

# Individualisierte Medizin – die Grenzen des Machbaren Implikationen für Planung und Durchführung klinischer Studien

PD DR. STEFAN LANGE / PROFESSOR DR. JÜRGEN WINDELER / STELLVERTRETENDER LEITER UND LEITER DES INSTITUTS FÜR QUALITÄT UND WIRTSCHAFTLICHKEIT IM GESUNDHEITSWESEN (IQWiG), KÖLN



*Im gegenwärtigen Fokus der „individualisierten Medizin“ steht der keineswegs neue Ansatz, die Therapieauswahl anhand von Patientencharakteristika vorzunehmen. Neu ist allenfalls, dass sich diese Auswahl vornehmlich an molekularbiologischen Markern aus den Bereichen der Genomik, Proteomik, Metabolomik, Transkriptomik etc. oder an moderner Bildgebung orientieren soll. Die aktuelle Diskussion bietet die Chance, dass sich die ebenfalls nicht neue Erkenntnis, die Wertigkeit von diagnostischen Markern und therapeutischen Konsequenzen als Einheit zu betrachten, breiter durchsetzt. Dabei lässt sich sehr einfach verdeutlichen, dass sich der wesentliche Nutzen einer Information allein aus der Wechselwirkung zwischen dieser Information und dem Effekt von Handlungsalternativen speist. Unmittelbar handlungsleitend ist eine qualitative Wechselwirkung, die vorliegt, wenn sich für Patienten mit einem bestimmten Merkmal eine klare Überlegenheit der einen Handlungsalternative über die andere ergibt. Um es sehr klar zu sagen: Individualisierte und Evidenz-basierte Medizin sind keine Gegensätze!*

## Einleitung

Der Begriff „individualisierte Medizin“ (oder auch „personalisierte Medizin“) steht für eine Reihe von Ansätzen medizinischer Interventionen: Von der im wahrsten Sinne Wortes individualisierten Anfertigung von Medizinprodukten (z.B. Prothesen) über die Therapiesteuerung mittels Biomarker (z.B. eine am INR-Wert ausgerichtete Antikoagulation) bis hin zu der Therapieauswahl anhand von Patientencharakteristika. Letzteres wird auch als „stratifizierende Medizin“ oder „Präzisionsmedizin“ bezeichnet.

Eine Therapie anhand von Patientencharakteristika auszuwählen, ist keineswegs neu und reicht bis in die Antike zurück [Bondio MG, Michl S. 2010]. Und auch die Homöopathie hat seit je für sich reklamiert, die Therapieauswahl individualisiert anhand der Symptome und klinischen Erscheinungen der Patienten vorzunehmen: „The homeopathic case history aims to ascertain the totality of signs and symptoms of each patient, enabling the selection of an individualized homeopathic medicine“ [Adler UC et al. 2011]. Neu ist allenfalls, dass sich diese Auswahl vornehmlich an molekularbiologischen Markern aus den Bereichen der Genomik, Proteomik, Metabolomik, Transkriptomik, etc. orientiert bzw. orientieren soll. Auch moderne Bildgebung (z.B. mittels PET/CT) soll zunehmend diesem Zweck dienen. Der vorliegende Beitrag widmet sich ausschließlich solchen Ansätzen einer stratifizierenden Medizin. Das heißt nicht, dass nicht auch die anderen Ansätze einer jeweiligen Evaluation bedürfen. Aber die aktuelle Diskussion dreht sich im Wesentlichen um die stratifizierende Medizin als Ausdruck einer individualisierten Medizin, im Besonderen im Bereich der Onkologie.

Die Versprechungen, die mit der Individualisierten Medizin einhergehen, sind größtenteils enorm und erreichen mitunter Ausmaße einer Hybris. Proportional zu diesen Versprechungen gehen Befürchtungen über die Schwierigkeiten, den Nutzen entsprechender Methoden nachzuweisen. So ließ sich der Präsident der amerikanischen Krebsgesellschaft des Jahres 2011 vernehmen: „We can no longer think of cancer as one disease. Even something like lung cancer could be hundreds of distinct cancers, each defined by specific molecular characteristics requiring different treatment approaches. This makes research more challenging, but the payoff for patients will be enormous“ [Kris MG, Meropol NJ, Winer EP 2011].

In einer Broschüre der Koalition für personalisierte Medizin aus den USA heißt es gar: „Patients with melanoma, metastatic lung, breast, or brain cancers, and leukemia are now being routinely offered a ‘molecular diagnosis’ in some clinical centers, allowing their physicians to select tailored treatments that can greatly improve their chances of survival“ [Personalized Medicine Coalition. The Case for Personalized Medicine].

