

Die Bedeutung von Biomarkern für eine personalisierte Medizin

Die moderne Therapieforschung stellt den Ärzten ein Armamentarium von gut untersuchten Arzneimitteln zur Verfügung, die sie im Kampf gegen Krankheit und Leiden zum Einsatz bringen können. Dennoch scheint es eine unumstößliche Tatsache zu sein, dass alle wirksamen Medikamente auch Nebenwirkungen haben, die, oft unvorhersehbar, bestimmten Patienten Schaden zufügen. Die im Rahmen der Zulassung von neuen Arzneimitteln durchgeführten Studien sollen die Wirksamkeit der experimentellen Therapie belegen und eine Datenbasis zur Verfügung stellen, die ein positives Nutzen-Risiko-Verhältnis abschätzen lässt. Das bedeutet, dass der zu erwartende Nutzen den möglichen Schaden der Anwendung des Arzneimittels überwiegen muss. Sofern erkenntlich, werden Untergruppen in der Fach- und Gebrauchsinformation (Langversion des Beipackzettels) kontraindiziert, für die dieses Nutzen-Risiko-Verhältnis nicht positiv erscheint. Beispiele liefern Medikamente, die bei Patienten mit Niereninsuffizienz nicht oder nur in reduzierter Dosis eingenommen werden dürfen, oder die Klasse der Sartane, die bei Farbigen nicht wirksam sind. Sind derartige Untergruppen nicht identifizierbar, so müssen die beobachteten Nebenwirkungen benannt werden, sodass der Patient informiert ist, welche Risiken bestehen, wenn er ein für ihn erforderliches Arzneimittel zur Behandlung seiner Erkrankung einnimmt. Darüber hinaus gibt es viele Erkrankungen,

die sich auch weiterhin einer wirksamen medikamentösen Behandlung entziehen, weil es möglicherweise nicht gelingt, den Teil der Patientenpopulation zu identifizieren, der von einem bestimmten Arzneimittel oder Wirkprinzip profitieren würde (Alzheimer, vgl. [1]). Die klassische Chemotherapie onkologischer Erkrankungen ist mit substanziellen und unter Umständen sogar lebensbedrohlichen Nebenwirkungen verbunden, die alle Patienten treffen, während bekanntermaßen nur eine bestimmte Subpopulation davon profitiert. So fand sich beispielsweise in einer multizentrischen, randomisierten Studie von Scagliotti et al. [2] bei 1725 Patienten mit fortgeschrittenem nichtkleinzelligem Bronchialkarzinom (NSCLC) trotz einer globalen Nicht-Unterlegenheit von Pemetrexed im Vergleich zu Gemcitabine die Untergruppe der Patienten mit squamösen Tumoren. In dieser Untergruppe wurde ein signifikant schlechteres Überleben der Patienten unter Pemetrexed beobachtet.

All diese Beispiele verbindet die große Hoffnung, dass durch die Identifizierung von Biomarkern (objektiv am Patienten messbare Merkmale) vorhergesagt werden kann, ob in einer bestimmten Patientengruppe bei der Einnahme eines Medikaments eine schwere Nebenwirkung zu erwarten ist, das Arzneimittel nicht wirkt und nur die Nebenwirkungen toleriert werden müssen oder ob eine Teilpopulation identifiziert werden kann, in der ein Medikament wirkt, das global wirkungs-

los zu sein scheint. Biomarker stehen im Interesse der Forschung und sind Teil der Zukunftsvision der personalisierten Medizin, in der Therapien individualisiert für einzelne Patienten bzw. Patientengruppen dargestellt werden können. Erste Beispiele existieren mit Herceptin zur Behandlung des HER+-Brustkrebses [3], EGFR-Mutationen des NSCLC, die mit Gefitinib [4], und die K-RAS-Mutation beim kolorektalen Karzinom, die mit Cetuximab [5] behandelt werden können.

Gemäß der Terminologie der FDA heißt ein Biomarker „Companion Diagnostics“, wenn er geeignet ist,

1. Patienten zu identifizieren, die mit großer Wahrscheinlichkeit von einem Arzneimittel profitieren werden,
2. Patienten zu identifizieren, die mit großer Wahrscheinlichkeit durch die Anwendung des Arzneimittels einen Schaden erleiden werden, oder
3. wenn er geeignet ist, den Behandlungserfolg zu monitorieren und die Therapie zu steuern [6].

Nicht nur weil Companion Diagnostics Kosten verursachen, sondern vor allem weil sie kritisch für die Anwendung eines Arzneimittels sind, müssen sie im Rahmen der Arzneimittelentwicklung validiert werden, da sie quasi Teil der Behandlungsvorschrift sind.

Während die prinzipiellen Anforderungen an einen validen Wirksamkeitsnachweis für ein neues Arzneimittel relativ klar umrissen sind (und damit auch

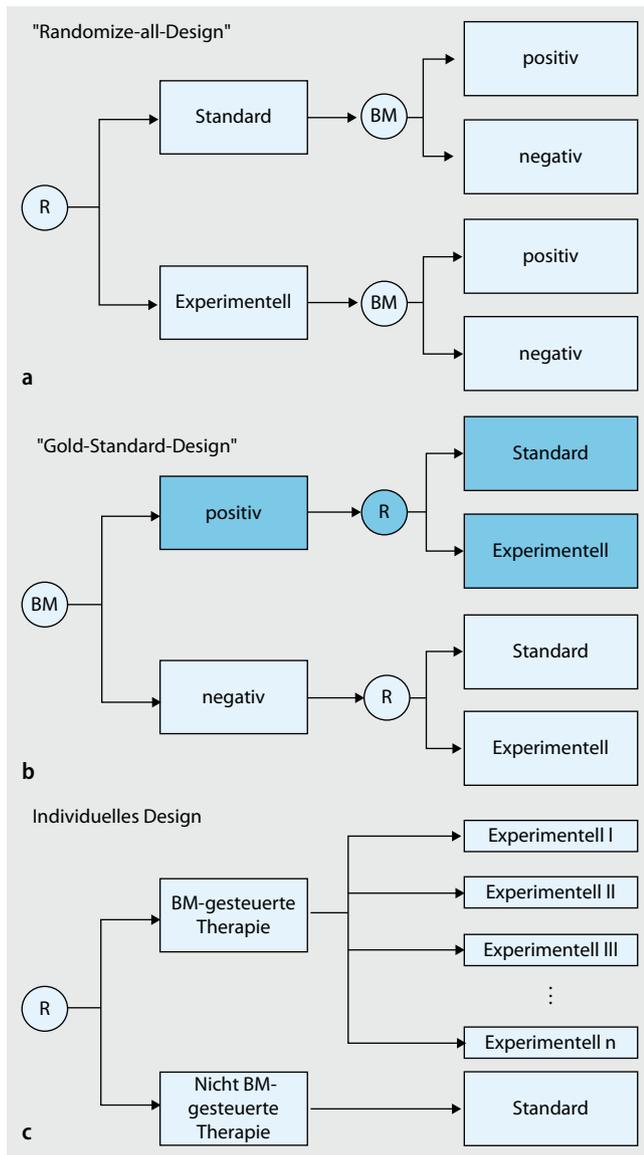


Abb. 1 ◀ Studiendesigns für Biomarkerstudien. *R* Randomisierung, *BM* Biomarker, *n* Anzahl experimentelle Therapien

klar ist, was in einer Population von Biomarker-positiven Patienten zu belegen ist), ist bisher nicht viel darüber ausgesagt worden, wie der Nutzen der Biomarker-Testung zu belegen ist.

