IQWiG Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Institute for Quality and Efficiency in Health Care

Herbst-Symposium 2014 ▪ 28. und 29. November 2014 ▪ IQWiG, Köln

# Evaluation von Diagnostik: Eine Evolution vor dem Hintergrund der „personalisierten“ Medizin am Beispiel des primären Mammakarzinoms

**Martina Markes, Alexander Mensch, Juliane Hardt, Fülöp Scheibler**
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

## Personalisierte Medizin - Stratifizierende Medizin

Der Begriff „Personalisierte Medizin“ wird von seinen Befürwortern mit einer Revolutionierung der Medizin durch eine zielgerichtete Behandlung gleichgesetzt, die den individuellen genetischen und molekularbiologischen Voraussetzungen der Patienten Rechnung tragen soll. Die Methode, eine Population (z. B. unter Nutzung von Biomarkern) in Subgruppen zu unterteilen, ist allerdings nicht neu. Daher wurde hierfür der neutrale Begriff der „stratifizierenden Medizin“ vorgeschlagen [1]. Die Neuerungen finden vielmehr in der Anwendung vielfältiger Designs randomisiert kontrollierter Studien (RCTs) statt, mit denen der Nutzen verschiedener Strategien der stratifizierenden Medizin (z. B. der Einsatz von Biomarkern) bewertet werden kann: dies soll hier am Beispiel der systemischen Therapie des primären Mammakarzinoms gezeigt werden.

## Prognostische und prädiktive Biomarker

Mit prognostischen Biomarkern kann der Verlauf einer Erkrankung abgeschätzt werden: beim Brustkrebs beeinflussen beispielsweise Tumorgröße und -ausdehnung, Tumorbefall von Lymphknoten, das Vorliegen von Fernmetastasen sowie das Alter der Patientin den Krankheitsverlauf nachweislich und gelten deshalb als prognostische Marker. Aber nur wenn die Prognose extrem günstig oder ungünstig ist, kann auf dieser Grundlage eine Therapieentscheidung getroffen werden.

Prädiktive Marker können hingegen Aussagen zum Ansprechen auf eine bestimmte Behandlung oder bezüglich des Auftretens unerwünschter Effekte treffen. Hierfür eignen sich besondere Eigenschaften des Tumors. Beispielsweise zeigt der Hormonrezeptorstatus, ob der Tumor hormonabhängig wächst und somit durch eine endokrine Therapie effektiv behandelt werden kann. Prädiktive Marker eignen sich für die Steuerung der weiteren Behandlung, also der Therapieauswahl.

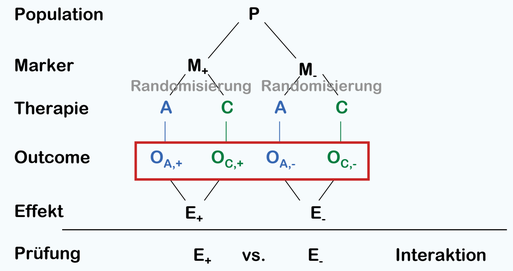
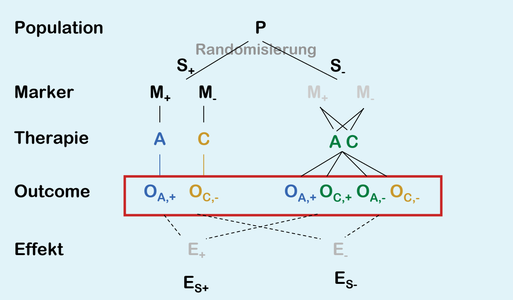
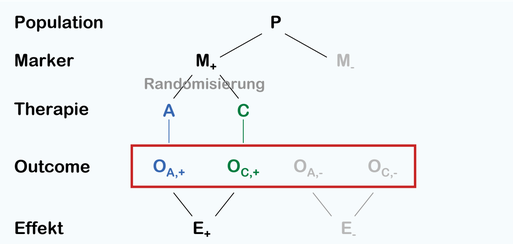
## Stratifizierende Medizin beim primären Mammakarzinom

Eine wichtige Frage in der Therapie von Brustkrebs-Patientinnen ist noch nicht endgültig beantwortet: Welche Patientinnen mit mittlerem Rezidivrisiko haben mit hoher Wahrscheinlichkeit einen Nutzen von der Chemotherapie und welche werden auch ohne Chemotherapie kein Rezidiv erleiden und profitieren somit nicht von der Chemotherapie?

Die beiden Proteine Urokinase-Typ Plasminogen-Aktivator (uPA) und sein Inhibitor PAI-1, zwei maßgeblich an Tumor-Stoffwechselprozessen beteiligte Proteine, sind ein Beispiel für potenziell prädiktive Biomarker, auf deren Basis die Entscheidung für oder gegen eine Chemotherapie getroffen werden kann. Die Nutzenbewertung dieser beiden Marker wurde kürzlich abgeschlossen [2]. Weitere Marker in dieser Indikation untersucht ein aktuell laufender IQWiG-Bericht [3].

