

# Früherkennung, Screening – Zielsetzung und Gesetzmäßigkeiten, Teil I\*

## *Early Recognition, Screening – Objectives and Concept, part I*

Heinz-Harald Abholz<sup>1</sup>, Christian Lerch<sup>1</sup>

**Zusammenfassung:** In dem Text werden die wesentlichen Punkte dargestellt und erklärt, die eine Früherkennung/ Screening von herkömmlicher Diagnostik unterscheidet. Anhand von Ergebnissen aus Screening-Studien werden insbesondere die biostatistischen Hintergründe und ethischen Aspekte von Screenings erklärt. Schlussfolgerungen für die Gesundheitsversorgung werden gezogen. Teil 2 wird in Heft 2/2010 erscheinen.

**Summary:** This paper reviews the main topics of screening and shows that screening is a different issue compared to diagnosing. Especially biostatistical and ethical problems are exposed using the findings in several studies on screening. Consequences for health care are pointed out. This is the first of two parts; the second part will be published in issue 2/2010.

**Schlüsselworte:** Früherkennung, Screening, Biostatistik, Ethik, Versorgungssystem

**Keywords:** screening, biostatistics, ethic, health care

### Definitorisches

Früherkennung hat zum Ziel, Krankheit, Vorstufen von Krankheit oder Risikofaktoren für Krankheit frühzeitig zu erkennen und dann zu behandeln. Die im Hintergrund stehende Grundannahme ist: Je früher etwas erkannt wird, umso nutzbringender ist es zu behandeln und umso weniger Nebenwirkungen entstehen bei einer Therapie.

Sollte diese Annahme auch in ihrer Allgemeinheit richtig sein, ergeben sich doch zahlreiche Probleme, die Anlass zu einer differenzierteren Betrachtung sein müssen.

Beinhaltet „Früherkennung“ generell jedes frühe Erkennen („früher als üblich“ – also auch bei den ersten Symptomen), so wird das als Screening (oder auch systematische Früherkennung) verstanden, das sich hinsichtlich der Zielerkrankung nur auf asymptotische Personen bezieht.

Zudem gilt für systematische Screeningverfahren, dass bestimmte Vorgaben eingehalten werden müssen. Diese betreffen die Definition der zu untersuchenden Population, die Zeitintervalle und die Art der Untersuchung sowie die Interpretation der Befunde. Damit

soll – meist auf Studienbasis belegt – ein Optimum zwischen Nutzen und Schaden sowie Nutzen und Aufwand gewährleistet werden.

Die folgende Darstellung ist ausschließlich auf die systematische Früherkennung (Screening) bezogen; nur in diesem Sinne wird von Früherkennung gesprochen. Dabei werden überwiegend Beispiele aus dem Bereich der Krebs-Früherkennung benutzt, da hier die Studienlage am fundiertesten ist.

### Screeningprinzip und Zielkrankheiten

Screening-Untersuchungen sind auf die asymptomatische Phase einzelner Krankheiten, Vorstufen von Krankheiten sowie Risikofaktoren ausgerichtet. Die jeweiligen Störungen, die im Visier der Früherkennung stehen, werden als Zielkrankheiten bezeichnet. Es geht also z. B. um eine gestörte Glukosetoleranz (als Vorform von Krankheit), um einen asymptomatischen Diabetes mellitus (frühe, asymptomatische Phase), um große Polypen im Dickdarm (als potenzielle Vorform eines Dickdarmkarzinoms), um einen arteriellen Hochdruck

(Risikofaktor für kardiovaskuläre Erkrankung) oder um die Intimadicke der Carotis (Risikoindikator für kardiovaskuläre Erkrankung).

Das Grundprinzip von Früherkennung und damit auch von Screening ist einfach und erscheint erst einmal plausibel: „Je früher, desto besser“. Es bedeutet, dass ein frühes Erkennen in der Regel die Chance für eine erfolgreiche Therapie erhöht und/oder die unerwünschten Wirkungen einer Therapie eher niedrig zu halten erlaubt.

Bei genauerer Betrachtung aber ist das Prinzip voller Probleme, die im Folgenden deutlich gemacht werden sollen.

Um diese Problematik zu verstehen, muss man sich den sogenannten natürlichen Verlauf (natural history) der jeweils im Hintergrund stehenden Krankheit ansehen [1]. Eine solche natural history lässt sich auf einer übergeordneten, grundsätzlichen Ebene abhandeln – so wie dies hier geschehen soll. Es ist aber zu betonen, dass der natürliche Verlauf jeder einzelnen Krankheit in seinen Stadien sowie deren Verlauf und Dauer unterschiedlich ist bzw. dass dazu nicht immer alle Einzelheiten bekannt sind.