Individualisierte und Evidenz-basierte Medizin sind keine Gegensätze, wie manchmal kolportiert wird, getreu dem Motto: Evidenz-basierte Medizin sei etwas für die Massenversorgung, individualisierte Medizin dagegen etwas auf den Einzelnen passgenau zugeschnittenes (sozusagen ein „Premiumprodukt“, für das dann konsequenter Weise auch „Premiumpreise“ verlangt werden können – doch das ist eine andere Geschichte). Die immer noch vergleichsweise wenigen, real existierenden Beispiele für klinisch etablierte oder gar behördlich zugelassene bzw. vorgeschriebene molekularbiologische Marker für die The-

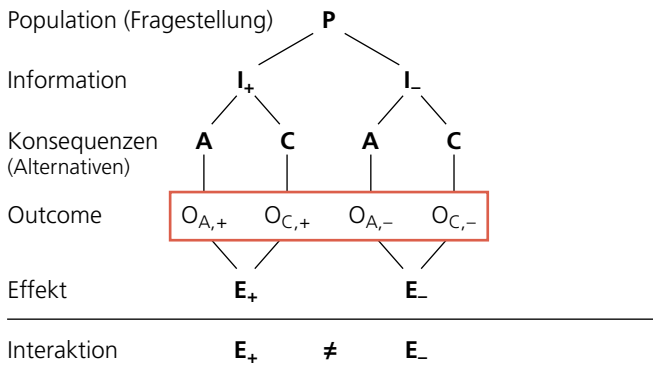
rapieauswahl in der Onkologie belegen das Gegenteil: Erstens wurden diese nur auf Basis von aufwändigen klinischen Studien identifiziert und etabliert, und zweitens werden damit weiterhin Gruppen von Patienten charakterisiert, so wie zuvor und immer noch Gruppen von Patienten anhand von klinischen Merkmalen (z.B. Symptomen) charakterisiert werden.

Allerdings birgt die gegenwärtige Aufmerksamkeit für eine individualisierte Medizin doch eine Chance für etwas Neues: Dass sich nämlich die keineswegs neue Erkenntnis, die Wertigkeit von diagnostischen Markern und therapeutischen oder auch präventiven Ansätzen als Einheit zu betrachten, breiter durchsetzt. Denjenigen, die sich vor Jahrzehnten mit der Evaluierung diagnostischer Verfahren befasst haben, war bereits damals klar, dass Informationsgewinn allein keinen Wert an sich hat, sondern dieser sich in aller erster Linie an den möglichen Konsequenzen bemisst, die aus dieser Information resultieren.

### **Der Wert von Information**

Was ist es wert, zu wissen, innerhalb der nächsten 12 Monate mit einer Wahrscheinlichkeit von etwa 80% zu sterben? Eine solche Information kann beispielsweise ein Biomarker bei einer bestimmten Krebserkrankung liefern. Nun, das mag einen Wert haben, wenn es nachfolgend unterschiedliche Optionen des Handelns gibt und der Erfolg dieser Handlungsoptionen vom Risiko (der Prognose) abhängt. Dieser Zusammenhang von (diagnostischer oder prognostischer) Information und „Erfolg“ einer Intervention – in der Evidenz-basierten Welt sollte eher von einem „Effekt“ gesprochen werden – wird statistisch-technisch als Wechselwirkung oder Interaktion bezeichnet.

### Der Wert von Information

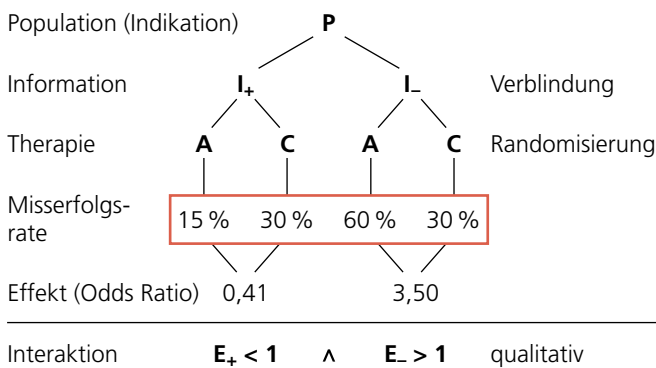


Quelle: Stefan Lange, Jürgen Windeler



**Abbildung 1:** Bei einer Interaktion handelt es sich um den Zusammenhang zwischen einer diagnostischen oder prognostischen Information und dem Effekt einer Intervention.

### Qualitative Wechselwirkung mit gegenläufigen Effekten



Quelle: Stefan Lange, Jürgen Windeler



**Abbildung 2:** Diese Situation einer qualitativen Wechselwirkung ist selten, da in Studien aufgrund der Vorannahmen oft der Vergleich von A und C auf Patienten in nur einem Stratum beschränkt wird.

