

Beurteilung der klinischen Wirksamkeit in HTA

Einleitung

Die Beurteilung der Auswirkungen einer Technologie erfordert umfassende Informationen, die widerspiegeln, was in einem Gesundheitssystem oder der Gesellschaft mit Wahrscheinlichkeit passiert. Eine gute Analyse erfordert die Nutzung von Expertenrat und Methoden aus den verschiedenen Disziplinen, die als Input verwendet werden.

HTA-Beurteilungen sind vergleichende Analysen, die die bestehende Standard-Grundversorgung mit der neuen Technologie vergleichen, um festzustellen, welchen Wert die neue Technologie liefern würde (den sogenannten „Mehrwert“). Ländern, Regionen und Krankenhäuser setzen HTA-Beurteilungen auf verschiedene Weisen um. Alle berücksichtigen das Gesundheitsproblem in ihrem lokalen Kontext und beurteilen dann die Behandlung für das Anwendungsgebiet, auf das sich die Regulierungsbehörden für das Arzneimittel geeinigt haben.

Innerhalb des Anwendungsgebiets untersuchen HTA-Gremien verfügbare Daten, um zu bewerten, wie gut die Behandlung im Vergleich mit der besten Standardversorgung (in Bezug auf die Sicherheit und klinische Wirksamkeit) funktionieren kann. Einige HTA-Gremien beurteilen auch die Kosten und die Wirtschaftlichkeit eines Arzneimittels. Und während einige Gremien formale Beurteilungen der ethischen, organisatorischen, sozialen und rechtlichen Aspekte durchführen, berücksichtigen andere diese Probleme in der Beurteilung einfach implizit.

Der „Mehrwert“ einer Technologie wird von jeder einzelnen HTA-Organisation in einer unterschiedlichen, vielfältigen Weise

bestimmt. Schlussfolgerungen zum „Mehrwert“ einer Technologie können sich je nach HTA-Organisation unterscheiden. Das europäische Netzwerk für HTA (EUnetHTA) hat einen Rahmen entwickelt, mithilfe dessen der „Mehrwert“ beurteilt werden kann – das sogenannte HTA Core Model[®].¹ Es gibt neun Bereiche im HTA Core Model[®]:

1. Gesundheitsproblem
2. Technische Beschreibung der Technologie
3. Sicherheit
4. Klinische Wirksamkeit
5. Kosten und finanzielle Bewertung (Wirtschaftlichkeit)
6. Ethische Analyse
7. Organisatorische Aspekte
8. Soziale Aspekte
9. Rechtliche Aspekte

EUnetHTA hat die Beurteilung der ersten vier Bereiche als eine „Beurteilung der relativen Wirksamkeit“ definiert, bei der zum Beispiel eine neue Behandlung mit der bestehenden Behandlung/den bestehenden Behandlungen verglichen wird.

Beurteilung der klinischen Wirksamkeit in HTA

Eine Beurteilung der klinischen Wirksamkeit untersucht die Wirkung einer neuen Technologie auf die Gesundheit von Patienten unter klinischen Standardbedingungen im Vergleich zur derzeitigen Standardversorgung. Die Wirkung, die eine Technologie auf die Gesundheit hat, wird in der Regel mithilfe einer weiteren Untersuchung der gesundheitlichen Folgen analysiert. Patienten wünschen Zugang zu neuen Arzneimitteln, die:

- Ergebnisse, die als „schlecht“ empfunden werden, reduzieren – wie etwa Herzinfarkte, Krankenhausaufenthalte und Nebenwirkungen **und/oder**

- Ergebnisse, die als „gut“ empfunden werden, erhöhen – wie etwa eine verbesserte Funktionalität und schmerzfreie Tage.

Während der Beurteilung der klinischen Wirksamkeit nutzen HTA-Gremien bewährte Methoden im Zusammenhang stehender Disziplinen der Medizin. Insbesondere die Bewertung der klinischen Wirksamkeit wird mithilfe von Prinzipien durchgeführt, die aus Epidemiologie und Medizin (die sogenannten „klinische Epidemiologie“) geliehen werden.

