

1 Prinzipien und Methoden von Früherkennungsuntersuchungen

Bernt-Peter Robra und Norbert Schmacke

C. Günster | J. Klauber | B.-P. Robra | N. Schmacke | C. Schmuker (Hrsg.) Versorgungs-Report Früherkennung
DOI 10.32745/9783954664023-1, © MWV Medizinisch Wissenschaftliche Verlagsgesellschaft Berlin 2019

Abstract

Früherkennung von Erkrankungen ist nur dann legitim, wenn durch hinreichend große randomisierte kontrollierte Studien belegt ist, dass sie tatsächlich einen klinisch relevant günstigeren Verlauf der Erkrankung bewirkt. Eine naive Gleichsetzung von „früh(er)“ mit „wirksam(er)“ und „kostengünstig(er)“ ist nicht haltbar. In dem Umfang, in dem Primärprävention möglich ist, z.B. durch Verzicht auf Noxen, oder kurative Behandlung erfolgreich ist, verliert Früherkennung an Bedeutung.

Der vorliegende Beitrag beschreibt Begriffe und Prinzipien, die bei Vorbereitung, Einführung und Evaluation von Früherkennungsprogrammen berücksichtigt werden müssen. Selbst bei Vorliegen hochwertiger Evidenz aus (internationalen) Studien verlangt die Entscheidung, ein Früherkennungsprogramm flächendeckend einzurichten, mutige Verallgemeinerungen auf die Funktionsfähigkeit des eigenen Gesundheitswesens. Das unterstreicht den Stellenwert einer begleitenden formativen Evaluation. Die Möglichkeiten dazu haben sich in Deutschland schrittweise verbessert. Zu wesentlichen Fragen, nicht zuletzt zum Stellenwert des Screenings auf Haut- und Prostatakrebs und der Früherkennung in der Primärversorgung, gibt es aber inter-

nationaler Klärungsbedarf, zu dem eine Beobachtung laufender Aktivitäten der GKV nur indirekt beitragen kann.

Welche Information darf ein individueller Versicherter daher in Sachen Screening erwarten, welche kann er bekommen und welche erhält er? Standards einer die Gesundheitskompetenz der Bürger fördernden Informationspolitik werden zunehmend erkennbar. Doch ohne Polemik ist festzustellen: Der Nutzen selbst gut begründeter und durchgeführter Screening-Programme wird i.d.R. in öffentlich verbreiteten Aufklärungsmaterialien und in der individuellen Arzt-Patienten-Kommunikation deutlich überschätzt. Das große Thema „paternalistische vs. versichertenorientierte Aufklärung“ ist kein Thema allein der Patientenrechte. Im Bereich der Früherkennung ist es auch eine Frage der Rechte der Bürger.

Early detection of diseases requires evidence of clinically relevant improvements in the natural course of the disease, generally provided by sufficiently powered randomized controlled trials. A simplistic equation of „earlier“ with „more effective“ or „more efficient“ is not tenable. Early detection loses relevance to the extent that primary prevention, e.g. by avoiding hazards, or curative interventions become possible.

This introduction presents terminology and principles which must be considered in preparation, implementation, and evaluation of screening programmes. Even if high grade (international) evidence is available the decision to implement a population-based screening programme requires some courageous extrapolations towards the functional capability of the targeted health system. This points to the significance of concomitant formative evaluation. The evaluative potential in the German health system has gradually improved. Major clarifications, however, are necessary in the international realm, e.g. concerning the effectiveness of screening for skin or prostate cancer and the working of multiphasic screening in primary care. Here monitoring of ongoing health services in Germany can contribute, albeit marginally.

Which information regarding screening measures may an individual expect, which could he possibly get and which will eventually be available to him? Standards of communication fostering health literacy of the public are becoming available. But it can be stated without being polemic that the benefits of screening programmes, even those based on evidence and conducted well, is grossly overvalued in publicly disseminated information materials and in doctor-patient-encounters. The major topic „paternalistic vs. shared decision making“ concerns not alone patients' rights. In screening, this is also a question of the citizens' rights.