In Abbildung 1 ist der verallgemeinerte Verlauf einer Krankheit (natural

<sup>1</sup> Abt. für Allgemeinmedizin, Universitätsklinikum Düsseldorf

\* Der Text ist auf Grundlage eines Expertengutachtens im Rahmen des Sondergutachten 2009 des „Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen“ entstanden: <http://www.svr-gesundheit.de/Startseite/Startseite.htm>

Peer reviewed article eingereicht: 17.09.2009; akzeptiert: 29.10.2009

DOI 10.3238/zfa.2009.0039

REPARIERENDE FAKTOREN	KRANKHEITS- STADIUM	FÖRDERNDE FAKTOREN
Reparierende Faktoren ⇒	<b>Disposition, Genetik</b> ⇓	Risikofaktor 1 ⇐
Reparierende Faktoren ⇒	<b>Vorformen von Krankheit</b> ⇓	Risikofaktor 2 ⇐
Reparierende Faktoren ⇒	<b>Asymptomatischer Beginn der Krankheit</b> ⇓	Risikofaktor 3 ⇐
Reparierende Faktoren ⇒	<b>Krankheits- manifestation</b> ⇓	Risikofaktor 4 ⇐
	<b>Krankheitsfolgen/ Organschäden</b>	

**Abbildung 1** Formale Genese einer chronischen Krankheit (natural history).

history) von den ersten Veränderungen bis zur symptomatischen Krankheit wiedergegeben. Dabei ist an jeder Stelle dieser Entwicklung dargestellt,

- dass sich eine Vorstufe oder eine erste asymptomatische Krankheitsmanifestation weiterentwickeln kann, aber nicht muss und
- unter welchen Einflüssen dies in die eine oder die andere Richtung geht.

Eine solche sogenannte formale Krankheitsentwicklung gibt es für jede Krankheit, vom Karzinom bis zum Diabetes mellitus, einer Hormonstörung oder einer Gefäßveränderung. Dabei sind hier primär nur Entwicklungen für chronische Erkrankungen relevant, weil sich ein Früherkennungsprogramm in der Regel auch nur auf diese bezieht [2].

Die Abbildung zeigt, dass an jedem Übergang von der einen zur anderen Stufe Reparaturvorgänge des Körpers erfolgen können, die diesen Weg zur nächsten Stufe verhindern oder verlangsamen können. Auf der anderen Seite gibt es – zumindest konzeptionell – für alle Übergänge auch Risikofaktoren, die den Weg ermöglichen oder beschleunigen.

Am Beispiel einer Krebserkrankung kann dies verdeutlicht werden: Von ersten Zellentartungen, deren Entstehung unterschiedlich begründet sein kann (genetisch, Bestrahlungsfolge, andere „Karzinogene“), geht es zu ersten Zellverbänden, die als Vorstufen eines Krebses anzusehen sind. Oder es werden diese ersten zellulären „Entartungen“ vom Körper wieder repariert – dann endet der Ablauf hier. Von den entarteten Zellverbänden kann es aber auch weiter zu makroskopisch sichtbaren Krebsveränderungen kommen, die immer noch asymptomatisch sind. Es kann jedoch auch hier noch eine körpereigene Korrektur erfolgen. Alternativ schreitet entartetes Wachstum fort, eine Krebserkrankung wird symptomatisch [2].

Zum Beispiel weiß man, dass ein ductales carcinoma in situ (DCIS) die Vorstufe eines Mammakarzinoms sein kann, was hier bereits eine gewebliche, nicht nur eine zelluläre Veränderung darstellt. Dennoch entwickeln sich nur etwa 20 bis 30 % dieser DCIS auch zu einem Mammakarzinom [4].

Für den Befall mit Humanen Papilloma-Viren (HPV) im Bereich der Gebärmutter der Frau wissen wir, dass nur ein

minimaler Anteil der Frauen, die einen solchen Befall aufweisen, im Verlauf des Lebens zelluläre Entartungen und schließlich ein Gebärmutterhalskarzinom entwickelt [4]. Etwa 90 % der Frauen weisen in ihrem Leben eine solche Besiedlung mit HPV auf, nur bei etwa 5 % aber kommt es zu zellulären Entartungen, die zu einem operativen Eingriff Anlass geben können; bei noch deutlich weniger Frauen entsteht ein Gebärmutterhalskarzinom [4].

Einschränkungen in Bezug auf die Aussagen zur Reparaturfähigkeit des Körpers stellen die genetisch fixierten Erkrankungen sowie schon sehr ausgeprägte Krankheitsentwicklungen dar. Aber selbst bei den ersteren weisen nur wenige eine annähernd hundertprozentige Determination auf.

Damit ergibt sich die erste und entscheidende Einschränkung zu der Grundannahme von Screenings, dass „früher entdeckt“ immer auch „besser behandelt“ heißt: Denn je früher etwas entdeckt und behandelt wird, umso eher verhindert man auch, dass der Organismus sich selbst „repariert“. Damit aber taucht die für die Früherkennung typische Problematik auf, dass bei sehr früher Erkennung auch sehr viele unter den entdeckten Fällen „umsonst“ therapiert werden. Denn ein nennenswerter Teil der dann therapierten Personen würde „sich selbst reparieren“.