Es gibt vier zugrundeliegende Prinzipien für eine gute Beurteilung der klinischen Wirksamkeit:

1. Informationen suchen,
2. Relevante Fragen stellen,
3. Unterschiede verstehen **und**
4. Unterschiede bewerten.

Informationen suchen

HTA-Gremien verwenden klinische Informationen, um abzuschätzen, welche gesundheitlichen Folgen den Patienten widerfahren könnten, wenn sie ein neues Arzneimittel erhalten. Zuerst jedoch müssen sie entscheiden, wie sie die Informationen sammeln. Es gibt drei verschiedene Arten, auf die HTA-Gremien klinische Informationen zu neuen Technologien erhalten können:

1. Überprüfen von bereits bestehenden Informationen zur Wirksamkeit des Arzneimittels,
2. Durchführen einer neuen Studie zur Erfassung von Informationen und Bewertung der Wirksamkeit des Arzneimittels unter realen Bedingungen **oder**
3. Befragen von Ärzten und Patienten („Experten“) zu ihren Erwartungen an das Arzneimittel.

HTA-Gremien verwenden oft eine Kombination aus diesen Ansätzen. Zum Beispiel:

- Sie können Informationen vom Zulassungsinhaber der Technologie nutzen, um ihre eigenen unabhängigen Überprüfungen und Analysen zu untermauern.
- Wenn Informationen fehlen, kann eine Expertenmeinung erforderlich sein – zum Beispiel um herauszufinden, ob Änderungen bei den kurzfristigen Ergebnissen (wie etwa die Senkung des Cholesterins) Vorhersagen zu Änderungen bei den langfristigen Ergebnissen (wie etwa Vermeiden von Krankenhausaufenthalten) machen können.

HTA-Gremien beauftragen selten neue Studien, da die erforderliche Zeit für Planung und Zulassung einer Studie typischerweise zu lang ist. In einigen Fällen erlauben die verantwortlichen Gremien, ein Arzneimittel bedingt auf Grundlage weiterer Informationen zu erstatten (das wäre ähnlich der Erteilung einer bedingten Zulassung durch die Regulierungsbehörden, welche die Erfassung weiterer Informationen erfordert). Das Risiko, dass ein neues Arzneimittel unter realen Bedingungen schlechter als erwartet wirkt, kann dann vom Zulassungsinhaber und verantwortlichen Gremium durch Preisverhandlungsmechanismen oder andere Änderungen an den Bedingungen im Zusammenhang mit dem erstatteten Zugang (wie etwa weitere Beschränkungen bei der Patientenpopulation, die einen erstatteten Zugang erhalten kann), während die Patienten unmittelbarerem Zugang erhalten.

Relevante Fragen stellen

Bei der Beurteilung der klinischen Wirksamkeit einer neuen Gesundheitstechnologie muss das HTA-Gremium alle damit im Zusammenhang stehenden Ergebnisse sorgfältig berücksichtigen. Es ist wichtig, diese Ergebnisse zu kennen, um die relevanten Fragen zur Wirksamkeit der Technologie zu stellen.

Es besteht ein steigendes Verständnis dafür, dass die Ergebnisse, die Ärzten wichtig erscheinen, nicht immer jene sind, die von den Patienten als am wichtigsten erachtet werden. Aus diesem Grund ist es für Patienten wichtig, am

Design von Studien beteiligt zu sein, um sicherzustellen, dass Informationen zu den Ergebnissen gesammelt werden, die für sie von Bedeutung sind. In den letzten Jahren wurde beispielsweise anerkannt, dass die Lebensqualität ein wichtiges Ergebnis für Patienten ist. Dies hat zur Entwicklung spezifischer Methoden zur Schaffung von Maßen für die Lebensqualität und sogenannten „Patient Reported Outcomes“ (vom Patienten berichtete Ergebnisse) in klinischen Studien geführt.