1.1 Früherkennung in der gesetzlichen Krankenversicherung

Der Gesetzgeber definiert Umfang und Qualität der Krankenversorgung, die den Versicherten zusteht und die Krankenkassen und Leistungserbringer sicherzustellen haben, durch offene Rechtsbegriffe (ausreichend, wirtschaftlich, zweckmäßig, notwendig; § 12 SGB V). Zu Früherkennungsprogrammen dagegen hat er im vierten Abschnitt des Sozialgesetzbuchs V Vorgaben gemacht (§§ 25, 25a und 26 SGB V). Der Gemeinsame Bundesausschuss (G-BA) legt mit seiner Richtlinienkompetenz nach § 92 SGB V das Nähere über Früherkennungsmaßnahmen, organisierte Früherkennungsprogramme und die Qualitätssicherung der Früherkennungsuntersuchungen fest:

- Richtlinie zur Empfängnisregelung und zum Schwangerschaftsabbruch (vormals Sonstige-Hilfen-Richtlinie) mit der Vorgabe eines jährlichen genitalen Chlamydien-Screenings bei Frauen bis zum abgeschlossenen 25. Lebensjahr
- Richtlinien über die ärztliche Betreuung während der Schwangerschaft und nach der Entbindung (Mutterschafts-Richtlinien)
- Richtlinie über die Früherkennung von Krankheiten bei Kindern (Kinder-Richtlinie)
- Richtlinien über die Früherkennungsuntersuchungen auf Zahn-, Mund- und Kieferkrankheiten (Zahnärztliche Früherkennung)
- Richtlinie zur Jugendgesundheitsuntersuchung
- Richtlinie über die Gesundheitsuntersuchungen zur Früherkennung von Krankheiten (Gesundheitsuntersuchung-Richtlinie)
- Richtlinie über die Früherkennung von Krebserkrankungen (Krebsfrüherkennungsrichtlinie)
- Richtlinie für organisierte Krebsfrüherkennungsprogramme (oKFE-RL)
- Richtlinie Ultraschall-Screening auf Bauch-aortenaneurysmen (US-BAA-RL)

Das im Einheitlichen Bewertungsmaßstab geregelte sog. hausärztlich-geriatrische Basisassessment ist eine weitere komplexe Früherkennungsmaßnahme, allerdings ohne Richtlinie des G-BA. Insgesamt hält die gesetzliche Krankenversicherung also ein lebensbegleitendes Früherkennungsangebot vor (siehe dazu auch die Übersicht des Gemeinsamen Bundesausschusses: <https://www.g-ba.de/institution/themenschwerpunkte/frueherkennung/ueberblick/>).

Wie sind diese Regulierungen der sog. Selbstverwaltung im Gesundheitswesen in den Stand des Wissens um Früherkennungsprinzipien und -methoden einzuordnen? Ein Blick auf die Früherkennungsmaßnahmen anderer entwickelter Länder zeigt: Zielkrankheiten, Einladungsverfahren, eingesetzte Suchtests und die weitere Abklärungsdiagnostik, der Doku-

mentationsaufwand und – nicht leicht zu beurteilen – wahrscheinlich auch die erreichten Ergebnisse und damit die Wirtschaftlichkeit sind international unterschiedlich. So verzichtet z. B. selbst das sonnenverwöhnte Australien auf die Früherkennung von Hautkrebs (Cancer Council Australia's National Skin Cancer Committee 2014) und Schweizer Kantone schaffen die Einladung zum Mammografie-Screening wieder ab (Föry 2017), nachdem das unabhängige Swiss Medical Board sich wegen eines konstatierten ungünstigen Nutzen-Risiko-Verhältnisses und eines ungünstigen Kosten-Wirksamkeits-Verhältnisses gegen das systematische Mammografie-Screening ausgesprochen hatte (Fachgremium Swiss Medical Board 2013). Internationale Vergleiche verlangen also eine kritische Überprüfung der deutschen Programme mit dem Ziel einer Verbesserung von Angebot, Organisation und Durchführung.