Insbesondere ist bisher nicht spezifiziert, welche und wie viel Evidenz in der Population der Biomarker-negativen Patienten vorzulegen ist, bevor die Testung des Arzneimittels in der durch den Biomarker identifizierten Untergruppe fortgeführt werden kann. Da hier die Epidemiologie der Erkrankung, die Erfolgsrate der therapeutischen Alternativen und viele andere Aspekte eine wichtige Rolle spielen, können die entsprechenden Überlegungen nur am Beispiel dargestellt und methodische Lösungsansätze

darüber vorgestellt werden, wie innerhalb eines Studienprogramms alle relevanten Informationen zusammengetragen werden können.

Im vorliegenden Beitrag wird zunächst eine Reihe von Studiendesigns vorgestellt, mit deren Hilfe der Nutzen eines Biomarkers untersucht werden kann. An einem Beispiel werden die Überlegungen vorgestellt, die zum Zeitpunkt der Planung eines gemeinsamen Entwicklungsprogramms für einen Biomarker zur Patientenselektion und ein Arzneimittel zu beachten wären. Im Rahmen der Diskussion werden Alternativen untersucht, die die Validierung des Biomarkers in das gesamte Entwicklungsprogramm des Arzneimittels mit einbinden.

Studiendesigns zur Untersuchung der Relevanz von Biomarkern

Studiendesigns zur Bewertung von Biomarkern und Therapien

Eine umfangreiche Übersicht über mögliche Studiendesigns bei Biomarkerstudien gibt [7]. Darüber hinaus haben wir in einer eigenen Übersichtsarbeit die wichtigsten Vor- und Nachteile dieser Studiendesigns diskutiert [8].

Die klassische Vorgehensweise ist das sog. „Randomize-all-Design“ (■ **Abb. 1a**), in dem zunächst alle Patienten randomisiert dem Standard- oder dem experimentellen Therapiearm zugeteilt werden. Später wird, unabhängig von der Randomisierung, nach dem Befund des Biomarkers adjustiert, und es werden ggf. Interaktionen und/oder Untergruppen betrachtet. Dieses Design eignet sich sowohl für eine retrospektive als auch prospektive Untersuchung eines Biomarkers. Insbesondere kann mit dieser Herangehensweise aufgezeigt werden, dass der Biomarker eine Untergruppe identifizieren kann, die besonders von der Therapie profitiert. Ferner kann der Algorithmus, in dem der Biomarker zur Therapiesteuerung verwendet werden soll, entwickelt und verbessert werden. Jedoch kann der Effekt des Biomarkers nicht unabhängig vom Therapieeffekt geschätzt werden. Die Schlussfolgerungen hängen darüber hinaus von der Vollständigkeit der Daten ab und bedürfen mindestens einer unabhängigen Validierung. Speziell bei retrospektiven Biomarkerstudien, bei denen die Biomarker nicht vorab spezifiziert sind, sondern ganze Panel untersucht und/oder zunächst Cut-offs identifiziert werden sollen, muss anhand zusätzlicher Maßnahmen wie unabhängiger Replikation die Glaubwürdigkeit der Ergebnisse unter Beweis gestellt werden. Nur mit mindestens einer weiteren unabhängigen positiven Studie kann ausgeschlossen werden, dass die detektierte Untergruppe ein Zufallsbefund ist.

Das Pendant zum Randomize-all-Design sind Studiendesigns, in denen der Biomarker vor der Randomisierung betrachtet wird (■ **Abb. 1b**). Sind die Ressourcen knapp, wird eine neue Therapie

mit einem Selektionsdesign untersucht. In diesem werden zunächst die Biomarker-positiven Patienten ausgewählt und anschließend dem Standard- oder dem experimentellen Therapiearm randomisiert zugeteilt (dunkelblau hinterlegter Studienteil in **Abb. 1b**). Jedoch kann dadurch, dass die Biomarker-negativen Patienten nicht behandelt werden, der Nutzen des Biomarkers nicht bewertet werden.

Der Nutzen eines Biomarkers für die Therapieentscheidung kann somit nur in einer Studie belegt werden, in der sowohl Biomarker-positive als auch Biomarker-negative Patienten untersucht werden. Daher werden im sog. „Gold-Standard-Design“ die Patienten zunächst mithilfe eines Biomarkers stratifiziert und anschließend randomisiert dem Kontroll- oder dem experimentellen Therapiearm zugeteilt (**Abb. 1b**). Auf diese Weise können anhand der Studienergebnisse sowohl Aussagen zur Wirksamkeit der Therapie als auch zum Nutzen der Biomarkertestung abgeleitet werden. Alle anderen Designs können jeweils nur Teilaspekte validieren und müssen dann zusätzliche Annahmen treffen, um die gewünschte Schlussfolgerung in Bezug auf die Wirksamkeit einer Therapie und den Nutzen des Biomarkers zur Populationsselektion zu belegen.

Ein Extrem ist der Fall, dass nicht mehr nur für einige Subgruppen von Patienten mithilfe eines Biomarkers eine spezifische Therapie ausgewählt wird, sondern, dass für wenige oder sogar einzelne Patienten eine spezielle Therapie vorgeschlagen wird (**Abb. 1c**). In diesem Fall kann dann nicht mehr die Wirksamkeit einer Therapie, sondern lediglich der Algorithmus zur Therapiesteuerung validiert werden. Anhand einer solchen Studie kann gezeigt werden, dass ein global besseres Ergebnis als für eine fixe Therapieentscheidung unter bestimmten Bedingungen erzielt wird.

Spezielle Biomarkerstudiendesigns zur gleichzeitigen Auswahl einer Subgruppe und zum Wirksamkeitsnachweis einer Therapie

Im klassischen Fall werden in einer frühen Phase des klinischen Entwicklungsprogramms Biomarker ausgewählt und die Auswahl einer Subgruppe mithilfe eines oder mehrerer Biomarker definiert. Nachfolgend wird in einer großen Phase-III-Studie die Wirksamkeit in dieser Subgruppe nachgewiesen, u. U. nach Durchführung weiterer Phase-II-Studien, z. B. zur Bestimmung der Dosis-Wirkungs-Beziehung. Zur Beschleunigung des klinischen Programms wurden nun adaptive Designs vorgeschlagen, die Elemente der Phase II mit einer Wirksamkeitsstudie der Phase III verbinden. Beispielsweise kann eine solche Phase-II/III-Studie mehrere Dosisgruppen untersuchen und die sinnvollste Dosis in einer Zwischenauswertung auswählen, die Studie mit dieser Dosis weiterführen und hierfür den eigentlichen Wirksamkeitsnachweis erbringen. Dabei müssen entsprechende statistische Methoden angewandt werden, um die Anzahl der möglichen zu testenden Hypothesen zu berücksichtigen (Problem der Multiplizität). Solche Studien wurden bereits mehrfach erfolgreich durchgeführt.