In Abbildung 1 wird dies für den einfachen Fall einer binären Information (positiv/negativ) und dem Vorliegen zweier (möglicher) Konsequenzen (Handlungsoptionen) C (für Kontrolle) und A (für Alternative) schematisch dargestellt. Um das Bild komplett zu machen, bedarf es noch der Festlegung eines Kriteriums (Outcome) anhand dessen der Erfolg, der Effekt (E) gemessen wird. Wenn der Effekt in der Gruppe I<sub>+</sub> ein anderer als in der Gruppe I<sub>-</sub> ist, dann liegt eine Wechselwirkung vor.

Und aus dieser Abbildung wird auch sofort deutlich, was für ein Studiendesign erforderlich ist, um den Nutzen eines Biomarkers zu evaluieren: Idealerweise handelt es sich um einen randomisierten Vergleich der beiden Alternativen A und C, denn nur in einem randomisierten Vergleich lässt sich (theoretisch) verzerrungsfrei ein Effekt schätzen. Und in einer solchen Studie wird gleichzeitig die Information I erhoben, die möglichst für die Dauer der Studie verblindet bleibt. Durch die möglichen Ausprägungen von I ergeben sich im einfachen binären Fall zwei Strata, für die die Effekte in I<sub>+</sub> und I<sub>-</sub> getrennt geschätzt werden. Daher rührt die alternative Bezeichnung stratifizierende Medizin.

### Wechselwirkung

Abbildung 2 kennzeichnet die ideale Situation einer qualitativen Wechselwirkung: Für Patienten in Stratum I<sub>+</sub> ergibt sich eine klare Überlegenheit der Therapie A gegenüber Therapie C (gemessen an dem Kriterium „Misserfolgsrate“ und operationalisiert durch das Effektmaß Odds Ratio), während es in Stratum I<sub>-</sub> genau umgekehrt ist (Überlegenheit von C gegenüber A). Eine solche Situation findet sich selten, da mehr oder weniger klare pathophysiologische Vorstellungen häufig dazu (ver-)führen, von vornherein in Studien den Vergleich von A und C auf Patienten einzuschränken, die ein gewisses Merkmal (z.B. das Merkmal I<sub>+</sub>) aufweisen (so genanntes „Anreicherungsdesign“).

Dennoch gibt es ein recht aktuelles Beispiel, nämlich den Vergleich des Tyrosinkinashemmers Gefitinib gegenüber der Kombinationschemotherapie aus Carboplatin und Paclitaxel bei Patienten mit einem fortgeschrittenen nichtkleinzelligen Lungenkarzinom. Bei Vorliegen einer EGFR-Mutation im Tumorgewebe zeigte sich in einer Studie ein klarer Vorteil für Gefitinib in Bezug auf das progressionsfreie Überleben (Hazard Ratio 0,48 [95%-Konfidenzintervall: 0,36-0,64]). Wenn eine solche Mutation nicht nachgewiesen werden konnte, ergab sich ein klarer Nachteil (Hazard Ratio 2,85 [2,05-3,98]) [Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpawaravong P, Han B, Margono B, Ichinose Y, Nishiwaki Y, Ohe Y, Yang JJ, Chewaskulyong B, Jiang H, Duffield EL, Watkins CL, Armour AA, Fukuoka M. 2009].

Eine qualitative Wechselwirkung kann auch dann angenommen werden, wenn ein (beispielsweise) positiver Effekt in dem einen Stratum nachgewiesen werden kann, nicht aber im anderen. Abbildung 3 zeigt eine derartige Konstellation, bei der sich im Stratum I kein Unterschied

im Outcome zwischen den beiden Alternativen zeigen lässt. Klassisches Beispiel für eine derartige Interaktion ist die Gabe von selektiven Östrogenrezeptormodulatoren (z.B. Tamoxifen) nur bei Patientinnen mit Brustkrebs, bei deren Tumore Hormonrezeptoren nachgewiesen werden können. Bei Rezeptor-negativen Patientinnen entfalten solche Medikamente keine Wirkung [Early Breast Cancer Trialists' Collaborative Group (EBCTCG) 2005].

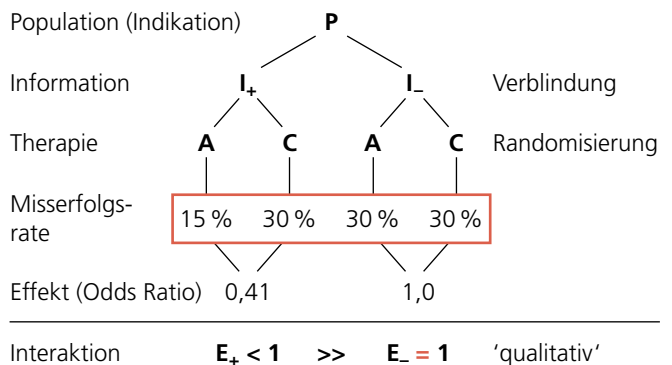
Problematisch wird die Situation, wenn zwischen Information und Therapieeffekt eine quantitative Wechselwirkung vorliegt (Abbildung 4). Hier würde die Wahl für Therapie A auf Basis der Information dazu führen, dass den Patienten im Stratum I. mit dem kleineren Therapieeffekt eine dennoch wirksame Therapie vorenthalten wird. Die Entscheidung für Therapie A muss sich dann also auf weitere Kriterien stützen. Sind es allein monetäre Erwägungen, wäre das schlicht ein Fall von Rationierung.