Die Früherkennung des häufigsten extrakraniellen Tumors im Kindesalter, des Neuroblastoms, hat dies deutlich gemacht: Man entdeckte deutlich mehr Tumore als je zuvor manifest wurden. Diese wurden aber alle – teilweise sehr eingreifend – therapiert. Damit hatte man in der Bilanz mehr Schaden als Nutzen erreicht – insbesondere, weil die Behandlung sehr eingreifend, also mit nennenswerten Nebenwirkungen bis hin zur Todesfolge verbunden ist [5, 6].

### **Ethische Bedingungen eines Früherkennungsprogramms (Screenings)**

Es gibt einige ethische Bedingungen, die vor Einführung eines Früherkennungsprogramms geprüft werden müssen – so wie dies im Auftrag der WHO vor nun 40 Jahren orientierend festgehalten wurde.

Die Notwendigkeit einer besonders gründlichen ethischen Prüfung begrün-

1)	Die Behandlung der gesuchten Erkrankung muss nachgewiesenermaßen einen Nutzen haben (möglichst mit Endpunktstudien belegt).
2)	Der natürliche Verlauf der Krankheit und die möglichen Interventionszeitpunkte mit günstigem Einfluss müssen bekannt sein.
3)	Früherkennungsdiagnostik muss, auch bei den Konsequenzen einer notwendig werdenden Abklärungsdiagnostik, wenig eingreifend sein.
4)	Die gesuchte Krankheit muss relativ häufig sein bzw. ein relevantes Gesundheitsproblem in einer Gesellschaft darstellen.
5)	Aufwand und Nutzen – für beteiligte Personen und von ökonomischer Seite – müssen in einem günstigen Verhältnis zueinander stehen.
6)	Bei einer Betrachtung von Nutzen und Schaden muss – bezogen auf alle am Screening Beteiligten – eine positive Bilanz vorliegen.
Heutzutage würde man noch ergänzen:	
7)	Personen müssen vor Durchführung einer Früherkennung auf deren Rahmenbedingungen (Aussagekraft und mögliche Folgen, Abklärungsdiagnostik und Befunderhebung, Therapien und deren Nebenwirkungen) aufgeklärt werden, um dann erst eine Entscheidung zur Teilnahme auszusprechen.

**Tabelle 1** Ethische Grundprinzipien von Früherkennung (Gekürzt und umformuliert nach Wilson/Jungner) [7].

det sich darin, dass hier primär und ganz überwiegend gesunde Personen einer medizinischen Maßnahme und ihrer Folgediagnostik unterzogen werden; also der Maßstab in Bezug auf Nebenwirkungen strenger als bei der Therapie eines Patienten sein muss [8, 9]. Denn bei einer Diagnostik eines symptomatischen Patienten ist es bekanntermaßen das Ziel, „Leid“ oder Beschwerden zu lindern. Zum anderen ist in der Gruppe von Patienten mit Symptomatik die Wahrscheinlichkeit, dass eine Krankheit gefunden wird, ungleich höher als in einer Screening-Gruppe (bei letzterer geht es ja – definitionsgemäß – um asymptomatische Personen in Bezug auf die Zielerkrankung und damit um solche mit geringerer Krankheitswahrscheinlichkeit). Die Wahrscheinlichkeit, über potenziell auch gefährliche Diagnostik wirklich etwas Hilfreiches für die versorgte Person tun zu können, ist bei Screenings also ungleich niedriger.

Zu den oben aufgeführten Punkten (Tab. 1) sind zwei Anmerkungen nötig:

**Zu Punkt 2:** Der natürliche Verlauf muss bekannt sein: Wenn, wie beim Neuroblastom-Screening, dieser nicht ausreichend bekannt ist, kann ein „falscher“ Zeitpunkt eines Screenings zu mehr Schaden als Nutzen führen. Ähnlich verhielt es sich beim Screening auf Bronchialkarzinom in den 80er und

90er Jahren des letzten Jahrhunderts: Man fand im jährlichen, dann halbjährlichen Röntgen-Screening und im Sputum-Screening zwar mehr Karzinome, aber diese ganz überwiegend in einer Phase, in der Therapiemaßnahmen nicht mehr Nutzen hatten als in einer späteren Phase des Krankheitsverlaufs [10, 11].

Umgekehrt, aber auf das gleiche Problem hinauslaufend, verhält es sich beim Screening mittels Spiral-CT der Lunge: Man entdeckt überwiegend Veränderungen, die sich im Verlauf als harmlos herausstellen; mit der Methode lassen diese sich aber nicht ausreichend von den problematischen Veränderungen trennen [12].

Ein anderes Beispiel für das gleiche Problem: Ein Screening bei Schwangeren mittels Glukosebelastungstest wird eine Gruppe mit dem Label „Gestationsdiabetes“ belegen. Von diesen Frauen wird später aber nur ein kleiner Teil tatsächlich einen Diabetes mellitus entwickeln oder ihr „werdendes Kind“ gefährden [13]. Solange keine genaue Angabe möglich ist (durch Studien belegt), ab welchem Grad der Glukose-Stoffwechselstörung (also ab welcher Zuckerbelastung) die Wahrscheinlichkeit einer solchen Entwicklung höhergradig ist, besteht das Problem der faktischen Übertherapie, zumindest aber das des „Labelings“.