Ähnlich zu diesen adaptiven Studien mit Dosisselektion wurden in jüngerer Vergangenheit auch Studiendesigns diskutiert, die Subgruppen auswählen und gleichzeitig den Wirksamkeitsnachweis in dieser Subgruppe erbringen. Unabhängig davon, ob die Auswahl der Subgruppe in einer Zwischenauswertung oder am Ende der Studie erfolgt, muss ebenso wie im Fall der Dosisselektion dem Problem der Multiplizität mit entsprechenden statistischen Methoden begegnet werden. Dabei sollte die Anzahl der Möglichkeiten sinnvollerweise klein gehalten werden. Würde z. B. am Ende der Studie eine Subgruppe wegen hoher Wirksamkeit ausgewählt werden, wäre die sich aus denselben Daten ergebende Wirksamkeitsschätzung offensichtlich zugunsten der neuen Therapie verzerrt, da die beobachtete Wirksamkeit zum Teil ein Zufallsbefund darstellt. Der zugehörige statistische Test wäre invalide, da der Fehler 1. Art, d. h. die

Bundesgesundheitsbl 2013 · 56:1480–1488
DOI 10.1007/s00103-013-1828-9
© The author(s) 2013. This article is published with open access at link.springer.com

A. Großhennig · N. Benda · A. Koch

Die Bedeutung von Biomarkern für eine personalisierte Medizin

Zusammenfassung

Im vorliegenden Beitrag wird die Relevanz von Biomarkern zur Bewertung des Nutzen-Risiko-Verhältnisses eines Arzneimittels und deren Validierung im Rahmen von klinischen Studien dargelegt. Zu diesem Zweck werden die wichtigsten Studiendesigns zur Validierung vorgestellt und diskutiert. Am Beispiel der Entwicklung von Rosuvastatin und der Validierung von hs-CRP (JUPITER-Studie) werden Aspekte, die bei der Planung eines gemeinsamen Entwicklungsprogramms für ein Arzneimittel und einen Biomarker notwendig sind, erörtert. Darüber hinaus werden Alternativen diskutiert, wie die Validierung des Biomarkers in das gesamte Entwicklungsprogramm des Arzneimittels eingebunden werden kann.

Schlüsselwörter

Biomarker · Personalisierte Medizin · Companion diagnostics · Versuchsplanung · Arzneimittelentwicklung

The relevance of biomarkers for personalised medicine

Abstract

In this paper the relevance of biomarkers for the benefit/risk assessment of pharmaceuticals and the validation of biomarkers within clinical trials is presented. For this purpose the most important study designs for validation are described and discussed. Using the example of the development of rosuvastatin and the validation of hs-CRP (JUPITER study), aspects that are necessary for the planning of a joint development program for a pharmaceutical drug and a biomarker are discussed. In addition, alternatives considering how the validation of the biomarker could be involved in the entire development program of the pharmaceutical drug are presented.

Keywords

Biomarker · Personalised medicine · Companion diagnostics · Design of experiments · Drug development

Wahrscheinlichkeit, per Zufall ein statistisch signifikantes Ergebnis zu erhalten, obwohl keine Wirksamkeit vorliegt, höher ist als nominal angenommen.

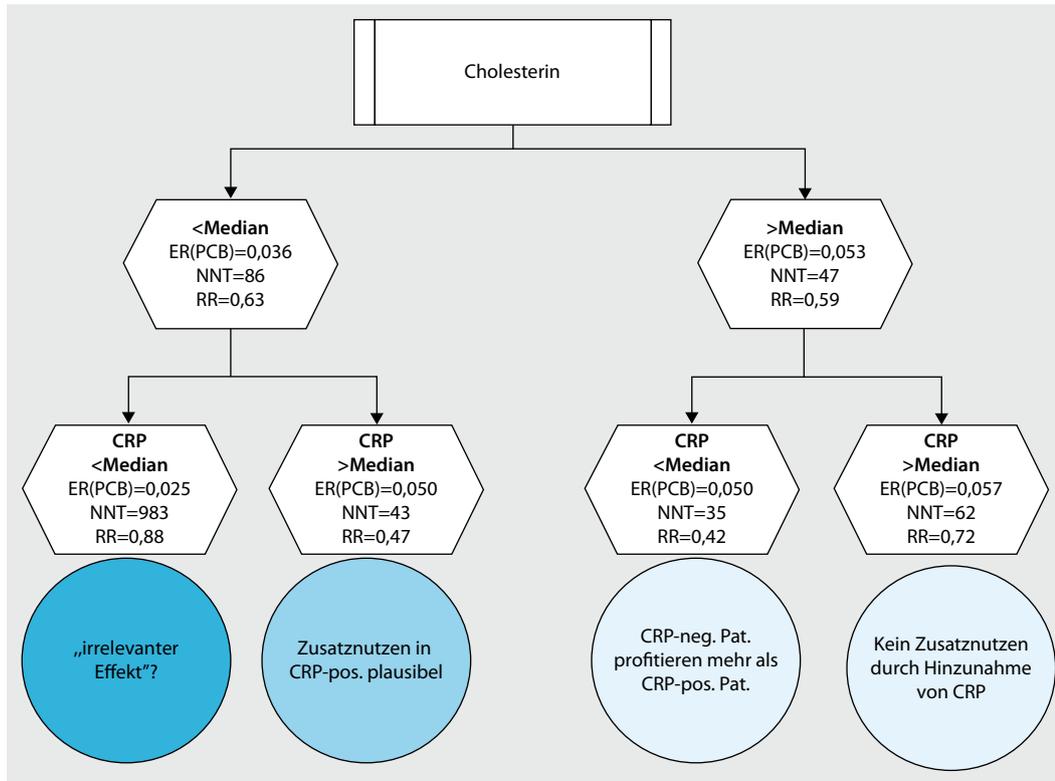


Abb. 2 Ergebnisse der Validierungsstudie AF-CAPS/TexCAPS [13] in Subgruppen nach dem CRP-Status. CRP C-reaktives Protein, ER(PCB) Ereignisrate im Placeboarm, NNT Number needed to treat, RR relatives Risiko

Ebenso erhielt man eine Verzerrung falls die Auswahl in einer Zwischenauswertung erfolgt und die Daten, die hierfür erhoben wurden, für den Wirksamkeitsnachweis mit benutzt würden, falls keine korrekte Adjustierung für die Anzahl möglicher Subgruppen vorgenommen würde. Würde z. B. aus einer größeren Anzahl von Biomarkern ein Biomarker und damit eine optimale Selektionsregel (z. B. mittels eines Grenzwertes für die Expression dieses Biomarkers) ausgewählt werden, so wäre die Anzahl der Möglichkeiten unendlich oder zumindest sehr groß. Eine entsprechende Adjustierung wäre somit unmöglich oder ineffizient. Stattdessen müsste die Anzahl der präzise definierten Subgruppen sehr klein sein, damit ein entsprechendes Studiendesign sinnvoll sein kann. Im einfachsten Fall wurden der Biomarker und die Selektionsregel vor der Studie definiert, während in der Studie die Frage beantwortet wird, ob diese Subpopulation oder die gesamte für die entsprechende Indikation infrage kommende Population ausgewählt wird, d. h., die Anzahl der Möglichkeiten beschränkt sich auf zwei. Eine solche Möglichkeit wurde in einem Artikel von Brannath et al. [9] untersucht.