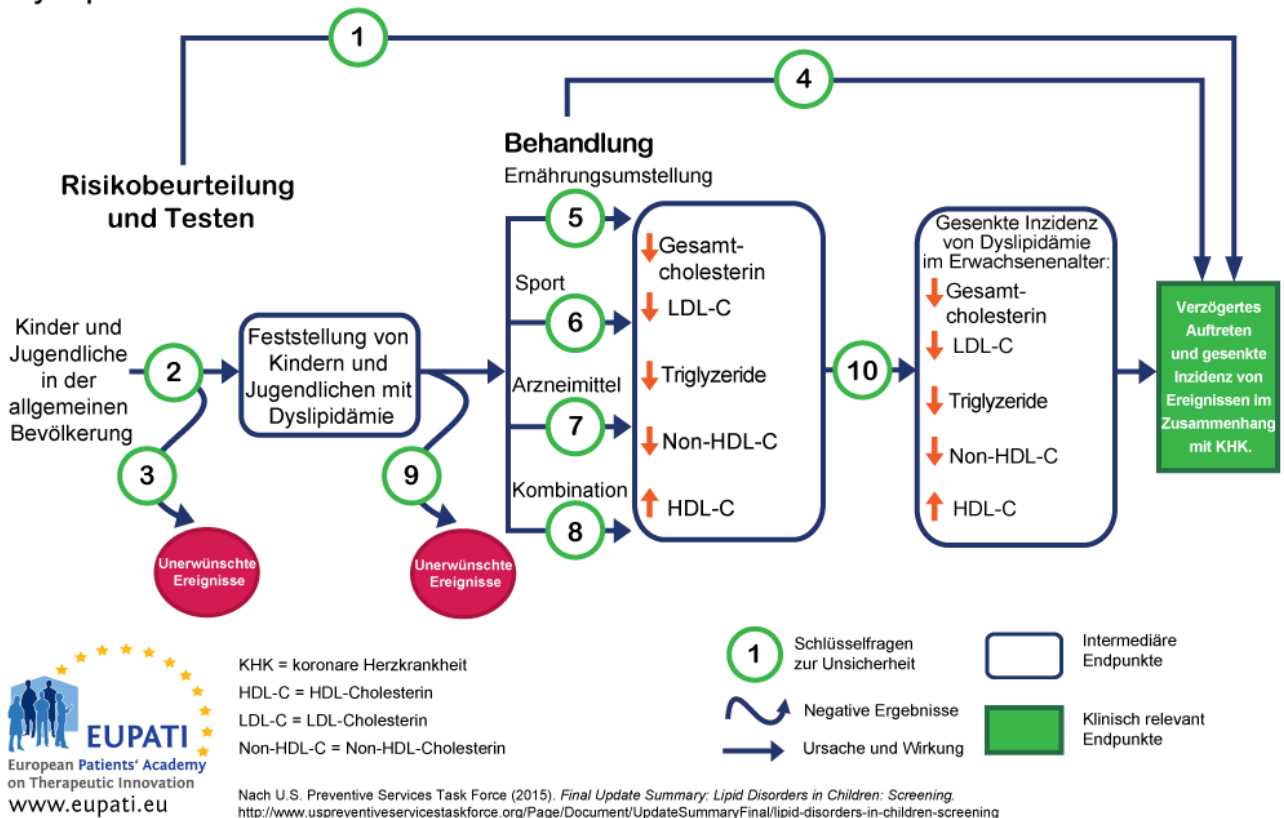
Ein Ansatz, um sicherzustellen, dass alle wichtigen Ergebnisse einer bestimmten Technologie untersucht werden, ist die Anwendung eines analytischen Rahmens – zum Beispiel in Form eines Schaubildes wie in Abbildung 1.² Analytische Rahmen sind hilfreich, um alle Ergebnisse im Zusammenhang mit einer Intervention darzustellen und hervorzuheben, welche Unsicherheiten es gibt.

