

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.