Bei Vorliegen einer qualitativen Wechselwirkung ist die aus der Bestimmung des Biomarkers resultierende Information unmittelbar handlungsleitend. In einem solchen Fall wird auch von einem prädiktiven Biomarker gesprochen, in Abgrenzung von einem Marker, der lediglich prognostische Information (z.B. etwa 80%-ige Wahrscheinlichkeit innerhalb eines Jahres zu sterben) liefert [Sargent DJ, Conley BA, Allegra C, Collette L. 2005]. Während es nur sehr wenige prädiktive Biomarker gibt, ist die Fülle an (Literatur zu) prognostischen Biomarkern unübersehbar [Ioannidis JP, Tzoulaki I. 2010]. Der Kontrast ist bedrückend: Nur prädiktive Biomarker haben einen Patienten-relevanten Nutzen, prognostische Biomarker letztlich nicht.

### Anreicherungsdesign

Das in den Abbildungen 1-4 verwendete Schema ermöglicht es, alternative Studiendesigns und insbesondere auch deren Schwächen aufzuzeigen. In Abbildung 5 ist das – bereits oben kurz angesprochene – Anreicherungsdesign dargestellt. Bei einem solchen Design wird eine qualitative Wechselwirkung – zumeist auf Basis von pathophysiologischen Vorstellungen – vorausgesetzt. Beispiele dafür sind die Studien, die zur Zulassung von Trastuzumab beim Her2/neu-positiven Mammakarzinom [Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, Swain SM, Pisansky TM, Fehrenbacher L, Kutteh LA, Vogel VG, Visscher DW, Yothers G, Jenkins RB, Brown AM, Dakhil SR, Mamounas EP, Lingle WL, Klein PM, Ingle JN, Wolmark N 2005] oder

### Qualitative Wechselwirkung mit Null-Effekt in einem Stratum

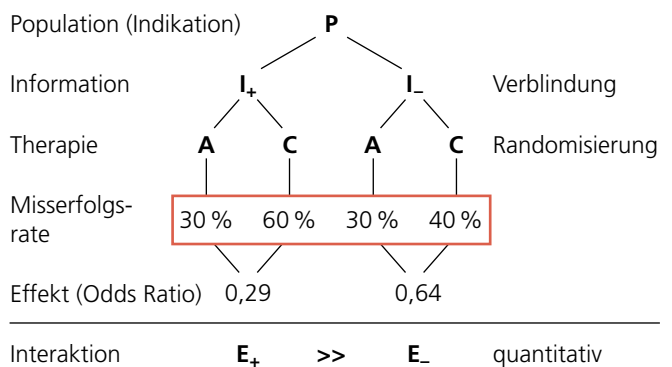


Quelle: Stefan Lange, Jürgen Windeler



**Abbildung 3:** In dieser Konstellation wird in einem Stratum ein positiver Effekt nachgewiesen, in dem anderen aber nicht.

### Quantitative Wechselwirkung



Quelle: Stefan Lange, Jürgen Windeler

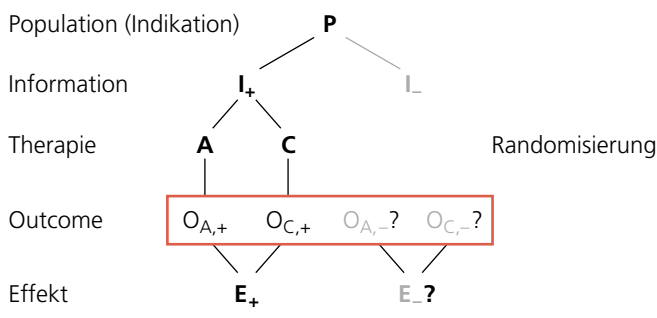


**Abbildung 4:** Die Wahl der Therapie A würde auf Basis der Information dazu führen, dass den Patienten im Stratum I. mit dem kleineren Therapieeffekt eine dennoch wirksame Therapie vorenthalten wird.

von Vemurafenib beim BRAF-positiven fortgeschrittenen malignen Melanom [Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D, Lorigan P, Lebbe C, Jouary T, Schadendorf D, Ribas A, O'Day SJ, Sosman JA, Kirkwood JM, Eggermont AM, Dreno B, Nolop K, Li J, Nelson B, Hou J, Lee RJ, Flaherty KT, McArthur GA 2011] geführt haben.