Im analytischen Rahmen in Abbildung 1:

- Ursache und Wirkung werden als Pfeile dargestellt.
 - Gekrümmte Pfeile weisen auf negative Ergebnisse hin.
- Die Ergebnisse einer Verbesserung der Gesundheit (wie etwa gesenkte Mortalität) werden als Rechtecke dargestellt.
 - Rechtecke mit spitzen Ecken zeigen klinisch relevante Endpunkte (jene, die vom Patienten wahrgenommen werden, wie etwa Brustschmerzen).
 - Rechtecke mit runden Ecken zeigen intermediäre Endpunkte, einschließlich Surrogat-Endpunkte (die nicht vom Patienten wahrgenommen werden können, wie etwa der Cholesterinspiegel im Blut).
- Schlüsselfragen zur Unsicherheit können dann numerisch dargestellt werden.

Beispiel eines analytischen Rahmens

Screening und Behandlung von Kindern und Jugendlichen im Hinblick auf Dyslipidämie:



Dieser analytische Rahmen wurde verwendet, um die Stärken und Grenzen der Evidenz für die Wirksamkeit der Untersuchung von Kindern und Jugendlichen auf Dyslipidämie (Störungen des Lipidstoffwechsels) als Teil der routinemäßigen Primärversorgung festzustellen. Dyslipidämien sind wichtige Risikofaktoren für koronare Herzkrankheiten (KHK).

Die Schlüsselfragen im Zusammenhang mit diesem analytischen Rahmen lauten wie folgt:

- **Schlüsselfrage 1:** Ist das Screening im Hinblick auf Dyslipidämie bei Kindern/Jugendlichen wirksam, um das Auftreten von Ereignissen im Zusammenhang mit KHK (koronare Herzkrankheit) zu verzögern und ihre Inzidenz zu reduzieren?
- **Schlüsselfrage 2:** Wie akkurat ist das Screening im Hinblick auf Dyslipidämie bei der Identifizierung von Kindern/Jugendlichen mit einem erhöhten Risiko für

Ereignisse im Zusammenhang mit KHK und andere Ergebnisse?

- **Schlüsselfrage 3:** Was sind die unerwünschten Ereignisse des Screenings (einschließlich falsch positiver und falsch negativer Ergebnisse, Etikettierung)?
- **Schlüsselfrage 4:** Wie wirksam sind Arzneimittel, Ernährungsumstellung, Sport und Kombinationstherapie bei der Senkung der Inzidenz von Dyslipidämie im Erwachsenenalter und der Verzögerung des Auftretens und der Senkung der Inzidenz von Ereignissen im Zusammenhang mit KHK und anderen Ergebnissen (einschließlich des optimalen Alters für den Behandlungsbeginn) bei Kindern und Jugendlichen?
- **Schlüsselfragen 5-8:** Wie wirksam sind Arzneimittel, Ernährungsumstellung, Sport und Kombinationstherapie bei der Behandlung von Dyslipidämie bei Kindern und Jugendlichen (einschließlich des Zusatznutzens einer Behandlung von Dyslipidämie in der Kindheit)?
- **Schlüsselfrage 9:** Was sind die unerwünschten Ereignisse im Zusammenhang mit dem Arzneimittel, Ernährungsumstellung, Sport und Kombinationstherapie bei Kindern/Jugendlichen?
- **Schlüsselfrage 10:** Senkt eine Verbesserung der Dyslipidämie in der Kindheit das Risiko für Dyslipidämie im Erwachsenenalter?
- **Schlüsselfrage 11:** (nicht bildlich dargestellt): Welche Kostenfragen sind mit dem Screening im Hinblick auf Dyslipidämie bei symptomfreien Kindern verbunden?