Das Gleiche gilt für die immer wieder vorgeschlagenen Messungen von Verkalkungsplaques an den arteriellen Halsgefäßen, Ganzkörper-CTs etc. [14, 15]. Auch hier weiß man nicht, welche „Phase“ eines Krankheitsgeschehens „gemessen“ wird und welche Verläufe mit welchen Wahrscheinlichkeiten für klinisch relevante Krankheiten anzunehmen sind.

**Zu Punkt 3:** Früherkennungsdiagnostik und deren Folgediagnostik muss zumutbar sein: Regelmäßige Ganzkörperuntersuchungen mittels Computertomographie zur Früherkennung sind möglich, nur sind sie als Strahlenbelastung – regelmäßig durchgeführt – nicht zumutbar, weil sie in der Summe wahrscheinlich mehr Schaden anrichten als Nutzen stiften [14].

Ein Screening auf Eierstockkarzinom mittels Ultraschall und Hormonbestimmungen wird international abgelehnt. Es würde – bei der Seltenheit dieses Karzinoms – deutlich mehr falsch-als richtig-positive Befunde ergeben [16]. Um dies aber im Falle eines krankhaften Screening-Befundes diagnostisch abzuklären, ist jeweils bei allen „Screening-Positiven“ eine Laparoskopie notwendig. Diese Untersuchung hat zwar nur selten – in der Masse dann aber doch sehr relevant – bedrohliche Nebenwirkungen. Zudem würde sich das Risiko solcher unerwünschten Effekte ganz überwiegend auf Frauen beziehen, die sich dann später als „gesund“ herausstellen (weil die Zahl falsch-positiver die der richtig-positiven Befunde übersteigt).

### Biostatistische Besonderheiten von Früherkennung

Jegliche Diagnostik, einschließlich der in Früherkennungsuntersuchungen angewendeten, ist in Bezug auf das Vorliegen oder Fehlen einer Erkrankung nie hundertprozentig sicher. Man bezeichnet die diagnostischen Kenngrößen eines Verfahrens mit Sensitivität und Spezifität.

Bei Screening-Untersuchungen versteht man unter

- Sensitivität den Prozentsatz der richtig entdeckten Fälle an allen im Untersuchungskollektiv vorhandenen Kranken. Und unter

**A: Diagnostik: 300 von 1000 sind krank**

(Prävalenz 30/100)

Sensitivität und Spezifität mit 90 % jeweils angenommen

		Test positiv	Test negativ
Krank	300	270	30
Gesund	700	70	630
Alle 1000 Positiven		340	

270 von 340 positiven Befunden sind richtig positiv.

Positive prädikative Wertigkeit (ppV) rund 80 %

**B: Screening: 3 von 1000 sind krank**

(Prävalenz 3/1000)

Sensitivität und Spezifität mit jeweils 90 % angenommen

		Test positiv	Test negativ
Krank	3	2,7	0,3
Gesund	997	99,7	897,3
Alle 1000 Positiven		102,4	

Somit sind von abgerundet 102 positiven Befunden abgerundet 3 richtig-positiv. PpV liegt damit bei 3 %.

**Abbildung 2** Positive Prädikative Wertigkeit von Mammographie bei Diagnostik- und Screening-Einsatz.

- Spezifität den Prozentsatz der richtig als gesund klassifizierten Personen unter allen Gesunden im Untersuchungskollektiv.

Abhängig vom Screening-Test und der gesuchten Krankheit liegen Sensitivität und Spezifität meist zwischen 70 und 90 %, manchmal jedoch noch niedriger, selten höher [17]. Dies wäre bei der Diagnostik bestehender Krankheitssymptome keine sehr erhebliche Einschränkung der Nutzbarkeit eines diagnostischen Instruments (Laborwert oder technisch vermittelte Diagnostik, Bildgebung). Denn aufgrund der Symptomatik können stets mehrere, sich ergän-

zende diagnostische Maßnahmen parallel oder nacheinander durchgeführt werden. Für den Patienten erscheint dies zumutbar, hat er doch das Anliegen, seine Symptomatik abzuklären. Darüber hinaus leitet die Art und Ausgestaltung der Symptomatik den Vorgehensweg zu unterschiedlichen diagnostischen Zugängen.

In einer Früherkennungspopulation kann man nicht aufgrund von Krankheitssymptomen Hinweise gewinnen, ob man diagnostisch „weiterrufen“ muss oder abbrechen kann, weil man mit hoher Wahrscheinlichkeit vor einem falsch-positiven Befund steht. Denn Früherkennung beinhaltet Diag-

nostik bei asymptomatischen Personen – von denen man weiß, dass die ganz überwiegende Mehrheit von ihnen gesund ist, also nicht die gesuchte Erkrankung haben wird.