In beiden Situationen wurden ausschließlich solche Patientinnen und Patienten eingeschlossen, die das je-

### Anreicherungsdesign



Quelle: Stefan Lange, Jürgen Windeler



**Abbildung 5:** Im Fall eines Anreicherungsdesigns wird eine qualitative Wechselwirkung zumeist auf Basis von pathophysiologischen Vorstellungen vorausgesetzt.

weilige molekularbiologische Charakteristikum aufweisen, und in den Studien wurden unbestreitbar Patienten-relevante Vorteile beobachtet. Dennoch existiert ein Problem: Pathophysiologische Ideen können falsch oder zumindest unvollständig sein; und dann liefern auch darauf basierende Studien falsche oder zumindest unvollständige Informationen.

Im Fall von Trastuzumab wurde zu Beginn des Jahres 2012 eine Studie veröffentlicht, in die Brustkrebs-Patientinnen mit einem Her2/neu-negativen Primärtumor eingeschlossen und mit Trastuzumab therapiert wurden. Dabei handelt es sich quasi um das fehlende Stratum aus Abbildung 5. Und in der Tat wurde in der Studie ein statistisch signifikanter Vorteil von Trastuzumab gegenüber Placebo in Bezug auf das krankheitsfreie Überleben beobachtet [Georgoulas V, Bozionelou V, Agelaki S, Perraki M, Apostolaki S, Kallergi G, Kalbakis K, Xyrafas A, Mavroudis D 2012].

Als Erklärung für diesen eigentlich überraschenden Befund wurden im Blut zirkulierende Tumorzellen angeführt, die – anders als der Primärtumor – Her2/neu-positiv waren. Bereits zuvor hatten andere Autoren darauf hingewiesen, dass das Bild von Trastuzumab beim Mammakarzinom möglicherweise unvollständig ist bzw. war [Mandrekar SJ, Sargent DJ. 2009]. Neben der Intensivierung einer Behandlung können in einem Anreicherungsdesign auch de-eskalierende Strategien im Sinne einer Nicht-Unterlegenheitsfragestellung überprüft werden.

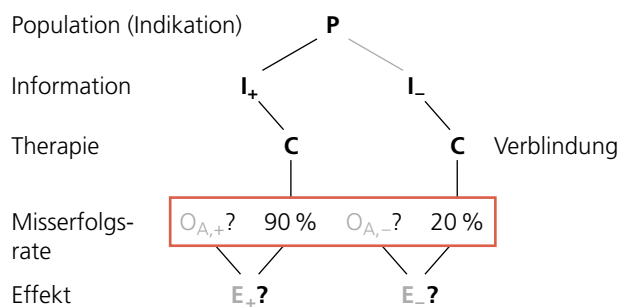
Die besondere Bezeichnung „Anreicherungsdesign“ ist im Grunde genommen ein wenig überspannt: Letztlich wird jede klinische Studie durch die Ein- und Ausschluss-

kriterien in diesem Sinne „angereichert“ durchgeführt. Und genau das wurde in der Vergangenheit und wird immer noch häufig beklagt, dass nämlich klinische Studien aufgrund ihrer zum Teil rigiden Selektionskriterien ein nur unvollständiges, häufig die (Alltags-)Realität nicht ausreichend widerspiegelndes Bild abgeben würden. Studien mit „Anreicherungsdesign“ können somit gleichzeitig zu einem doppelten Dilemma führen: Ist es ethisch und sozialrechtlich gerechtfertigt, Patienten, die die geforderten Ein- und Ausschlusskriterien der zugrunde liegenden Studien nicht erfüllen, die entsprechende Therapie – sofern sie sich als überlegen erwiesen hat – vorzuenthalten? Und ist es umgekehrt statthaft, solchen Patienten eine derartige Therapie anzubieten, weil Zweifel an der Übertragbarkeit bestehen?

Eine ernsthafte Auseinandersetzung mit diesem Dilemma hätte unter Umständen erhebliche Implikationen für unsere Studienkultur: Anstatt Studien immer effizienter zu planen und damit Fallzahlen zu minimieren, müssten vielmehr die Studien viel breiter und mit viel größeren Fallzahlen durchgeführt werden. Nur mit solch großen Fallzahlen und mit minimalen Ein- und Ausschlusskriterien versehenen Studien ließen sich die Wechselwirkungen zwischen relevanten Patientenmerkmalen und Therapieeffekt schätzen, die letztendlich handlungsleitend sind.