Selbst innerhalb Deutschlands gibt es erkennbar unterschiedliche Bewertungen des Notwendigen, Zweckmäßigen und Ausreichenden. So haben Vertreter der Deutschen Gesellschaft für Allgemeinmedizin (DEGAM) abweichende Voten in die Leitlinien der Fachgesellschaften zur Früherkennung von Hautkrebs (Chenot u. Egidi 2015) und von Prostatakarzinomen (AWMF Leitlinienprogramm Onkologie 2018; Kötter 2016) eingebracht, unter anderem mit dem Ziel, potenzielle Früherkennungsteilnehmer vor ungerechtfertigten Erwartungen und einer ungünstigen Nutzen-Risiko-Bilanz zu schützen.

Aus der Perspektive der bevölkerungsbezogenen Versorgungsforschung gibt es – wie die einzelnen Beiträge dieses Versorgungs-Reports zeigen – zahlreiche offene Fragen zur evidenzbasierten Begründung laufender Früherkennungsprogramme, zu ihrer patientenorientierten Durchführung und zur methodischen Bewertung, damit auch zur weiteren Entwicklung. Zum Beispiel fehlt dem im hausärztlich-geriatrischen Basisassessment enthaltenen Demenz-Screening die Evidenzbasis (s. Kap. 11). Darüber hinaus können praktisch täglich neue

Erkenntnisse der klinischen und der biomedizinischen Forschung potenziell auch zur Früherkennung von Krankheiten beitragen, weil sie ein besseres Verständnis biologischer Mechanismen, verbesserte diagnostische Verfahren oder wirksamere therapeutische Strategien zum Gegenstand haben. Diese Potenziale bedürfen allerdings erst einer aufwändigen kontrollierten Erprobung in symptomfreien Versichertengruppen, bevor sie flächendeckend und qualitätsgesichert umgesetzt werden können. Derartige grundlegende Erprobungen, insbesondere randomisierte kontrollierte Studien (RCTs, Goldstandard-Studien), fehlen für den Beginn des deutschen Krebsfrüherkennungsprogramms, was bis heute zu Unsicherheiten seiner Bewertung beiträgt und bei in Aussicht genommenen Modifikationen und Ergänzungen zu versorgungswissenschaftlichen, ethischen und rechtlichen Diskursen beiträgt.

Die folgenden Abschnitte dieses Beitrags sind wie folgt gegliedert: Zunächst werden grundsätzliche Prinzipien und Begriffe der Früherkennung geklärt und krankheitsübergreifend methodische Grundlagen der Bewertung von Früherkennungsprogrammen erläutert. Dabei wird unterschieden zwischen der Perspektive eines Individuums, das vor der Frage steht, an einem Früherkennungsprogramm für symptomlose Personen teilzunehmen (Individualperspektive), und der Perspektive der Entscheidungsverantwortlichen für Strukturen, Prozesse, Qualitätssicherung und Finanzierung im Gesundheitswesen (gesellschaftliche Perspektive). Weiter wird unterschieden zwischen der Perspektive vor Einführung eines systematischen Screenings (ex ante) und der Perspektive der Bewertung eines laufenden Screening-Programms (ex post). In der Situation ex ante sind die individuelle und die gesellschaftliche Perspektive strukturell nicht zu trennen. Sie können sich aber in der Bewertung von Risiken und Nutzen einzelner Gesundheitszustände unterscheiden. Die Abklärung von Verdachtsfällen und selbst invasive Eingriffe werden vorhersehbar Untergruppen der Teilnehmer betreffen, die

selber davon keinen Nutzen, sondern nur Nachteile haben werden. Daher muss ein symptomloser potenzieller Teilnehmer schon vor einer Entscheidung über die Teilnahme an einem wohldefinierten Früherkennungsprogramm umfassend über alle Folgen seiner Teilnahme aufgeklärt werden. Abschnitt 1.3 untersucht, wie Bürger in die Lage versetzt werden können, solche Entscheidungen selbstbestimmt, d.h. ggf. auch gegen den Hauptstrom der gesellschaftlichen Meinung und den Paternalismus in der Medizin, zu treffen.