Zum Beispiel ist die Prävalenz des Mammakarzinoms selbst in der Hochrisikogruppe der 50-jährigen und älteren Frauen nur zwischen 4 und 6 pro 1000 Frauen in der ersten Screeningrunde, bei der alle vorliegenden, aber bisher nicht bekannte Erkrankungen festgestellt werden (Prävalenzrunde). Bei den Folgerunden, bei der nur noch die Neuerkrankungen festgestellt werden (Inzidenzrunden), liegt diese dann nur um 3 pro 1000 untersuchte Frauen. Bei einer „symptomatischen Frau“ (z. B. bei deutlich tastbarem Knoten) hingegen liegt die „Krebswahrscheinlichkeit“ – je nach weiterer Konstellation – z. B. im Bereich von 30 bis 80 %.

Die anzunehmende Wahrscheinlichkeit einer gesuchten Pathologie, also die Prävalenz oder Inzidenz der gesuchten Krankheit, ist von entscheidender Bedeutung für die Prädiktion eines Testergebnisses: Ist die Krankheit häufig, dann ist die Wahrscheinlichkeit, dass sich hinter einem positiven Test-Ergebnis auch die Krankheit verbirgt, eher hoch. Ist aber die Wahrscheinlichkeit einer Erkrankung gering, dann ist selbst bei Vorliegen eines positiven Testergebnisses die Wahrscheinlichkeit, dass dieses falsch-positiv ist, eher hoch. Da in allen Screenings die Wahrscheinlichkeit gesuchter Erkrankung immer niedrig bis sehr niedrig ist, besteht hier ein besonders ausgeprägtes Problem: Es ist bei positiven Befunden immer davon auszugehen, dass der größte Teil falsch-positiv ist. Dies aber kann nur durch Abklärungsdiagnostik bestimmt werden. Es geht also um die Frage, mit welcher Wahrscheinlichkeit nun wirklich eine Erkrankung oder keine Erkrankung vorliegt, wenn man einen positiven – bzw. einen negativen – Screening-Befund vorliegen hat. Dies wird – neben Sensitivität und Spezifität – im Wesentlichen durch die zunehmende Häufigkeit der Erkrankung in der gesuchten Untersuchungspopulation (Prävalenz) bestimmt – das sog. Bayes Theorem.

Die positive Vorhersagekraft (positive prädikative Wertigkeit [ppW] oder positive predictive value [ppV]) eines Testes bezeichnet den Prozentsatz der wirklich Kranken an all denen, die einen

		Test positiv °	Test negativ
Alle Personen	10 000		
Frühe Mammakarzinome <u>und</u> dabei mit Nutzen versorgte°	5	5 (aufgerundet)	0
Ohne Mammakarzinom <u>oder</u> ohne „frühe Formen“	9 995	999	8996
Alle Fälle mit pos. oder neg. Befund		1004	8996
PpV bzw. npV		0,5 %	100 %
° = auf Basis einer Prävalenz von 3/1000 bzw. 30/10000 sind 30 „Fälle“ zu erwarten; davon 15 % als „Frühe Fälle mit Therapieerfolg“ einzustufen; dies sind 5 Fälle.			

**Tabelle 2** Vier-Felder-Tafel zur „wahren“ prädiktiven Wertigkeit eines Mammographie-Screenings (alle Berechnungen bei Annahme von 90 % Sensitivität und 90% Spezifität).

krankhaften Befund im Test (Screening) aufweisen.

Sie lässt sich am einfachsten in einer Vier-Felder-Tafel berechnen, in der die angenommene Prävalenz/Inzidenz einer Erkrankung eingetragen ist, und dann mittels der bekannten Sensitivität und Spezifität des eingesetzten Tests die entdeckten und übersehenen Fälle in den 4 Feldern berechnet werden können.

Dies ist in Abb. 2 für zwei unterschiedliche Prävalenzen am Beispiel einer Mammographie zur Diagnostik (A) oder einer zur Früherkennung (B) dargestellt. Dabei wird der Vereinfachung wegen von einer Sensitivität und Spezifität von jeweils 90 % ausgegangen. Dabei wird deutlich, dass – in Abhängigkeit von der im Hintergrund stehenden Prävalenz der gesuchten Krankheit – identische Diagnostik eine sehr unterschiedliche positive prädiktive Wertigkeit (ppW oder ppV) aufweist. Im gewählten Beispiel der Mammographie ist sie mit 80 % bei diagnostischem Einsatz hoch; beim Einsatz im Screening niedrig, nämlich 2 %. (Da in der Wirklichkeit die Sensitivität der Mammographie über 90 % liegt, ist die ppV auch eher um die 10 %).

Man sieht, dass in der Population von Kranken mit hoher Prävalenz (Abb. 2, Teil A) die (hier: positive) Vorhersagekraft sehr gut ist. In einer Screeningpopulation von Frauen ab 50 Jahren ist bei positivem Ausfall des „Testes“ (hier der Mammographie) die gleiche diagnostische Methode in ihrer Prädiktion höchst unzureichend.