## Studiendesigns in der stratifizierenden Medizin

Um Therapieentscheidungen treffen zu können, die einen Nutzen für Patientinnen bedeuten, werden Biomarker benötigt, deren prädiktiver Wert im Rahmen von RCTs evaluiert wurde. Für die Testung von prädiktiven Biomarkern war also eine Weiterentwicklung der bisher eingesetzten Studiendesigns erforderlich. Während früher einarmige Verlaufsbeobachtungen zur diagnostischen und prognostischen Güte von Biomarkern (Sensitivität, Spezifität) vermehrt zur Anwendung kamen, gelten nach internationalem Konsens mittlerweile unterschiedliche Varianten von RCTs als erforderlicher Standard für die Nutzenbewertung diagnostischer Verfahren [4]. Dabei wird die Anwendung eines Biomarkers innerhalb der diagnostisch-therapeutischen Kette geprüft. Der Variantenreichtum dieser Studien ist groß, mit jeweils unterschiedlichen Vor- und Nachteilen [5]. Hier sollen die drei gängigsten Varianten am Beispiel des primären Brustkrebs vorgestellt werden.

| Studientyp | Beschreibung | Schematische Darstellung | Vorteile | Nachteile |
|---|---|---|---|---|
| **Interaktionsdesign** | Bei allen Patienten werden zu Studienbeginn die Biomarker-Ausprägungen (diagnostischer Test) bestimmt. Die Zuteilung zur Behandlung erfolgt zufällig und unabhängig vom Testergebnis. Untersucht werden die Wechselwirkungen von Marker-Ausprägung und Therapieeffekt. Wird eine qualitative Interaktion (z. B. Test-Positive profitieren von einer Therapie, Test-Negativen schadet sie) oder eine starke quantitative Interaktion (z. B. Test-Positive profitieren deutlich, Test-Negative nur sehr wenig) gefunden, kann der Test zukünftig zur Entscheidung für oder gegen die Anwendung der untersuchten Therapie herangezogen werden. |  | ▪ Nicht nur 2, sondern auch mehrere Biomarker können zeitgleich miteinander verglichen werden.<br>▪ Kann mit geringem zusätzlichen Aufwand in jede Therapiestudie eingebettet werden.<br>▪ Effekt der Therapie kann klar vom (Interaktions-)Effekt der Marker-Anwendung unterschieden werden.<br>▪ Kann grundsätzlich auch retrospektiv durchgeführt werden. | ▪ Ein absoluter Effekt der Anwendung des Markers kann nicht abgeleitet werden.<br>▪ Die statistische Testung trennt nicht zwischen qualitativer und quantitativer Interaktion.<br>▪ Wird das Interaktionsdesign retrospektiv eingesetzt, dann können Vollständigkeit bzw. Vollzähligkeit ein Problem sein (attrition bias). |
| **Strategiedesign** | Bei diesem Studiendesign werden die Patienten zufällig zwei Behandlungsstrategien zugeteilt, einer markerbasierten Behandlung ($S_+$) oder einer Behandlung unabhängig vom Marker ($S_-$). Unterscheiden sich die Effekte in den beiden Behandlungsstrategien, kann dieser Unterschied kausal auf die Anwendung des Markers zurückgeführt werden. |  | ▪ Effekt der gesamten Strategie (diagnostisch-therapeutische Kette) wird getestet.<br>▪ Zusätzlich können direkte (Neben-) Wirkungen des Testens erhoben werden. Das kann besonders bei invasiven Tests, z. B. Biopsien, interessant werden oder auch dann, wenn Aussagen darüber getroffen werden, ob ein Behandlungserfolg überhaupt erwartet werden kann.<br>▪ Unverzerrte Antwort auf eine pragmatische Frage. | ▪ Im Allgemeinen weniger effizient, da im Unterschied zum Interaktionsdesign die gleichzeitige Evaluation mehrerer Marker nicht möglich ist. Außerdem ist eine höhere Fallzahl erforderlich, da der Marker nur in einem Studienarm getestet wird.<br>▪ Zeigt sich kein positiver Effekt der markerbasierten Strategie, bleibt unklar, ob der Biomarker nicht nützlich oder die Therapie unwirksam ist. |
| **Anreicherungsdesign** | Nur Patienten mit einer bestimmten Ausprägung des Markers werden randomisiert und nach-beobachtet. Zeigt sich im markerpositiven Arm ein Effekt, kann zukünftig für markerpositive Patienten eine klare Therapieempfehlung gegeben werden. Dieses Studiendesign ähnelt daher einer klassischen Therapiestudie sehr.<br>(Sofern auch die nicht randomisierte Teilpopulation nachbeobachtet wird, spricht man von einem Hybriddesign) |  | ▪ Effizientes Design, da die Patientenpopulation unter Berücksichtigung der vorliegenden Evidenz gezielt auf die Fragestellung hin ausgewählt werden kann. | ▪ Zeigt sich ein Effekt, bleibt unklar, ob er sich nicht auch in der anderen Ausprägung des Markers (also bei allen Patienten) gezeigt hätte. Daher sind starke Annahmen zum Fehlen eines Behandlungseffekts in der Gruppe der nicht randomisierten Patienten erforderlich. |