1.2 Prinzipien und Begriffe

Früherkennungsuntersuchungen, auch Filteruntersuchungen oder im internationalen Sprachgebrauch Screening genannt, richten sich definitionsgemäß an symptomlose Zielpersonen. Sobald ein Versicherter sich zur Abklärung von (selbst bemerkten) Symptomen an einen Arzt wendet, handelt es sich um ein Problem der Krankenversorgung. Dieser Unterschied ist von praktischer und ethischer Bedeutung. Denn wenn ein Versicherter ärztlichen Rat und Hilfe sucht, hat er die Initiative und muss nach Aufklärung (§ 630e BGB) selbst entscheiden, ob er die ärztlicherseits vorgeschlagenen Optionen akzeptiert oder nicht. Wenn aber ein ganzes Gesundheitssystem, eine Krankenversicherung oder ein Versorger symptomlosen Personen Untersuchungsverfahren aktiv anbietet oder die Inanspruchnahme nahelegt, müssen die Anbieter vorher allgemeinverständlich darlegen, dass für jeden potenziellen Teilnehmer und für die Zielgruppe insgesamt eine hinreichende Wahrscheinlichkeit für einen nicht trivialen medizinischen (Netto-) Nutzen besteht und welche unerwünschten Wirkungen gegen eine Inanspruchnahme sprechen.

An dieser Beweislast ändert sich nichts dadurch, dass das Ergebnis eines Screening-Tests (noch) nicht diagnostisch belastbar ist. Sein Zweck ist, die Wahrscheinlichkeit für das Vor-

liegen einer ernsthaften und wirksam behandelbaren Zielkrankheit so weit zu erhöhen, dass Abklärungsuntersuchungen gerechtfertigt sind und vorverlegte Therapiemaßnahmen für den Teilnehmer eine bessere Prognose erwarten lassen als die übliche Versorgung ohne Screening.¹

Früherkennung wird herkömmlich als „sekundäre Prävention“ der Primärprävention an die Seite gestellt. Primärprävention soll das Auftreten von Krankheiten verhindern, weniger wahrscheinlich machen oder verzögern (Walter et al. 2012, S. 196). Früherkennung verhütet keine Krankheit, sondern soll Langzeitergebnisse der Versorgung verbessern. Insofern ist der Begriff „sekundäre Prävention“ potenziell missverständlich. Früherkennung steht der Rehabilitation näher als der Primärprävention.

Wenn wir abgekürzt von „Screening-Programmen“ sprechen, ist ein mehrstufiges Vorgehen gemeint:

- a) die Auswahl einer geeigneten Zielkrankheit, eines zu ihrer Erkennung geeigneten Früherkennungstests und einer umschriebenen symptomlosen Zielgruppe
- b) die Einladung zur Teilnahme an diesem Früherkennungsprogramm
- c) die aufgeklärte Entscheidung einer symptomlosen Person zur Teilnahme
- d) die Durchführung notwendiger Folgeuntersuchungen zur Abklärung eines auffälligen Früherkennungstests
- e) die indizierte Frühintervention
- f) die begleitende Dokumentation, Erfolgssicherung und Bewertung (Evaluation) des Vorgehens

1 Das Oxford Textbook of Medicine definiert: „Medical screening is the systematic application of a test or inquiry to identify individuals at sufficient risk of a specific disorder to benefit from further investigation or direct preventive action (these individuals not having sought medical attention on account of symptoms of that disorder). Key to this definition is that the early detection of disease is not an end in itself; bringing forward a diagnosis without altering the prognosis is useless and may be harmful.“ (<http://oxfordmedicine.com/view/10.1093/med/9780199204854.001.1/med-9780199204854-chapter-030302>)

Früherkennung ist also kein Selbstzweck („je früher desto besser“), sondern erst gerechtfertigt, wenn sich am Ende eines komplexen Prozesses die gesundheitliche Perspektive von Menschen spürbar verbessern lässt. Der Nachweis einer solchen Verbesserung setzt kontrollierte Langzeitstudien voraus, die schwer zu organisieren und durchzuführen sind und in Deutschland fehlen (siehe dazu die einzelnen Beiträge in diesem Band). Die genannten Stufen des Vorgehens strukturieren die vorbereitende und die nachgehende Bewertung eines Früherkennungsprogramms. Sie werden im Folgenden vertieft.