Noch deutlich niedriger würde die positive prädiktive Wertigkeit (ppW) bei einem Screening ausfallen, wenn die

Prävalenz/Inzidenz noch deutlich niedriger läge – so wie bei Frauen zwischen 35 bis 45 Jahren mit einer Prävalenz des Mammakarzinoms von 1:1000.

Bei Kindern gibt es einen Teil von Früherkennungsuntersuchungen mit besonders hoher Sensitivität und Spezifität; dies gilt z. B. für die Testungen auf angeborene Stoffwechselstörungen (Sensitivität und Spezifität von nahezu 100 %). Allerdings ist hier die Prävalenz der gesuchten Erkrankungen noch weit aus niedriger (z. B. 1:10000) als bei allen im Erwachsenenalter gesuchten Erkrankungen. Es werden daher selbst mit diesen hoch-spezifischen Tests überwiegend erst einmal falsch-positive Befunde erhoben, die über Abklärungsdiagnostik dann wiederum als „gesund“ eingeordnet werden müssen.

Da hingegen die negative prädiktive Wertigkeit (npW oder npV) bei sehr niedriger Prävalenz immer sehr hoch ist, kann man bei negativem Testausfall fast immer von der Abwesenheit der gesuchten Erkrankung ausgehen.

### Worauf sollte sich Prädiktivität eigentlich beziehen?

Aus dem oben Gesagten, nach dem sich Früherkennungsprogramme auf die frühen Formen von Krankheit, deren Vorstufen oder Risikofaktoren beziehen, ergibt sich, dass die hier gemachten Aussagen zur positiven und negativen Prädiktivität eines Screening-Tests noch differenzierter dargestellt werden müssten. Sie beziehen sich, z. B. in Abb. 2, auf alle entdeckten Erkrankungen, also auch de-

ren spätere Formen und nicht nur deren Vorstufen und Frühformen. Mit anderen Worten, es werden bei Berechnungen der prädiktiven Wertigkeit alle Erkrankungsfälle (also nicht nur die frühen) sowie alle Vorformen und Risikofaktoren (also nicht nur die, die zur Krankheit führen würden) berücksichtigt.

Damit aber fokussiert man nicht auf das eigentliche Ziel der Früherkennung, nämlich die frühen Formen und Vorstufen oder gar Risikofaktoren. Mit anderen Worten: Wenn man das Ziel von Früherkennung, erfolgreicher behandeln zu können und zugleich Überbehandlung nicht zu sehr zu betreiben, im Auge hat, wird die prädiktive Wertigkeit des Screening-Tests mit einem solchen, durchaus üblichen Vorgehen stark überschätzt.

Dazu ist es wichtig festzuhalten: Bei einem Screening werden zahlenmäßig ganz überwiegend spätere Formen gefunden, die nicht oder nur mit einer geringeren Verbesserung der Therapiechance einhergehen (siehe Teil II). Zudem wird ein nennenswerter Teil der gefundenen „Fälle“ mit besserer Therapiechance nicht an der Zielerkrankung klinisch relevant erkranken oder gar daran versterben (s. Teil II).

Wenn man also die obige Vier-Felder-Tafel ergänzt um die Spalte „krank in früher Form“ oder gar „krank in früher Form und durch das Screening von Krankheit/Tod bewahrt“, so kommt man auf eine 8-Felder-Tafel. In dieser sind dann neben der klassischen Auswertung (wie Abb. 2) auch die Prädiktionszahlen in Bezug auf die eigentlichen Zielerkrankungen eines Screenings, die frühen Formen, ergänzt.

Für das Mammakarzinom ist dies hier einmal mittels angenommener, aber realistischer Zahlen dargestellt: Man geht in den sogenannten Wiederholungsrunden eines Screenings (Inzidenzrunden) davon aus, dass a) etwa 3 pro 1000 Frauen ein Inzidenzkarzinom aufweisen und b) maximal 15 % der gefundenen neuen Tumore solche sind, die eine bessere Therapierbarkeit haben und bei denen zugleich diese bessere Therapierbarkeit nicht durch andere Faktoren (wie zum Beispiel konkurrierende Todesursachen etc., s. Teil II) konkurrenzlos werden.

Im Folgenden ist der Übersichtlichkeit wegen nur die Berechnung für die „frühen Formen mit Nutzen profitie-



rend“ vorgenommen worden (Tab. 2). Man kann aber die Abb. 2, Teil B dagegen stellen, um den Unterschied in der prädiktiven Wertigkeit – bei ja ansonsten identischen Rechenbedingungen – zu sehen. Stellt man die beiden Abbildungen nebeneinander, dann kommt man zu der vergleichenden 8-Felder-Tafel. Betrug in Abb. 2 die positive prädiktive Wertigkeit aufgerundet 2 %, so ist sie nun, falls nur die „Frühen und mit Nutzen profitierenden Fälle“ in die Rechnung eingebracht werden, nur noch 0,5 %. Einfluss auf die negative prädiktive Wertigkeit hat diese unterschiedliche Sichtweise hingegen nicht: Sie ist in beiden Fällen 100 %.