Einschränkend kann dazu bemerkt werden, dass sehr große Fallzahlen nur dann vonnöten sein werden, wenn auch die Therapieeffekte in den relevanten Strata klein oder mäßig sind. Wenn die eingangs zitierte Behauptung der Koalition für personalisierte Medizin aus den USA zuträfe, dass sich durch die (neuen) „zielgerichteten“ („in-

### Prognosestudie



Quelle: Stefan Lange, Jürgen Windeler



**Abbildung 6:** Bei der Prognose-Studie erfolgt kein Therapievergleich, sondern es wird lediglich das Outcome der Patienten je nach Ausgang des Biomarkers beobachtet.

dividualisierten“) Therapien die Überlebenschancen von Patienten mit fortgeschrittenem malignem Melanom, metastasiertem Lungenkrebs, Brustkrebs, oder bösartigen Hirntumoren außerordentlich („greatly“) verbessert hätten, dann wären auch die entsprechenden Studien mit außerordentlich kleinen Fallzahlen ausgekommen. Die zitierte Behauptung trifft allerdings nicht zu: Wenngleich z.B. durch Vemurafenib beim BRAF-positiven metastasierten malignen Melanom ein als erheblich zu bezeichnender Zugewinn an Lebenszeit beobachtet werden konnte, ändert dies (bis jetzt) nichts an der zumeist infausten Prognose der Grunderkrankung, im Wesentlichen deshalb, weil die Tumore eine Resistenz entwickeln.

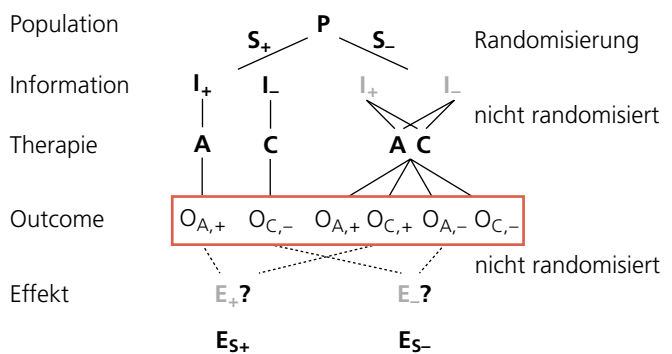
Eine besondere Variante des Anreicherungsdesign besteht darin, in beiden Strata unterschiedliche Therapieansätze miteinander zu vergleichen. Bei der HD18-Studie z.B., in die Patienten mit Hodgkin-Lymphom in einem fortgeschrittenen Stadium eingeschlossen werden, wird nach zwei Zyklen einer Chemotherapie ein Interims-Positronenmissionstomogramm (Interims-PET) durchgeführt. PET-positive Patienten, bei denen von einer ungünstigeren Prognose ausgegangen wird, werden in die Standard-Fortführung der Chemotherapie oder in eine zusätzliche Gabe von Rituximab – also eine Therapie-Eskalation – randomisiert (Ziel: Überlegenheitsnachweis), während bei PET-negativen Patienten eine Randomisierung zwischen der Standard-Fortführung der Chemotherapie und einer de-eskalierenden Strategie erfolgt (Ziel: Nicht-Unterlegenheitsnachweis) [Dietlein M, Eich HAT, Fuchs M, Borchmann P, Engert A, Kobe C. 2010].

Das Besondere an dieser Variante besteht in dem in beiden Strata für den Therapievergleich eingesetzten Standard-Arm, der eine Kontrolle der wesentlichen Grundannahme erlaubt, dass nämlich die PET-Untersuchung eine überaus überzeugende Prognose-Abschätzung erlaubt, die den jeweiligen Vergleichsarm in dem komplementären Stratum quasi verbietet. Dennoch bleibt vom Grundsatz her auch hier das Bild unvollständig.

### Prognose

Ein anderes unvollständiges Design ist die Prognose-Studie, bei der kein Therapievergleich erfolgt, sondern lediglich das Outcome der Patienten je nach Ausgang des Biomarkers beobachtet wird (Abbildung 6). In dem absichtlich extrem gewählten Zahlenbeispiel weist der Marker eine durchaus beachtliche prognostische „Güte“ auf: Das Odds Ratio für

### Strategie-Design



Quelle: Stefan Lange, Jürgen Windeler

**Abbildung 7: Beim Strategiedesign wird nicht zwischen zwei Therapien verglichen, sondern es wird die diagnostisch-therapeutische Kette als Ganzes geprüft.**

ein schlechtes Outcome beträgt für Marker-positive gegenüber Marker-negativen Patienten 36! Dennoch wäre der Marker nicht handlungsleitend, wenn bei einem putativen Therapievergleich in Stratum I<sub>+</sub> 80% der Patienten unter Therapie A ein schlechtes Outcome hätten und in Stratum I<sub>-</sub> 10% der Patienten. Denn dann existierte für das fiktive Zahlenbeispiel keine Wechselwirkung zwischen (Marker-basierter) Information und Therapieeffekt, und allen Patienten sollte Therapie A angeboten werden.

