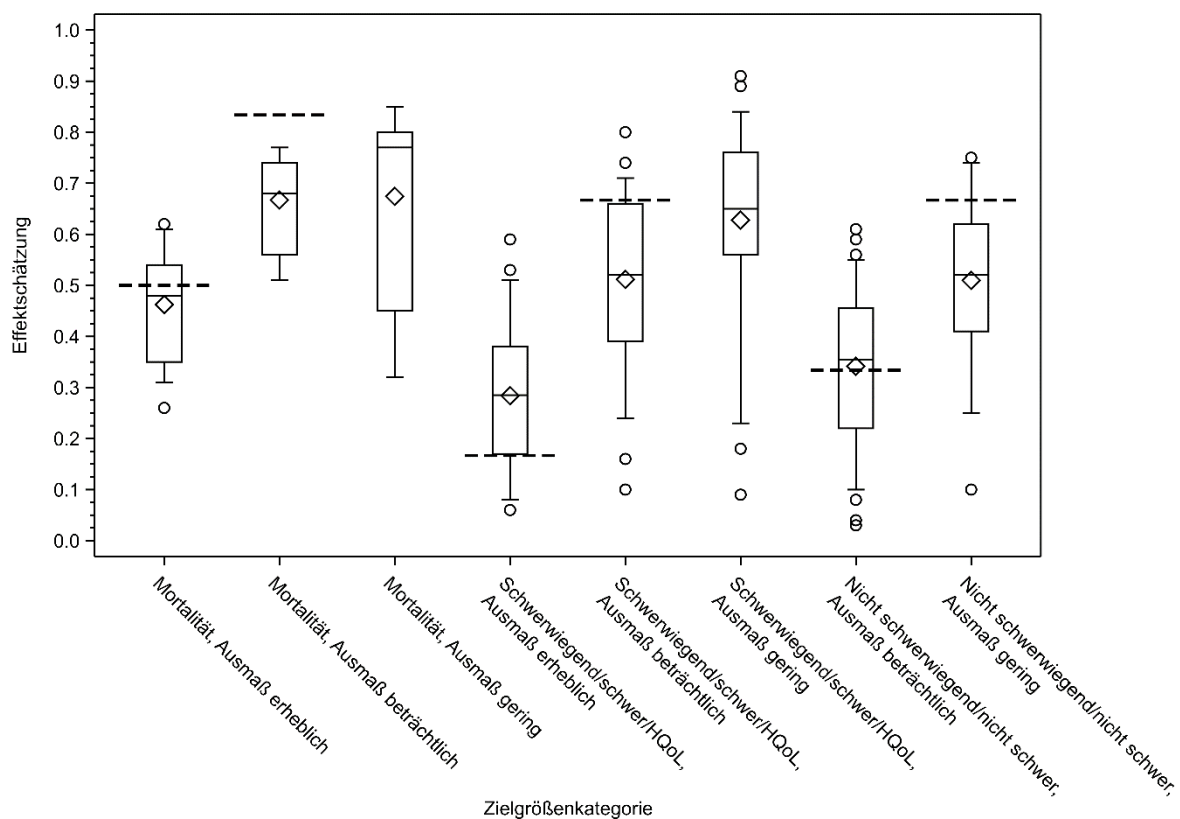


Erläuterung Boxplot:

Die Box entspricht dem Bereich des 25 %- bis 75 %-Quantils, die durchgezogene horizontale Linie innerhalb der Box dem Median, die Raute innerhalb der Box dem Mittelwert. Die Linien außerhalb der Box entsprechen dem 5 %- und 95 %-Quantil. Werte außerhalb dieser Quantile sind als Kreise dargestellt.

Die gestrichelten horizontalen Linien entsprechen den jeweils erwarteten Effektstärken.

Abbildung 33: Verteilung der tatsächlichen Effektschätzungen – Gesamtpopulation



Erläuterung Boxplot:

Die Box entspricht dem Bereich des 25 %- bis 75 %-Quantils, die durchgezogene horizontale Linie innerhalb der Box dem Median, die Raute innerhalb der Box dem Mittelwert. Die Linien außerhalb der Box entsprechen dem 5 %- und 95 %-Quantil. Werte außerhalb dieser Quantile sind als Kreise dargestellt.

Die gestrichelten horizontalen Linien entsprechen den jeweils erwarteten Effektstärken.

Abbildung 34: Verteilung der tatsächlichen Effektschätzungen – Subgruppenpopulation