## Diskussion

Keines der dargestellten Designs kann als alleiniger „Goldstandard“ zur Bewertung des Nutzens von Biomarkern betrachtet werden. Die Auswahl des Studiendesigns muss sich an der gegebenen Fragestellung und dem Vorwissen dazu sowie an den jeweiligen Rahmenbedingungen orientieren. Sowohl im Bereich der genetischen Biomarker (z. B. Brustkrebs) als auch in der bildgebenden Diagnostik wird aber immer breiter akzeptiert, dass in der Regel RCTs für einen Nutzennachweis erforderlich sind [6].

Studien im Interaktions- oder Strategiedesign können grundsätzlich in Nutzenbewertungen einfließen: Interessant erscheint beim Interaktionsdesign die Möglichkeit, bei bereits abgeschlossenen RCTs retrospektiv den Interaktionseffekt (mehrerer) neuer Marker (parallel) zu untersuchen, sofern ausreichend Tumorgewebe der Studienteilnehmer konserviert worden ist. Das Strategiedesign hat den Vorteil, den Effekt der gesamten diagnostisch-therapeutischen Kette klar und quantifizierbar zu bemessen. Zum Teil können auch Studien im Anreicherungsdesign hinreichend aussagekräftig sein. Hierfür muss gewährleistet sein, dass der Marker tatsächlich nur in der Gruppe der randomisierten Patienten die weitere Behandlung nutzbringend steuern kann. In Situationen, in denen ein neuer Marker einen anderen Marker ersetzen soll, kann auch das Diskordanzdesign (ohne Abbildung) angewendet werden, in dem nur die Patienten mit diskordanter Ausprägung der beiden Biomarker randomisiert werden.

Zu einigen klinischen Diagnostikfragen sind keine RCTs erforderlich. Wenn für einen Marker ein Nutzen bereits belegt werden konnte, dann kann es für einen zweiten Marker ausreichen zu prüfen, ob die Testergebnisse beider Marker hinreichend übereinstimmen [7]. Hierfür eignen sich Konkordanzstudien, in denen die Übereinstimmung der Testergebnisse, also der Anteil gleich klassifizierter Patienten an allen untersuchten Patienten, untersucht wird. Voraussetzung dafür ist allerdings, dass für den neuen Marker eben nicht der Anspruch erhoben wird, mehr „richtig Positive“ oder mehr „richtig Negative“ zu identifizieren.

**Kontakt: martina.markes@iqwig.de**

## Schlussfolgerungen

Die konsequente Anwendung der oben beschriebenen Studiendesigns setzt sich immer mehr durch. Obwohl die methodischen Grundlagen zur Bewertung diagnostischer Interventionen bereits lange bekannt sind, hat mit der raschen Verbreitung von Biomarkern und stratifizierender Medizin auch die Studienmethodik erheblich an Bedeutung und Aufmerksamkeit gewonnen. Während noch vor wenigen Jahren diagnostische Querschnittstudien in vielen Köpfen als Goldstandard galten, lassen sich aus neuen, vergleichenden Studiendesigns in effizienterer Weise relevantere Informationen über den Nutzen prädiktiver Biomarker ableiten. Auch in der Behandlung von Frauen mit Brustkrebs könnte auf diese Weise eine zielgerichtetere und damit wirksamere oder nebenwirkungsärmere Behandlung ermöglicht werden.

## Literatur

1. Windeler J, Lange S. Nutzenbewertung personalisierter Interventionen: methodische Herausforderungen und Lösungsansätze. Ethik Med 2013; 25: 173-182.
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Bestimmung der Antigenexpressionslevel von uPA und PAI-1 beim primären Mammakarzinom mit intermediärem Rückfallrisiko nach R0-Primäroperation: Abschlussbericht; Auftrag D13-02 [online]; 20.10.2014.
3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Biomarkerbasierte Tests zur Entscheidung für oder gegen eine adjuvante systemische Chemotherapie beim primären Mammakarzinom: vorläufiger Berichtsplan; Auftrag D14-01 [online]; 05.09.2014.
4. Köbberling J, Trampisch HJ, Windeler J. Memorandum for the evaluation of diagnostic measures. J Clin Chem Clin Biochem 1990; 20: 873-879.
5. Lijmer JG, Bossuyt PM. Various randomized designs can be used to evaluate medical tests. J Clin Epidemiol 2009; 62: 364-73.
6. Scheibler F, Zumbé P, Janssen I, Viebahn M, Schröer-Günther M, Grosselfinger R, Hausner E, Sauerland S, Lange S. Randomized controlled trials on PET: a systematic review of topics, design, and quality. J Nucl Med 2012; 53: 1016-25.
7. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Entwurf für Version 4.2. Köln: IQWiG; 2014.