Aus verschiedenen Gründen hat sich eine solche Darstellung jedoch beim konkreten Umgang mit einem Screening nicht durchgesetzt. Grund dafür ist, dass die Grenze zwischen den Gruppen mit besserer vs. schlechterer Therapiechance nicht gut festlegbar ist, da sie von vielen Bedingungen aufseiten der teilnehmenden Personen, der Screening-Methode sowie der ärztlichen Einschätzung besserer Therapierbarkeit abhängt.

Dennoch ist eine solche Darstellung didaktisch hilfreich, weil sie deutlich macht, dass für die eigentliche Zielerkrankung eines Screenings, also die frühe Form mit besserem und realisiertem Therapieerfolg, die positive prädiktive Wertigkeit (ppW) nochmals deutlich kleiner ist. Dies aber heißt, dass sich deutlich weniger Personen als nach klassischer ppW-Berechnung unter denen mit einem positiven Befund befinden, denen dann wirklich über die Früherkennung geholfen werden kann.

### Unterschiedliche Definitionskriterien für „krankhaft“

Da es bei Früherkennung um Vorformen und Risiken als auch um möglichst frühe Formen von Krankheit geht, wird nach einer anderen „Krankheitsentität“ als der klinisch manifesten Krankheit gesucht.

Sowohl in bildgebenden Verfahren als auch in Untersuchungen von Gewebeschnitten im Rahmen von Screenings ist es so, dass die Charakteristika des Krankhaften spärlicher, weniger entwickelt sowie weniger eindeutig sind. In einem Bild gesprochen: Zwischen

schwarz und weiß ist einfach zu unterscheiden, zwischen nuancierten Grautönen weitaus schwieriger; aber gerade um diese Nuance mehr an „grau“ – und nicht mehr weiß – geht es beim Screening [2].

Dies hat zweierlei zur Folge: Einmal ist die Befundung weitaus schwieriger, weil sie in dem „grauen“ Übergangsbereich zwischen krank und gesund stattfindet und es beim Screening eben um gerade diesen Übergangsbereich geht. Gesucht wird das „Nur-wenig-Kranke“, nicht das Vollbild der Krankheit. Diagnostiker, die ihren Blick und ihre Erfahrung an dem zumeist „deutlich Krankhaften“ geschult haben, müssen das Erkennen dieser Übergänge erst erlernen.

Deutlich wurde dies zum Beispiel, als erfahrene Mammographie-Untersucher in Deutschland Mammographie-Screening-Bilder aus den Niederlanden beurteilen sollten; dies geschah in Vorbereitung zum Deutschen Mammographie-Programm. Diese Bilder waren über den Zeitverlauf der Nachbeobachtung nach Herstellung der Bilder in Bezug auf „wirklich krankhaft“ bzw. „wirklich gesund“ validiert; man kannte also die „wahren Befunde“. Die deutschen Mammographie-Befunder versagten mehrheitlich bei ihrer Befundung und begriffen nun, dass hier eine ganz neue Sichtweise geschult werden muss (von Karsa, persönl. Mitteilung). Aus einem solchen Hintergrund leitet sich auch die Forderung nach Zweitbefundung ab.

Aber auch bei Zweitbefundungen gibt es ein Problem: Entscheidet man sich für „krankhaft“, wenn mindestens einer der beiden Befunder so urteilt, oder tut man dies nur, wenn beide Befunder bzw. drei (nach Hinzuziehung eines dritten Befunders) mehrheitlich für „krankhaft“ stimmen? Der jeweils zu erreichende Grenznutzen kann aus Studienkollektiven berechnet werden. Dann kann entschieden werden, was jede Verbesserung an Erkennen kostet – ausgedrückt in Geld, aber auch in Belastung für die Screening-Teilnehmer, die zusätzliche, ganz überwiegend ja mit dem Befund „gesund“ endende Abklärungsdiagnostik über sich ergehen lassen müssen [18]. Da diese Abklärungsdiagnostik auch schwere Folgen haben kann, ist dies kein banales Thema, wie am Beispiel des Screenings auf Down-Syndrom gezeigt wurde [19].

Aber selbst bei den Laborbefunden, die ja sehr exakt anzugeben sind, gibt es – neben Normwertabweichungen – ein Problem bei der Definition des Krankhaften. Der Bereich, in dem das Krankhafte beginnt, ist auch der Bereich, in dem sich das Gesunde noch zum Teil befindet. Die Verteilungskurven von Kranken mit ihren Labor-Werten und Gesunden mit ihren Labor-Werten überlappen sich nennenswert. Und gerade in diesem Überlappungsbereich sind die sehr frühen Formen von Krankheit zu suchen; nicht aber bei den deutlichen Abweichungen von den Normwerten [17].

Man hat also bei Früherkennung die besondere Aufgabe, zwischen „noch gesund“ und „schon krank“ einzustufen. Geht man dabei eher in die eine Richtung, dann erhöht man die Sensitivität eines Screening-Tests (und vermindert die positive prädiktive Wertigkeit). Geht man eher in die andere Richtung, dann vermindert man die Sensitivität, übersieht aber nennenswert gesuchte Personen mit früher Krankheit und erhöht die positive prädiktive Wertigkeit (zu weiteren Problemen der Sensitivität bei Screenings siehe [20]).