Hierbei beziehen sich die Autoren auf eine Ereigniszeit als primären Endpunkt, d. h. im speziellen Beispiel auf die Zeit bis zum Auftreten einer Tumorphase oder dem Tod des Patienten („progression free survival“, PFS) zur Untersuchung einer neuen Krebstherapie.

Aus statistischer Sicht sind solche adaptiven Designs durch 2 methodische Grundelemente gekennzeichnet. Dies ist zum einen die Entscheidungsregel, die auf Grundlage der Zwischenauswertung eine Entscheidung für die Einschränkung auf eine prädefinierte Subpopulation trifft (die Selektionsregel), und zum anderen die multiple statistische Teststrategie für den eigentlichen Wirksamkeitsnachweis, die unter Berücksichtigung der verschiedenen Möglichkeiten sicherstellen muss, dass die Wahrscheinlichkeit einer falsch positiven Studie den prädefinierten Fehler 1. Art nicht überschreitet. Für beide Elemente sind verschiedene Methoden denkbar.

Es ist möglich und erscheint sinnvoll, für die multiple Teststrategie allein die p-Werte der entsprechenden Tests nach Zwischen- und Endauswertung zu verwenden, was es erlaubt, die Selektionsregel unabhängig hiervon zu definieren,

denn in diesem Fall hat die Selektionsregel keine Auswirkung auf die Validität des Wirksamkeitsnachweises. Brannath et al. [9] schlagen eine sog. Bayesische Entscheidungsregel vor, die für die Wirksamkeit in der Subpopulation und in der gesamten Population eine Vorannahme ausgedrückt in einer Wahrscheinlichkeitsverteilung macht, die entsprechend den in der Zwischenauswertung gewonnenen Daten angepasst wird. Bemerkenswert dabei ist, dass für den eigentlichen Wirksamkeitsnachweis ein solches mit einer subjektiven Komponente versehenes Vorgehen nicht den üblichen Anforderungen einer pivotalen Phase-III-Studie entsprechen würde, diese Entscheidungsregel aber den letztendlichen Wirksamkeitsnachweis nicht tangiert und damit auch aus regulatorischer Sicht akzeptabel ist, solange die Kriterien für den unabhängigen und unvoreingenommenen Wirksamkeitsnachweis eingehalten werden (Einhaltung des Fehlers 1. Art für eine falsch positive Entscheidung durch die Teststrategie, Entscheidung für den Studienerfolg ausschließlich auf Grundlage der Studiendaten).

Die Auswahl der Subpopulation, d. h. die Entscheidung in der Zwischenauswer-

tung, sich im weiteren Verlauf der Studie nur noch auf Patienten dieser Subpopulation zu beschränken oder nicht, lässt sich mit der Entscheidung, die Studie wegen Erfolg oder Unwahrscheinlichkeit eines Erfolges („futility“) abzubrechen, kombinieren, wobei für den ersten Fall eine weitere Anpassung der statistischen Tests bezüglich Multiplizität vorzunehmen ist.

Entscheidend bei einem adaptiven Design, in dessen Verlauf man sich für eine durch Biomarker definierte Teilpopulation entscheidet und an dessen Ende ein Wirksamkeitsnachweis erbracht werden soll, sind die Auswahl einer validen Teststrategie, die die Multiplizität der möglichen Entscheidungen berücksichtigt – d. h. die Ablehnwahrscheinlichkeit unter der Nullhypothese der Unwirksamkeit durch die vorgegebene Schranke beschränkt – und eine sinnvolle Entscheidungsregel für die Auswahl. Letztere wird durch den Ansatz der kombinierten p-Werte erleichtert. Wie bereits erläutert, ist es zwar nicht erforderlich, statistisch signifikant zu zeigen, dass die Therapie im komplementären Teil der Patientenpopulation unwirksam oder geringer wirksam ist, jedoch sollten eindeutige Hinweise hierfür vorliegen. In diesem Sinne kann eine Bayesianische Strategie für die Auswahl, wie von Brannath et al. [9] vorgeschlagen, zwar sinnvoll sein, aufgrund der subjektiven Komponente sollten jedoch die Daten für den Vergleich beider Patientengruppen auch für sich allein sprechen. Ebenso sind andere Regeln, soweit eine sinnvolle Entscheidung nachzuvollziehen ist, möglich. Entscheidend im Sinne der Zulassung ist die valide statistische Testprozedur der Wirksamkeit in der ausgewählten Population, d. h. je nach Entscheidung entweder in der gesamten oder in der Teilpopulation.

Beispiel: JUPITER-Studie

Studiendesign und wichtigste Ergebnisse

JUPITER [10] ist eine randomisierte doppelblinde placebokontrollierte multizentrische Studie zum Nachweis der Wirksamkeit von Rosuvastatin in der Primärprävention kardiovaskulärer Erkrankungen. In dieser Studie wurde jedoch eine

Population untersucht, die üblicherweise nicht mit einem Statin zur Reduktion des LDL-Cholesterins behandelt werden würde. Insgesamt 17.802 Frauen und Männer mit LDL-Werten unter 130 mg pro Deziliter (mg/dl) und einem hs-CRP-Wert (hoch-sensitives C-reaktives Protein) von mindestens 2,0 mg pro Liter (mg/l) wurden entweder in den Rosuvastatin- (20 mg täglich) oder Placebostudienarm randomisiert. Der zusammengesetzte primäre Endpunkt beinhaltete das erstmalige Auftreten eines kardiovaskulären Ereignisses (Herzinfarkt, Schlaganfall, arterielle Revaskularisation, stationäre Aufnahme aufgrund einer instabilen Angina pectoris oder Tod aufgrund eines kardiovaskulären Ereignisses). Obwohl eine Nachbeobachtungszeit von 5 Jahren geplant war, wurde die Studie bereits nach einer medianen Follow-up-Zeit von 1,9 Jahren vorzeitig beendet. In der Rosuvastatin-Gruppe wurde eine erhebliche Reduktion der kardiovaskulären Ereignisse beobachtet (142/8901 vs. 251/8901; Hazard Ratio: 0,56, 95%-Konfidenzintervall: 0,46–0,69).