Zu a) die Auswahl einer geeigneten Zielkrankheit, eines zu ihrer Erkennung geeigneten Früherkennungstests und einer umschriebenen symptomlosen Zielgruppe

Unter den Zielkrankheiten laufender Screening-Programme sind die Karzinome von Gebärmutterhals, Brust und Dickdarm besonders prominent. Eine Teilnahme daran wird u.a. vom „European Code against Cancer“ unterstützt (Armaroli et al. 2015). Haut- und Prostatakrebs finden weniger Unterstützung. Die systematische Suche nach unentdeckten Neuroblastomen im Kindesalter hat sich nach Evaluation als nicht wirksam erwiesen (Schilling et al. 2002; Woods et al. 2002). Weitere Zielkrankheiten, darunter fachlich umstrittene oder nicht empfehlenswerte, sind Gegenstand weiterer Kapitel dieses Bandes.

Als Geburtsstunde heutiger Gesunden-Untersuchung kann die Suchen nach psychiatrischen Erkrankungen bei Musterungen der USA-Armee im Jahr 1917 gelten (Morabia u. Zhang 2004). Musterungsuntersuchungen hatte es natürlich vorher schon gegeben, doch tauchte in diesem Rahmen erstmals die Frage nach der Validität der Suchverfahren, den falsch positiven und falsch negativen Probanden auf. Ziel war allerdings nicht Frühtherapie, sondern eine Eignungsuntersuchung. Mit der zunehmenden

Verfügbarkeit von Laborautomaten richteten sich Hoffnungen auf ein multimodales Screening mit Kombination von Anamnese, körperlicher Untersuchung, Laborwerten und Gesundheitsberatung. Sie haben sich nicht erfüllt (z.B. South East London Screening Study 1977; Caley et al. 2014; Jørgensen et al. 2014), der deutsche Check-up 35 ist umstritten und revisionsbedürftig (siehe auch Mühlhauser 2014 und Kap. 9). Der Einsatz von EKG und Belastungs-EKG zur Früherkennung von Herzerkrankungen will selbst bei Menschen mit deutlich erhöhtem Risiko für eine koronare Herzerkrankung gut überlegt sein (Bhatia u. Dorian 2018). Auch übertragbare Krankheiten sind Gegenstand systematischer Suchtests, z.B. Syphilis, HIV/AIDS, Röteln oder MRSA. Screening darauf unterscheidet sich von dem auf andere Zielkrankheiten ethisch und ökonomisch dadurch, dass es unter Umständen Nutzen für Dritte stiften kann, die bisher nicht betroffen sind. Trotzdem wird in Deutschland auf eine Röntgen-Reihenuntersuchung der Wohnbevölkerung auf Tuberkulose, viele Jahre Teil öffentlicher Gesundheitspflege, verzichtet – Folge einer Risiko-Nutzen-Abwägung, die bei Migranten aus Ländern mit hohem Vorkommen an Tbc anders ausfällt.

Wilson und Jungner (1968, S. 26–27) haben die weitreichenden Vorbedingungen dieses definitorischen Schritts wie folgt zusammengefasst (vgl. auch § 25 Abs. 3 SGB V):

1. The condition sought should be an important health problem.
2. There should be an accepted treatment for patients with recognised disease.
3. Facilities for diagnosis and treatment should be available.
4. There should be a recognisable latent or early symptomatic stage.
5. There should be a suitable test or examination.
6. The test should be acceptable to the population.
7. The natural history of the condition, including development from latent to declared disease, should be adequately understood.

8. There should be an agreed policy on whom to treat as patients.
9. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
10. Case-finding should be a continuing process and not a 'once and for all' project.

Für die Auswahl eines zur Früherkennung geeigneten Tests sind neben Praktikabilität und Kosten vor allem Parameter seiner Eignung (Gültigkeit, Validität) relevant: Sensitivität und Spezifität. Die Sensitivität Se gibt die Wahrscheinlichkeit an, dass eine vorhandene Zielkrankheit tatsächlich erkannt wird. Die Spezifität Sp gibt die Wahrscheinlichkeit an, dass eine Person ohne Zielkrankheit als unauffällig eingestuft wird.