### Strategie-Design

Um das Bild komplett zu machen, sei noch das „Strategie-Design“ erwähnt, bei dem nicht zwischen zwei Therapien, sondern zwischen zwei Strategien verglichen wird (Abbildung 7): Auf der einen Seite eine Strategie, bei der auf Basis der Marker-basierten Information die Therapie ausgewählt wird, auf der anderen Seite das Standardvorgehen, bei dem der Marker erst gar nicht bestimmt oder das Ergebnis nicht bekanntgegeben und für die Therapieauswahl herangezogen wird. Im Extremfall erhalten bei diesem Design die Patienten im Kontrollarm nur eine, nämlich die Standardtherapie. Es ist aber durchaus möglich, dass bereits beide Therapiealternativen verfügbar sind, die Auswahl sich aber nach anderen Kriterien richtet.

Bei einem Strategie-Design wird die diagnostisch-therapeutische Kette als Ganzes geprüft, so dass die Wechselwirkung zwischen Information und Therapieeffekt nicht mehr geschätzt werden kann. Dies hat insbesondere bei einem negativen Ausgang einer solchen Studie den Nach-

teil, dass unklar bleibt, ob der Marker nicht ausreichend prädiktiv ist oder die eine Therapie gegenüber der anderen keinen Vorteil hat. Darüber hinaus ist ein derartiges Design in gewisser Weise weniger effizient, da sich – wie die Abbildung 7 verdeutlichen soll – in beiden Vergleichsarmen identisch behandelte Patienten mit gleichen Charakteristika befinden, der Kontrast also verwischt wird. Je weniger Patienten durch den Marker identifiziert werden sollen bzw. können, für die ggf. eine alternative Therapie infrage kommt, desto schwieriger wird es sein, Unterschiede zwischen den Strategien in einer solchen Studie aufzudecken.

Ein Vorteil des Strategie-Designs ist demgegenüber, dass hierbei über den reinen Informationsgewinn hinausgehende Effekte einer Test-Anwendung geprüft werden können, seien es direkt physische Komplikationsmöglichkeiten bei invasiven Testverfahren oder auch durch die Information getriggerte psychische (Neben-)Wirkungen [Bossuyt PM, McCaffery K. 2009].

Typisches Einsatzgebiet für das Strategie-Design sind die großen Screening- bzw. Früherkennungsuntersuchungen, z.B. zum Nutzen eines Mammographie- oder PSA-Screenings. Doch auch der Nutzen von diagnostischen Tests wird im Rahmen solcher Designs geprüft, mitunter selbst dann, wenn der Test nahezu perfekte Test-Gütekriterien aufweist und theoretisch nur noch das „Offensichtliche“ nachzuweisen ist. Beispielsweise wurde im Rahmen einer randomisierten Studie untersucht, ob ein Screening auf das Vorliegen eines bestimmten genetischen Merkmals (HLA-B\*5701) vor einer geplanten Therapie mit Abacavir bei Patienten mit einer HIV-Infektion das Auftreten einer Hypersensitivitätsreaktion (die mit Abacavir in Zusammenhang gebracht wurde) verhindern kann. In der Interventionsgruppe wurde bei Vorliegen von HLA-B\*5701 auf ein anderes Präparat gewechselt, während in der Kontrollgruppe für alle Patienten Abacavir vorgesehen war [Mallal S, Phillips E, Carosi G, Molina JM, Workman C, Tomazic J, Jägel-Guedes E, Rugina S, Kozyrev O, Cid JF, Hay P, Nolan D, Hughes S, Hughes A, Ryan S, Fitch N, Thorborn D, Benbow A 2008].

Die Studie endete – erwartungsgemäß? – positiv: Immunologisch bedingte Hypersensitivitätsreaktion traten in der Interventionsgruppe nicht auf, gegenüber einem Anteil von knapp 3 % in der Kontrollgruppe. Für solche immunologisch bedingten Hypersensitivitätsreaktionen hatte der eingesetzte Screening-Test eine Sensitivität von 100%, bei einer Spezifität von annähernd 97%.