Diskussion des Studiendesigns und der Studienergebnisse

JUPITER belegt in einer durch einen Biomarker definierten Population die Wirksamkeit einer Therapie mit Rosuvastatin. Auch unter Berücksichtigung der Risiken einer Statintherapie wird insgesamt das Nutzen-Risiko-Verhältnis in dieser Population positiv eingeschätzt [10]. Da ein Selektionsdesign gewählt wurde, kann in JUPITER offensichtlich nicht gezeigt werden, dass hs-CRP-negative Patienten in der Primärprävention kardiovaskulärer Erkrankungen nicht ebenfalls von Rosuvastatin profitieren würden.

Erschwerend kommt hinzu, dass der in JUPITER beobachtete Therapieeffekt deutlich größer ausfällt, als das nach den bisherigen Outcome-Studien bei Patienten mit zu hohem Cholesterinspiegel (z. B. WOSCOP [11], PROSPER [12]) erwartet worden wäre. Selbst wenn also für Biomarker-negative Patienten der Therapieeffekt kleiner sein sollte, könnte er immer noch klinisch relevant sein.

Hinzunahme externer Information

Die Autoren der JUPITER-Studie [10] begründen anhand der früher durchgeführten und publizierten AFCAPS/TexCAPS-Studie [13], dass in JUPITER eine nochmalige Untersuchung der hs-CRP-negativen Patienten nicht erforderlich gewesen wäre. In der AFCAPS/TexCAPS-Studie wurde bereits anhand eines Randomize-all-Designs der Zusammenhang zwischen Cholesterin und CRP und dem Therapieeffekt von Lovastatin im Vergleich zu Placebo untersucht. In dieser Studie wurde bei Patienten mit niedrigem Cholesterinspiegel und niedrigem hs-CRP lediglich eine Reduktion des relativen Risikos für ein kardiovaskuläres Ereignis von rund 12 relativen Prozent gezeigt und als nicht relevant eingestuft. Die Autoren rechtfertigen das Weglassen der CRP-negativen Patienten ferner mit einer Nichtmachbarkeit hinsichtlich der notwendigen Fallzahl, wenn man die Wirksamkeit einer Statintherapie auch für diese Subpopulation würde führen wollen. Die Ergebnisse sind in **Abb. 2** zusammengestellt.

Offensichtlich ist sowohl die absolute als auch die relative Verbesserung des primären Endpunkts bei Patienten mit niedrigem Cholesterinspiegel und niedrigem hs-CRP am geringsten (hohe NNT und geringe Verbesserung des relativen Risikos). Wohl aus diesem Grund bemerken die Autoren der JUPITER-Studie, dass ein Wirksamkeitsnachweis in dieser Population nicht machbar wäre [10]. Inkonsistent ist jedoch, dass Patienten mit niedrigem hs-CRP, aber hohem Cholesterinspiegel mehr profitieren als die Gruppe der Patienten mit hohem Cholesterinspiegel und hohem hs-CRP. In dieser Studie ist der Marker hs-CRP folglich kein unabhängiger Prädiktor für einen großen Nutzen, sondern muss mit dem Gesamtcholesterin zusammen betrachtet werden. Ebenso ist hier nicht belegt, dass lediglich hs-CRP zur Patientenselektion geeignet ist oder ob auch andere Tests (z. B. die Framingham-Kriterien) angewendet werden können.

Sollen die hier gefundenen Ergebnisse für die Anwendung des Biomarker-Tests im Zusammenhang mit Rosuvastatin genutzt werden, wird jedoch auch gleichzeitig implizit angenommen, dass die gezeigt

Tab. 1 Powerkalkulation für Biomarker-negative Patienten der JUPITER-Studie

Szenario	Ereignisrate Nullhypothese (%)	Nichtunterlegenheitsschranke (%)	Ereignisrate Alternativhypothese (%)	Fallzahl	Power (%)
1	2,5	20	2,000	500	79
2	2,5	20	2,000	750	84
3	2,5	15	2,125	500	72
4	2,5	15	2,125	1000	79
5	2,5	15	2,125	1500	84
6	2,5	10	2,250	500	65
7	2,5	10	2,250	1000	70
8	2,5	10	2,250	1500	74
9	2,5	10	2,250	2500	80

ten Effekte nicht spezifisch für ein Statin sind, sondern einen Klasseneffekt darstellen. Dann wäre tatsächlich der Nachweis zu erbringen, dass hs-CRP für mehr als ein Statin genutzt werden kann, und damit diejenigen Patienten zu wählen, die von einer Statintherapie einen deutlichen Vorteil haben.

Wirksamkeitsbeleg für die Biomarkertestung und die Therapie aus einer Studie

Aus den vorgenannten Gründen steht die prospektive Untersuchung eines Biomarkers im direkten Zusammenhang mit der Prüfung der Wirksamkeit der „Gold-Standard“-Ansatz. Dafür sind die folgenden Kriterien zu diskutieren.

Vorüberlegungen: Unter welchen Bedingungen würde man hs-CRP-negativen Patienten eine Behandlung vorenthalten können?

Letztlich gibt das Arzneimittelgesetz den geeigneten Rahmen vor. Allgemein kann eine Zulassung verweigert werden, wenn für eine Population

- I. keine signifikante Überlegenheit zu Placebo gezeigt werden könnte (hier: fehlende Wirksamkeit in der Population der Biomarker-negativen Patienten) oder wenn
- II. das Nutzen-Risiko-Verhältnis nicht positiv einzuschätzen wäre (weil a) nur eine irrelevante Überlegenheit zu Placebo bei gleichzeitiger Unsicherheit bezüglich der Risiken der Therapie gezeigt ist oder b) die Risiken im direkten Verhältnis zum belegten Nutzen zu groß sind).

Dabei ist (I) zum Zeitpunkt der Studienplanung nicht direkt diskutierbar, da nicht signifikante Therapieeffekte immer noch klinisch relevant sein können. Folglich ist zu klären, ob es möglich ist, eine Übereinkunft bezüglich einer Grenze für die minimale Größe des Therapieeffektes zu treffen, die, wenn unterschritten, eine Behandlung nicht mehr rechtfertigt. Diese Grenze könnte in Analogie zur Diskussion im Rahmen der Festlegung von Nichtunterlegenheitsschranken (Irrelevanzschranken) mit den Zulassungsbehörden „ausgehandelt“ werden. In ähnlicher Weise können spezifische Risiken üblicherweise zum Zeitpunkt der Studienplanung nicht oder nur sehr begrenzt eingeschätzt und einer bestimmten Subpopulation zugeordnet werden.