Das Komplement der Sensitivität ($1-Se$), die Wahrscheinlichkeit, eine vorhandene Ziel-läsion zu übersehen, ist ein bedeutsames Problem für Früherkennungsteilnehmer und Leistungserbringer, das möglichst klein gehalten werden sollte. Aber auch die Unspezifität ($1-Sp$), die Wahrscheinlichkeit eines falsch positiven Testergebnisses bei Teilnehmern ohne Zielläsion, führt zu unerwünschten medizinischen Folgen. Sie sind nicht zuletzt deswegen bedeutsam, weil die Teilnehmer ohne Zielläsion die große Mehrheit aller Teilnehmer stellen. Nach 10 Jahren jährlicher Screening-Mammografie wird z.B. mehr als die Hälfte aller teilnehmenden Frauen wenigstens eine falsch positive Nachuntersuchung erhalten, 7% bis 9% eine falsch positive Biopsieempfehlung. Ein Screening nur alle zwei Jahre reduziert diese kumulative Wahrscheinlichkeit, erhöht aber die Wahrscheinlichkeit einer Diagnose erst in einem späten Stadium (Hubbard et al. 2011).

Sensitivität und Spezifität eines Tests hängen typischerweise gegenläufig miteinander zusammen. Die Wahl des Grenzwertes, von dem ab ein Testergebnis als auffällig beurteilt

wird und deswegen zu Konsequenzen führt, beeinflusst beide Testgüte-Parameter, damit Aufwand, Kosten und Erfolg des ganzen Programms. Das „richtige“ Verhältnis von Sensitivität und Spezifität hängt nicht zuletzt davon ab, wie die Regulatoren oder die Zielgruppe das Verhältnis des Nutzens richtig erkannter Zielkrankheiten zum Schaden durch Testaufwand, Abklärungsaufwand und falsch positive, d.h. schädliche Folgeinterventionen bewerten. Ein schmerzhaftes Beispiel ist die notwendige Inkaufnahme des Todes gesunder Föten bei der systematischen Früherkennung von Fehlbildungen in der Schwangerschaft (Felder et al. 2003; Felder u. Robra 2006). Erstmals zur Verwendung immunologischer Tests auf Blut im Stuhl (iFOBT) hat der Gemeinsame Bundesausschuss explizit Mindestgrenzen für Sensitivität und Spezifität vorgegeben (Bundesministerium für Gesundheit 2016). Für die Entdeckung von Darmkrebsen und fortgeschrittenen Polypen muss in nur einer Stuhlprobe eine Sensitivität von mindestens 25% bei einer Spezifität von mindestens 90% erreicht werden. Referenztest ist das Ergebnis einer Koloskopie. Mit diesen zurückhaltenden Mindestvorgaben offenbart der G-BA implizit eine Güterabwägung von Nutzen und Schäden im Rahmen des Darmkrebsfrüherkennungsprogramms.

Ein weiteres, unvermeidbares Problem ist als Überdiagnostik bekannt. Es bedeutet, dass in Bevölkerungen mit Früherkennungsprogramm für einen bestimmten Anteil der entdeckten und dann auch behandelten Fälle die Zielkrankheit zu Lebzeiten nicht mehr klinisch manifest oder gar lebenszeitbegrenzend geworden wäre. Es handelt sich um ein bevölkerungsbezogenes Konzept, d.h. individualmedizinisch tätige Kliniker und Pathologen können es nicht erkennen – sie sehen und behandeln einen nachgewiesenen Fall. Überdiagnostik trifft vor allem bei Zielkrankheiten zu, die bevorzugt im höheren Alter auftreten. Sie kann in Studien quantifiziert werden, die Neuerkrankungsziffern und Sterblichkeit in Bevölkerungen mit und ohne Möglichkeit zur Teil-

Tab. 1 Erwarteter Nutzen eines Screening-Programms – Ergebnis des Screening-Tests, Vorliegen der Krankheit und Bewertungen des Nettonutzens der vier Zustände