## Fazit

Durch die sogenannte individualisierte Medizin ergeben sich bis auf Detailfragen keine besonderen Herausforderungen für die Planung und Durchführung klinischer Studien. Vielmehr müssen lediglich bereits seit langem bekannte Prinzipien der Einheit von Diagnostik und Therapie stärker in der klinischen Forschung verinnerlicht werden, auch wenn manche medizinische Fachgesellschaften hier noch einem gewissen Anachronismus verfallen zu sein scheinen. Die Deutsche Gesellschaft für Nuklearmedizin hat zum Beispiel ihre Position vertreten, dass die Verknüpfung von diagnostischer Methode und therapeutischem Ansatz zum Nachweis eines Patienten-relevanten Nutzens grundsätzlich abgelehnt werde [Kotzerke J, Dietlein M, Grünwald F, Bockisch A. 2010].

Dabei soll nicht verschwiegen werden, dass sich eine bestimmte Analyse-Methode, nämlich die Analyse von Subgruppen, aus der biometrischen „Schmuddel-Ecke“ befreien muss. Während es früher – etwas überspitzt formuliert – hieß: „Subgroups kill people“ [Van Gijn J. 1995], gilt heute eher: „Not doing subgroup analyses has very probably killed more people“ [Rothwell PM. 2005]. Letztlich ist individualisierte Medizin nichts anderes als „Subgruppen-Medizin“.

Im Übrigen bedarf die Prüfung stratifizierender Therapie-Entscheidungen nicht immer, zumindest nicht sofort der ausschließlich prospektiven Validierung. Beispielsweise wurde die Bedeutung des KRAS-Mutationsstatus für die Therapieauswahl beim (fortgeschrittenen) kolorektalen Karzinom auf Basis bereits durchgeführter randomisierter Therapiestudien entdeckt und verifiziert [Ginsburg GS, Kuderer NM 2012]. Voraussetzung dafür ist das Aufbewahren von Tumorproben möglichst aller Patienten aus klinischen Studien. Dies dürfte dann schon eine weitgehend logistische und juristische Herausforderung für die Zukunft sein, nämlich das Sammeln, Aufbewahren und Zur-Verfügung-Stellen solcher Proben.

Allerdings muss konstatiert werden, dass der statistische Nachweis von tatsächlich qualitativen Wechselwirkungen gegenwärtig nicht zum Standard-Repertoire biometrischer Analysen gehört. Allenfalls wird auf das Vorliegen irgendeiner (auch quantitativen) Wechselwirkung getestet, was im Grunde genommen recht uninteressant ist. Methodisch angemessener wäre es, eine mehr als irrelevante Wechselwirkung zu definieren und diese mit Hilfe eines

statistischen Hypothesentests (mit üblichem Irrtumsniveau) auszuschließen. In letzter Konsequenz führt das dazu, jegliche quantitative Wechselwirkung als irrelevant zu deklarieren, um dann zu einem Test auf qualitative Wechselwirkung zu gelangen [Gail M, Simon R. 1985]. Hier gibt es einen gewissen Nachholbedarf und auch einen Bedarf an Weiterentwicklungen, da entsprechende statistische Tests eine vergleichsweise geringe Power besitzen.

Eines ist auf jeden Fall klar: Individualisierte und Evidenzbasierte Medizin sind keine Gegensätze!

Literatur bei den Autoren

E-Mail-Kontakt: stefan.lange@iqwig.de und juergen.windeler@iqwig.de

#### PD DR. MED. STEFAN LANGE

*Dr. Lange studierte von 1982 bis 1989 Humanmedizin an der Universität Düsseldorf und promovierte dort 1994. Von 1989 bis 1993 war er, zunächst als Arzt im Praktikum, später als Assistenzarzt, an der Medizinischen Klinik des Ferdinand-Sauerbruch-Klinikums in Wuppertal tätig. 1993 wechselte er an die Abteilung für Medizinische Informatik, Biometrie und Epidemiologie der Ruhr-Universität Bochum. Im Jahr 2003 habilitierte er sich an der Ruhr-Universität und erhielt die Venia legendi für das Fachgebiet „Medizinische Biometrie und Klinische Epidemiologie“. Im November 2004 erfolgte der Wechsel an das IQWiG. Dort ist er seit 2005 stellvertretender Leiter.*



#### PROFESSOR DR. MED. JÜRGEN WINDELER

*Professor Windeler hat Humanmedizin an der Universität Göttingen und an der Medizinischen Hochschule Lünebeck studiert. Nach seiner Promotion arbeitete er ab 1988 als Assistenzarzt und Wissenschaftlicher Mitarbeiter an mehreren Universitäten, zuletzt war er – nach der Habilitation 1993 – stellvertretender Leiter der Abteilung für Medizinische Biometrie an der Universität Heidelberg. Zwischen 1999 bis 2004 war Windeler Leiter des Fachbereichs Evidenzbasierte Medizin beim Medizinischen Dienst des Spitzenverbandes Bund der Krankenkassen (MDS) in Essen, seit 2004 dort Stellvertretender Geschäftsführer und Leitender Arzt. Seit September 2010 leitet Windeler das IQWiG in Köln.*