Im aktuellen Beispiel ist klar, dass diese untere Schranke deutlich kleiner als der Wert sein muss, der in den früheren Statinstudien belegt wurde, da dieser letztlich die gängige therapeutische Praxis rechtfertigt. In der WOSCOP-Studie wurde beispielsweise eine 20%ige Verbesserung gezeigt [11]. Es müsste also theoretisch eine Nichtunterlegenheitsschranke von deutlich unter 20% angenommen werden. Bezüglich Punkt (II) wäre zum Zeitpunkt der Studienplanung demnach zu argumentieren, dass in der Biomarker-negativen Gruppe die Ereignisrate so niedrig ist, dass unabhängig von der Größe des Therapieeffektes eine Behandlung weder im Hinblick auf die schwierig zu quantifizierenden Risiken einer medikamentösen Behandlung noch aus wirtschaftlicher Sicht zu rechtfertigen ist. Wie im vorausgegangenem Abschnitt ausgeführt, ist für hs-CRP-negative Patienten

in der AFCAPS/TexCAPS-Studie lediglich eine Verbesserung um rund 12 relative Prozent zu beobachten.

Welche Stoppregeln können zur Anwendung kommen, sodass eine unnötige Behandlung von hs-CRP-negativen Patienten vermieden werden kann?

- a) Formaler Nachweis der Irrelevanz des Therapieeffektes: Prinzipiell ist es möglich, eine Studie so zu planen, dass formal nachgewiesen wird, dass der Behandlungseffekt kleiner als die vorgegebene Irrelevanzgrenze ist. Eine mögliche Stoppregel zum Zeitpunkt der Zwischenauswertung bestände darin, auf Basis des Effektschätzers einen relevanten Effekt in der Biomarker-negativen Gruppe auszuschließen. Für die JUPITER-Studie hätte man generell andere Präventionsstudien, wie z. B. WOSCOP, heranziehen können. In dieser wurde eine 20%ige Verbesserung gezeigt. Ein Punktschätzer von 10 oder 15% könnte, wie oben ausgeführt, als Grundlage der Planung dienen. Die Fallzahl wäre so einzurichten, dass mit hoher Wahrscheinlichkeit die obere Grenze des Konfidenzintervalls unterhalb der relativen Verbesserung um 20% liegen würde. In der AFCAPS/TexCAPS-Studie betrug die Ereignisrate der CRP-negativen Patienten mit Cholesterinwerten unter dem Median 2,5% (vgl. [Abb. 2](#), [13]). Schon bei einer Unterlegenheitsschranke von 20%, wären 15.306 CRP-negative Patienten pro Studienarm notwendig gewesen, um bei einem einseitigen Fehler 1. Art von 2,5%, formale Nicht-Unterlegenheit nachzuweisen. Ein solcher formaler Nachweis ist dementsprechend nicht praktisch umsetzbar.
- b) Ein formaler Nachweis, dass der Effekt bei den Biomarker-negativen Patienten irrelevant ist, ist prinzipiell nicht notwendig. Es reicht der begründete Verdacht, dass die Therapie bei den Biomarker-negativen Patienten nicht wirkt. Man könnte die Power so einstellen, dass der Schätzer mit hoher Wahrscheinlichkeit kleiner als diese Schranke ist. Wie bereits im letzten Abschnitt diskutiert, würde

die Nichtunterlegenheitsschranke für die JUPITER-Studie höchstens 20%, eher 15% oder 10% betragen müssen. Eine Abschätzung der Power für gegebene Fallzahlen zu diesen 3 Nichtunterlegenheitsschranken findet sich in **Tab. 1**.

- c) Es könnte zu einer Zwischenauswertung in der Studie untersucht werden, ob genug Evidenz dafür vorhanden ist, dass unterschiedliche Therapieeffekte in den beiden Strata CRP-positive und CRP-negative Patienten vorliegen. Um zu entscheiden, ob die Therapieeffekte heterogen sind, könnte beispielsweise der p-Wert der Cochran's Q-Statistik betrachtet werden [14]. Für 2 Strata entspricht der Cochran's Q-Test der Differenz der beiden Therapieeffekte. Der zugehörige p-Wert ist somit mit dem p-Wert der Interaktion identisch. Liegt dieser p-Wert unter einem bestimmten Grenzwert, so kann die Rekrutierung im Stratum mit dem geringeren Effekt, also im CRP-negativen Studienarm, abgebrochen werden.

— Unter Berücksichtigung dieser Abbruchregel, kann die Fallzahl für die Zwischenauswertung (und damit den möglichen Abbruchzeitpunkt des CRP-negativen Arms) wie folgt bestimmt werden: Basis sind alle Patienten der AFCAPS/TextCAPS-Studie, die ein Cholesterinratio kleiner dem Median haben (vgl.

Abb. 2). Bei den CRP-positiven Patienten wurde der primäre Endpunkt im Lovastatin-Arm bei 2,5% (17/650) und im Placeboarm bei 5% (35/696) der Patienten beobachtet. Bei den CRP-negativen Patienten wurde der primäre Endpunkt im Lovastatin-Arm bei 2,4% (19/762) und im Placebo-Arm bei 5% (20/763) der Patienten beobachtet. Der entsprechende p-Wert der Cochran's Q-Statistik ist 0,14. Somit lag in diesem Fall schon ein deutlicher Hinweis auf einen heterogenen Therapieeffekt vor.

- Je nachdem wie die Schranke für den p-Wert für den Test auf Heterogenität vorab festgelegt wird,

d. h., wie groß die Evidenz für heterogene Therapieeffekte in den beiden CRP-Strata sein soll, kann die Fallzahl für eine Zwischenauswertung festgelegt werden. Mit beispielsweise 1250 Patienten pro Stratum und Therapiearm, wäre der p-Wert 0,05. Da mit dem Cochran's Q-Test auf Heterogenität und nicht auf Homogenität getestet wird, ist prinzipiell auch eine größere p-Wert-Schranke denkbar. So ist beispielsweise ein p-Wert von 0,2 schon ein starkes Signal für unterschiedliche Therapieeffekte. Eine Fallzahl von ungefähr 550 Patienten pro Stratum und Therapiearm hätte also bei dieser Abbruchregel in der JUPITER-Studie im CRP-negativen Arm ausgereicht. Wird eine substanziale Interaktion identifiziert, muss zusätzlich noch gezeigt werden, dass der Therapieeffekt vernachlässigbar ist. Die Argumentation hierfür könnte beispielsweise anhand des Punktschätzers in den Biomarker-negativen Patienten (Regel a) geführt werden, oder es könnte damit gerechtfertigt werden, dass Biomarker-negative Patienten ein höheres Risiko (z. B. mehr Nebenwirkungen) haben.

Wie viele Patienten mit hs-CRP <2,0 mg/l wären möglich gewesen?

Anhand des Screening-Logs, das im Supplement der JUPITER-Studie [10] dargestellt ist, kann bestimmt werden, wie viele Patienten mit einem hs-CRP <2,0 mg/dl möglich gewesen wären. Insgesamt wären 25.993 Patienten mehr in die Placebo-Run-In-Phase gegangen. Unter der Annahme, dass der Anteil der Patienten mit guter Compliance bei den hs-CRP-negativen genauso groß ist wie bei den hs-CRP-positiven, also 7,9%, hätten 23.939 hs-CRP-negative Patienten randomisiert werden können. Für alle im vorherigen Abschnitt beschriebenen Stoppregeln wären demnach genug hs-CRP-negative Patienten in der Screeningphase identifiziert worden.