### Krankheitsdefinition mittels Verlauf und Therapiestudien

Die bei Früherkennung erfolgende Konzentration auf einen Übergangsbereich ist auch der Grund dafür, dass man bei sehr vielen Werten oder beschreibbaren Befunden Argumente finden kann, sie als gerade schon im Krankhaften oder aber eben nicht im Krankhaften einzustufen. Denn es liegt fast immer eine Kontinuität des Risikos für eine Krankheit vor: Ein Blutdruckwert von systolisch 145 mm Hg beinhaltet immer noch ein höheres Risiko als einer von 140 mmHg, einen Schlaganfall zu bekommen. Und ein Polyp im Darm von 2 mm Durchmesser hat ein kleineres Risiko zu entarten als einer mit 2,5 mm Durchmesser.

In einer solchen Situation ist es sinnvoll, für „Grenzwerte“ den Folgenutzen und Folgeschaden zu beschreiben. Gemeint ist damit, dass auf Basis derartiger Grenzwerte Kollektive von Personen resultieren, die man behandeln kann oder eben auch nicht. Für diese lassen sich dann in Studien bei Vergleich von „Behandlung“ vs. „keiner Behandlung“ die

**Prof. Dr. med. H.-H. Abholz ist ...**



... Facharzt für Allgemeinmedizin und Facharzt für Innere Medizin. Langjährig in Klinik und Allgemeinpraxis tätig. Seit 1998 Leiter der Abteilung für Allgemeinmedizin der Universität Düsseldorf.

sog. „Gewonnene Lebensjahre“ und „Verhinderter Tod durch Behandlung“ – in Abhängigkeit des Grenzwertes zwischen den Kollektiven – bestimmen. Dem können gegenübergestellt werden: Schaden und Kosten der Behandlung – ebenfalls in Abhängigkeit eines Grenzwertes.

Um dies aber als Entscheidungshilfe tun zu können, muss man entsprechende Studien durchgeführt haben. Diese aber stehen selbst bei sehr häufigen Gesundheitsproblemen, wie z. B. Definition des Diabetes, eines Risikos für eine diabetische Augenschädigung, Risiko der kognitiven Einschränkung im Alter etc. häufig nicht zur Verfügung. Dabei sind die „Überlappungen“ der Kollektive, die in der einen oder der anderen Weise als „krank“ definiert sind, nicht selten irritierend gering. In Bezug auf die Diagnose des Diabetes oder des Risikos einer späteren diabetischen Augenschädigung umfasst der Überlappungsbereich auf Basis unter-

schiedlich definierter „Kranken“- bzw. Risiko-Gruppen nicht einmal ein Drittel der Personen [21, 22]. D.h. dass unterschiedliche Definitionen und unterschiedliche Grenzwerte zur Definition einer Krankheit eines Risikos zu sehr unterschiedlichen Kollektionen von „Kranken“ führen. Welche Definition die adäquatere ist, müsste auch hier durch Verlaufsstudien der später dann auftretenden Schäden bzw. derer Verhinderung – im Fall einer Therapie – entschieden werden.

Für einige Fragen gibt es schon solche Studien: Das, was wir heute zum Beispiel als „krank“ in Bezug auf einen Blutdruck festgelegt haben, macht sich an durchgeführten Verlaufsstudien oder – noch tauglicher – randomisierten Therapiestudien fest. Es ging dabei immer darum festzustellen, wie das Risiko von Hochdruck-Folgeerkrankungen bei Nutzung unterschiedlicher Grenzwerte zur Therapieindikation wächst bzw. geringer wird. Bei gleichzeitiger Erfassung von Neben-

wirkungen und Kosten, der Berechnung der Zahl derjenigen, die behandelt werden müssen, um einem in Bezug auf Morbidität oder Mortalität zu helfen (NNT, number needed to treat), kann man sich dann auf einen allgemein akzeptierten Grenzwert zur Therapie gesellschaftlich einigen. Dass es hier um eine gesellschaftliche Einigung, nicht um eine biologische Größe geht, sieht man allein schon daran, dass unterschiedliche Länder und unterschiedliche Leitlinien unterschiedliche Grenzwerte für Therapieindikationen haben.

**Interessenkonflikte:** keine angegeben.

**Literatur**

Die hier zitierte Literatur wird am Ende von Teil II aufgeführt.

**Korrespondenzadresse:**

Prof. Dr. med. Heinz-Harald Abholz  
Abt. Allgemeinmedizin  
Heinrich-Heine-Universität  
Universitätsklinikum Düsseldorf  
Moorenstraße 5  
40225 Düsseldorf  
Tel.: 0211 / 8 11 77 71  
Fax: 0211 / 8 11 87 55  
E-Mail: abholz@med.uni-duesseldorf.de