Die Unterschiede zwischen Ergebnissen verstehen

Wenn alle wichtigen Ergebnisse erfasst sind, können immer noch etliche Herausforderungen beim Vergleich der Wirkungen einer neuen Technologie mit jener der Standardversorgung und anderer bestehender Behandlungen auftreten. Die Ergebnisse können auf unterschiedliche Weise gemessen werden oder zwei Technologien

können scheinbar ähnliche Ergebnisse erreichen, bis eine genauere Prüfung Unterschiede aufzeigt.

In Fällen, in denen die festgestellten wichtigen Ergebnisse schwer zu messen sind oder nie zuvor gemessen wurden, müssen die Wissenschaftler gewissenhaft ein Maß schaffen, das in einer Studie dann reproduziert werden kann. Ein Patient möchte zum Beispiel wissen, wie ein Arzneimittel ihm helfen kann, wieder arbeiten zu gehen oder das Bett zu verlassen. Wissenschaftler können eine numerische Skala zur Schmerzbeurteilung für Patienten mit Kreuzschmerzen schaffen. In anderen Fällen, wenn zum Beispiel eine Studie eine Veränderung bei einem Laborparameter misst, muss diese Veränderung in ein Maß uminterpretiert werden, das für Patienten von größerer Bedeutung ist – wie etwa die Fähigkeit, wieder zur Arbeit zu gehen.

Manchmal sind die Regulierungsbehörden, die Arzneimittel zulassen, damit zufrieden, wenn der Hersteller die Wirkung eines neuen Arzneimittels mit kurzfristigen Ergebnissen nachweist, wie etwa das Senken des Blutdrucks. Ein HTA-Gremium wird dieses kurzfristige Ergebnis in für Patienten relevantere Ergebnisse uminterpretieren müssen, wie etwa das Verhindern eines vorzeitigen Todes.

Einige Ergebnisse können zunächst eingängig erscheinen, doch bei näherer Untersuchung schwierig zu interpretieren sein. Eine Senkung des Risikos der Fünf-Jahres-Mortalität (Tod innerhalb von fünf Jahren) um 50 % heißt zum Beispiel nicht, dass das Arzneimittel einen vorzeitigen Tod verhindern kann. Es könnte es schlicht bedeuten, dass:

- die Lebenserwartungen bei einigen Patienten von 4,9 auf 5,1 Jahre (oder noch schlechter von 4,99 Jahren auf 5,01 Jahre) erhöht wird **oder**
- die Krankheit bei einigen sehr wenigen geheilt, bei anderen das Überleben jedoch in keiner Weise verlängert wird.

Auch wenn Unterschiede bei Messungen, die für Patienten von Bedeutung sind, beobachtet werden, können diese trotzdem schwierig zu interpretieren sein. Studien können zum Beispiel darauf hinweisen, dass ein neues Arzneimittel das Risiko für Krankenhausaufenthalte aufgrund einer Infektion um 33 % senkt. Das kann jedoch Verschiedenes heißen. Es könnte bedeuten, dass:

- 33 von 100 Personen, die das Arzneimittel einnahmen und ansonsten ins Krankenhaus eingewiesen worden wären, einem Krankenhausaufenthalt entgangen sind (dies wird als eine **absolute** Risikoreduktion bezeichnet), **oder**
- die Wahrscheinlichkeit eines Krankenhausaufenthalts um 33 % im Verhältnis zur Wahrscheinlichkeit eines Krankenhausaufenthalts ohne Arzneimittel gesenkt wird (dies wird als eine **relative** Risikoreduktion bezeichnet). Wenn die Wahrscheinlichkeit eines Krankenhausaufenthalts ohne das neue Arzneimittel 3 von 1 000 beträgt, dann senkt eine Reduktion um 33 % diese auf 2 von 1 000. Das heißt, dass 1 von 1 000 Personen, die das Arzneimittel einnehmen, einen Nutzen erzielt. Dies ist ein recht großer Unterschied zu den 33 von 100 Personen, die im Beispiel oben von dem Arzneimittel profitieren.