Nachweis des Nutzens

Neben dem Kriterium, dass anhand des Biomarkers eine bestimmte Patientengruppe identifiziert werden kann, die von dem untersuchten Arzneimittel besonders profitiert, sollte der prognostische Wert des Biomarkers mit den prognostischen Werten klinischer Parameter verglichen werden.

So ist z. B. in der JUPITER-Studie nicht klar, ob die Hinzunahme des CRPs zu den klassischen kardiovaskulären Risikofaktoren eine Verbesserung der Auswahl der Patienten, die ohne eine klassische Erhöhung des Cholesterinspiegels eine Statintherapie erhalten sollten, bewirken würde. Ohne einen substanziellen Zugewinn würde sich prinzipiell nicht rechtfertigen, dass der hs-CRP-Test, der Geld kostet, der Bestimmung eines kardiovaskulären Risikoscores (z. B. Framingham Risk Score oder SCORE) vorgezogen werden sollte.

Generell existieren an dieser Stelle kaum Vorgaben. Die FDA gibt in ihrer Draft Guidance „Qualification Process for Drug Development Tools“ verschiedene Kriterien vor [15], legt aber keine methodischen Aspekte fest. Die EU hat im Moment diesbezüglich keine allgemeinen Vorgaben, hat aber im Beispiel der JUPITER-Studie retrospektive Analysen zum Framingham Risk Score und zum SCORE gefordert [16]. Die Ergebnisse deuten darauf hin, dass das hs-CRP prinzipiell mehr Risikopatienten identifiziert als die beiden etablierten kardiovaskulären Risikoscores. Für einen direkten Vergleich der Prognosewerte der Scores und des hs-CRPs wären jedoch auch an dieser Stelle CRP-negative Patienten notwendig gewesen.

Interessanterweise wurden die Ergebnisse der JUPITER-Studie von den amerikanischen, den kanadischen und den europäischen Zulassungsbehörden in sehr unterschiedliche Label übersetzt [17]. Die Tatsache, dass im Rahmen der JUPITER-Studie keine Informationen zu den hs-CRP-negativen Patienten erhoben wurden, hat vermutlich zu einem großen Interpretationsspielraum für die Zulassungsbehörden geführt.

Diskussion

In dieser Übersichtsarbeit wurden die wichtigsten Studiendesigns, die bei der Validierung eines Biomarkers im Rahmen von Arzneimittelstudien im Moment verwendet werden, dargestellt. Es wird deutlich, dass nur bei der Verwendung des Gold-Standard-Designs, in dem sowohl Biomarker-positive als auch Biomarker-negative Patienten randomisiert werden, eine umfängliche Nutzen-Risiko-Bewertung möglich ist. Auch wenn, wie im Beispiel der JUPITER-Studie, für die Biomarker-negativen Patienten ein geringer Nutzen zu erwarten war, sollte zumindest ein kleiner Teil der Biomarker-negativen Patienten in die Studie eingeschlossen werden. Nur so kann im Rahmen einer einzigen klinischen Prüfung Evidenz dafür gesammelt werden, dass es statthaft ist, einer Subpopulation das Testarzneimittel auf der Basis eines Biomarker-basierten Tests vorzuenthalten. Wir haben zu diesem Zweck mögliche Entscheidungskriterien vorgeschlagen. Die Charakteristika dieser und möglicher weiterer Regeln müssen in jedem Fall noch weiter untersucht werden. Inwieweit im Rahmen des Entwicklungsprogramms eines Arzneimittels und eines zugehörigen Biomarkers bereits in früheren Phasen überzeugend belegt werden kann, dass Biomarker-negative Patienten keinen Nutzen von einer Behandlung mit dem Testarzneimittel haben, bleibt zu diskutieren.

Mit hs-CRP wurde in der JUPITER-Studie ein umstrittener Biomarker untersucht. Es gab große epidemiologische Studien, die das CRP zur Risikostratifizierung infrage stellen. In der Studie von Woloshin und Schwartz mit 9000 amerikanischen Männern und Frauen wurde z. B. beobachtet, dass die Hälfte der Personen, die älter als 20 Jahre waren, ein CRP von mindestens 2 mg/l aufwiesen [18]. Kritiker behaupten ferner, dass das CRP kein unabhängiger Risikofaktor für die Entwicklung kardiovaskulärer Erkrankungen ist, sondern einfach nur stark mit den klassischen Risikofaktoren korreliert. Auf der anderen Seite gibt es auch große Studien, die die Hypothese, dass CRP ein unabhängiger Risikofaktor ist, stützen (z. B. [19]). Da in der JUPITER-Studie zwar Personen

mit erhöhten LDL-Werten ausgeschlossen waren, aber andere kardiovaskuläre Risikofaktoren nicht berücksichtigt wurden, ist die Mehrzahl der Teilnehmer übergeordnet, mehr als die Hälfte hat Bluthochdruck und 12% haben eine positive Familienanamnese für KHK. Es kann also nicht davon ausgegangen werden, dass die untersuchte Studienpopulation kein Basisrisiko für kardiovaskuläre Erkrankungen aufwies, das auch auf anderem Wege (z. B. mit dem Framingham-Score) hätte entdeckt werden können.

Ein weiterer Ansatzpunkt, den Rogers und Shlipak in einem Kommentar zur JUPITER-Studie, diskutiert haben, besteht darin, CRP nicht zur Risikostratifizierung, sondern zur Titrierung der Statinmedikation zu verwenden [20]. In Abhängigkeit vom hs-CRP-Wert könnte jeder Patient seine individuelle Therapie erhalten. Im Studienkontext könnte dann jedoch nicht mehr die Wirksamkeit einer bestimmten Therapie nachgewiesen werden, sondern nur noch der Algorithmus zur Therapiesteuerung (Biomarker-gesteuerte Therapie vs. Standardtherapie). Anhand dieses Designs könnte sogar eine Kombination verschiedener Arzneimittel untersucht werden. Ein großes Problem ergibt sich dann jedoch für die Sicherheitsbewertung der Medikamente, denn vermutlich werden die Fallzahlen in einzelnen Medikamentenkombinationsgruppen zu klein, um diesbezüglich verlässliche Aussagen zu gewinnen. Da bislang noch keine Konzepte existieren, stößt man an dieser Stelle an eine große Herausforderung für die Erarbeitung von tragfähigen Regeln für die „Zulassung“ von personalisierten Arzneimitteln im strengen Sinne.

In der Beurteilung der Studienergebnisse müssen 2 Elemente unterschieden werden: die Auswahl eines Biomarkers, der eine sinnvolle Unterteilung der Patientenpopulation zulässt (und eine nachvollziehbare Datenbasis, die diese Auswahl begründet), und der Wirksamkeitsnachweis in der (jeweils) ausgewählten Patientenpopulation. Für Letzteres ist ein statistisch abgesicherter confirmatorischer Behandlungsvergleich erforderlich. Für die Vorgehensweise zur sinnvollen Unterteilung sind unterschiedliche Strategien möglich. Ein confirmatorischer Nachweis eines Unterschiedes

zwischen den Teilpopulationen ist keine sinnvolle Forderung, da sich die Therapieeffekte in 2 Subpopulationen „signifikant“ unterscheiden können, aber dennoch beide Therapieeffekte (d. h. in der Biomarker-positiven wie auch in der Biomarker-negativen Gruppe) eine Behandlung rechtfertigen können. Ebenso kann es keine sinnhafte Forderung sein, dass confirmatorisch nachgewiesen wird, dass der Therapieeffekt in der Biomarker-negativen Gruppe „irrelevant“ ist (d. h. ein als relevant angesehener Therapieeffekt formal ausgeschlossen werden kann).

Unterschiedliche Designs weisen je nach Modellannahmen eine unterschiedliche Effizienz auf. Unterschieden werden muss dabei wiederum die effiziente Separation von dem effizienten Wirksamkeitsnachweis. Während eine Biomarkerauswahl und Subgruppeneinteilung unter Umständen bereits in früheren Phasen erfolgen kann, können in der abschließenden Phase der Arzneimittelentwicklung z. B. adaptive Designs relativ effizient bezogen auf die erforderliche Fallzahl für den finalen Wirksamkeitsnachweis sein, wenn es zusätzlich darum geht, den Nutzen des Biomarkers für die Auswahl der Patientenpopulation zu belegen. Die Charakteristika solcher Designs einerseits, die Validität andererseits ihre Effizienz betreffend müssen u. U. im Einzelfall genauer untersucht werden.

In der Regel beschränkt sich somit die Individualisierung der Therapie auf die Auswahl einer Teilpopulation. Für vollständig personalisierte Arzneimittel, also Arzneimittel, welche sich von Patient zu Patient unterscheiden (im einfachsten Fall in der Dosis oder dem Dosisschema), kann die Therapie nur zusammen mit der Auswahl des personalisierten Arzneimittels bewertet werden, d. h., es würde die Wirksamkeit der Strategie bewertet, die bei einem einzelnen Patienten zu einer bestimmten Therapie führt. Dadurch würde sich die Anwendung des Konzepts jedoch auf Therapien und Therapiebausteine einschränken, von deren prinzipieller Unbedenklichkeit ausgegangen werden kann (auch dann, wenn sie nach unter Umständen komplexen Regeln kombiniert werden).

In der Summe kann konstatiert werden, dass die Entwicklung geeigneter Bio-

marker für die Therapieauswahl und die Therapiesteuerung eines der wichtigsten Zukunftsthemen im Bereich der Arzneimittelentwicklung darstellt, weil wesentliche Verbesserungen der Wirksamkeit und Sicherheit der Anwendung von Arzneimitteln für den einzelnen Patienten und damit auch für die Gesamtheit der Patienten zu erwarten stehen. Zulassungsbehörden und Industrie müssen gemeinsam Strategien dazu entwickeln, wie dieses Konzept effizient, aber auch nachvollziehbar geprüft in die klinische Praxis eingeführt werden kann.

Vermutlich wird noch auf lange Sicht eine intensive fallweise Diskussion darüber erforderlich sein, was für die Qualifikation eines Biomarkers und die im direkten Bezug stehende Entwicklung eines Arzneimittels oder einer Kombination von Arzneimitteln untersucht und nachgewiesen werden muss. Im Rahmen der wissenschaftlichen Beratungsgespräche bei der EMA sind hierfür Prozeduren etabliert, wie Vertreter der Zulassungsbehörden und Antragsteller, die durchaus auch aus dem akademischen Umfeld kommen können, diesen Prozess vorbereiten können (http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/general/general_content_000349.jsp).

Korrespondenzadresse

Prof. Dr. A. Koch

Institut für Biometrie, Medizinische Hochschule Hannover
Carl-Neuberg-Str. 1, 30625 Hannover
koch.armin@mh-hannover.de

Einhaltung ethischer Richtlinien

Interessenkonflikt. A. Großhennig, N. Benda und A. Koch geben an, dass kein Interessenkonflikt besteht.

Dieser Beitrag beinhaltet keine Studien an Menschen oder Tieren.

Open Access

This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Literatur

- Hampel H, Frank R, Broich K et al (2010) Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. *Nat Rev Drug Discov* 9(7):560–574
- Scagliotti GV, Parikh P, von Pawel J et al (2008) Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* 26(21):3543–3551
- Viani GA, Afonso SL, Stefano EJ et al (2007) Adjuvant trastuzumab in the treatment of her-2-positive early breast cancer: a meta-analysis of published randomized trials. *BMC Cancer* 7:153
- Keedy VL, Temin S, Somerfield MR et al (2011) American Society of Clinical Oncology provisional clinical opinion: epidermal growth factor receptor (EGFR) mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. *J Clin Oncol* 29(15):2121–2127
- Karapetis CS, Khambata-Ford S, Jonker DJ et al (2008) K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* 359(17):1757–1765
- Food and Drug Administration (2011) In vitro companion diagnostic devices. <http://www.fda.gov/downloads/MedicalDevices/Device> (Zugegriffen: 17.05.2013)
- Buysse M, Michiels S, Sargent DJ et al (2011) Integrating biomarkers in clinical trials. *Expert Rev Mol Diagn* 11(2):171–182
- Ziegler A, Koch A, Krockenberger K, Grosshennig A (2012) Personalized medicine using DNA biomarkers: a review. *Hum Genet* 131(10):1627–1638
- Brannath W, Zuber E, Branson M et al (2009) Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Stat Med* 28(10):1445–1463
- Ridker PM, Danielson E, Fonseca FA et al (2008) Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med* 359(21):2195–2207
- Shepherd J, Cobbe SM, Ford I et al (1995) Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *N Engl J Med* 333(20):1301–1307
- Shepherd J, Blauw GJ, Murphy MB et al (2002) Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. *Lancet* 360(9346):1623–1630
- Downs JR, Clearfield M, Weis S et al (1998) Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. *JAMA* 279(20):1615–1622
- Cochran WG (1954) The combination of estimates from different experiments. *Biometrics* 10:101–129
- Food and Drug Administration (2010) Draft guidance: qualification process for drug development tools. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf> (Zugegriffen: 21.05.2013)
- Koenig W, Ridker PM (2011) Rosuvastatin for primary prevention in patients with European systematic coronary risk evaluation risk $\geq 5\%$ or Framingham risk $> 20\%$: post hoc analyses of the JUPITER trial requested by European health authorities. *Eur Heart J* 32(1):75–83
- Ridker PM (2011) A tale of three labels: translating the JUPITER trial data into regulatory claims. *Clin Trials* 8(4):417–422
- Woloshin S, Schwartz LM (2005) Distribution of C-reactive protein values in the United States. *N Engl J Med* 352(15):1611–1613
- Kaptoge S, Di Angelantonio E, Lowe G et al (2010) C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet* 375(9709):132–140
- Rogers AM, Shlipak MG (2009) C-reactive protein, statins, and cardiovascular risk: what can {JUPITER} teach us? *Am J Kidney Dis* 53(5):737–740