Eine letzte Herausforderung, um die Unterschiede zwischen einer neuen Gesundheitstechnologie und der Standardversorgung zu verstehen, ist die Verwendung und der Missbrauch statistischer Tests. Statistische Tests dienen dazu, den Forschern zu helfen festzustellen, ob die von ihnen bestimmten Unterschiede wahrscheinlich real sind. Oftmals wird dies in Form eines p-Wertes angegeben. Allerdings spiegeln p-Werte nicht wider, wie groß der Unterschied ist oder ob dieser Unterschied für Patienten von Bedeutung ist. Das bedeutet, dass p-Werte für Patienten und Versorger im Allgemeinen nicht nützlich sind, um Entscheidungen zu treffen.

Weitere statistische Maße sind die Konfidenzintervalle.

Konfidenzintervalle sind hilfreicher, da sie ein gewisses Gefühl für die Größe des Unterschieds zwischen der neuen Gesundheitstechnologie und der Standardversorgung geben. Konfidenzintervalle spiegeln auch jegliche Unsicherheiten bezüglich der Schätzung des Ausmaßes des Unterschiedes wider. Es kann zum Beispiel berichtet werden, dass ein neues Arzneimittel die Wahrscheinlichkeit, in der Zukunft einen Herzinfarkt zu erleiden, im Verhältnis zur derzeitigen Wahrscheinlichkeit, einen Herzinfarkt zu erleiden, um 33 % senkt (mit einem 95-%-Konfidenzintervall von 5 % bis 45 %).

Unterschiede bewerten

Die letzte Herausforderung ist es zu verstehen, wie die Unterschiede zwischen den Ergebnissen wahrzunehmen und zu bewerten sind. Wenn ein Arzneimittel das Leben um 0,2 Jahre verlängert, muss das HTA-Gremium trotzdem noch Folgendes wissen:

- wie hoch würde ein Patient angesichts der erwarteten Nebenwirkungen und anderen Probleme 0,2 Jahre bewerten
- ob alle Patienten in etwa die gleichen Erfolge erzielen oder ob deutliche Unterschiede bei den Patienten bestehen **und**
- ob alle Patienten diese Erfolge ähnlich bewerten.

Ein neues Arzneimittel, das die Lebenserwartung um durchschnittlich 0,2 Jahre verlängerte, würde unterschiedlich wahrgenommen werden, wenn es bei einigen Patienten wirkt, bei anderen jedoch nicht, wenn es mit einem Szenarium verglichen wird, in dem alle Patienten 0,2 Jahre mit geringfügigen Unterschieden zwischen den Patienten erzielen.

Es gibt verschiedene Mechanismen, die verwendet werden können, um den relativen Wert zu verstehen, den Patienten und Versorger Unterschieden bei Gesundheitsergebnissen beimessen. Einer ist eine qualitative Forschung, wie etwa Umfragen oder Fokusgruppen, die dazu dienen, ein Verständnis dessen zu

liefern, welche Ergebnisse für Patienten am wichtigsten sind. Ein anderer ist quantitative Forschung auf Grundlage von Umfragen unter Patienten, die präzise numerische Werte für die Wichtigkeit, die verschiedenen Gesundheitszuständen beigemessen werden, vergeben.

Kurz gesagt, die Beurteilung sollte die folgenden Fragen beantworten:

1. Wie umfassend waren die Informationen?
2. Wie präzise sind die Informationen?
3. Fehlt etwas?
4. Wie verständlich sind die Informationen?

Quellenangaben

1. *HTA Core Model*. Retrieved 7 December, 2015, from <http://www.eunethta.eu/hta-core-model>
2. U.S. Preventive Services Task Force (2015). Final Update Summary: Lipid Disorders in Children: Screening. Retrieved 7 December, 2015, from: <http://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/lipid-disorders-in-children-screening>

A2-6.03.1-v1.1