		Krankheit	
		ja	nein
Test	positiv	$Pr \cdot Se \cdot NU_{(T+K+)}$	$(1 - Pr) \cdot (1 - Sp) \cdot NU_{(T+K-)}$
	negativ	$Pr \cdot (1 - Se) \cdot NU_{(T-K+)}$	$(1 - Pr) \cdot Sp \cdot NU_{(T-K-)}$

Pr ist die Prävalenz der Zielläsionen, bei einem einzelnen Teilnehmer sein Krankheitsrisiko a priori; Se Sensitivität und Sp Spezifität sind Kennwerte der Testgüte; NU steht für Nettonutzen (net utility), der Nettonutzen der vier Zustände wird durch Indizes unterschieden: K+ bzw. K- signalisieren den wahren Gesundheitszustand, T+ bzw. T- das Testergebnis

nahme am Früherkennungsprogramm miteinander vergleichen. Der Anteil überdiagnostizierter Fälle an allen Fällen ist z.B. hoch beim Screening auf Prostatakarzinom (Robra et al. 2013), aber auch beim Brustkrebs-Screening nicht zu vernachlässigen (Autier et al. 2017) und reduziert natürlich den Nutzen der Früherkennung. Jedes Absenken diagnostischer Schwellen verstärkt Überdiagnostik (Welch et al. 2011). Forschung zu aussagefähigen Progressionsmarkern könnte helfen, unnötige Interventionen zu reduzieren.

Falsch positive und falsch negative Testergebnisse und Überdiagnostik sind nicht nur Herausforderungen der individuellen Teilnahmeaufklärung, sondern Probleme für die Konzeption eines Früherkennungsprogramms ex ante. Denn das Angebot eines Früherkennungsprogramms bedeutet de facto die Einladung zur Teilnahme an einer Lotterie: Einige Teilnehmer werden einen Nutzen realisieren können, andere werden Schäden davontragen. Es handelt sich also um einen Verstoß gegen das Gebot, nicht zu schaden (Non-Malefizien). Die hinter dem Programmangebot stehende Güterabwägung trifft der Regulator, noch bevor ein potenzieller Teilnehmer individuelle Präferenzen geltend machen kann.

Tabelle 1 illustriert vier elementare Zustände im Screening: Der Screening-Test kann positiv oder negativ ausfallen, die Krankheit liegt vor oder nicht. Zunächst vernachlässigen wir die dritten Multiplikatoren, die Nutzenparameter.

Die Wahrscheinlichkeit, einen der vier Zustände zu erreichen, ist eine Funktion von Prävalenz und Testgüte. Die Wahrscheinlichkeit eines Teilnehmers, sich im Feld links oben zu finden, ist gegeben durch das Produkt aus Prävalenz und Sensitivität. Das Feld darunter quantifiziert die Wahrscheinlichkeit eines falsch negativen Ausgangs, das Feld rechts oben die Wahrscheinlichkeit eines falsch positiven. Das Feld rechts unten beschreibt die Wahrscheinlichkeit, zur Gruppe der negativ getesteten Gesunden zu gehören. Das ist typischerweise die größte Gruppe.

Der dritte Parameter NU in jeder Zelle steht für den Nettonutzen, der jenem Zustand zugeschrieben wird. $NU_{(T+K+)}$ steht für den Nettonutzen richtig erkannter Kranker. Nettonutzen heißt, dass der medizinisch bewertete Nutzen dieses Zustands, z.B. ein Gewinn an Lebensdauer, schon um unerwünschte Effekte vermindert wurde, z.B. um einen Verlust an Lebensqualität relativ zum Status vor Testung. $NU_{(T+K+)}$ muss positiv sein, sonst macht das Screening keinen Sinn. $NU_{(T+K-)}$ steht für den Nettonutzen, mit dem ein falsch positiver Testausgang bewertet wird. Er ist typischerweise negativ. Je nach Perspektive kann die Nutzenbewertung zusätzlich direkte und indirekte Kosten einbeziehen und Nutzenströme über die Zeit diskontieren.

Die Summe der vier bewerteten Zustände ist der erwartete Gesamtnutzen von Früherkennungsprogrammen auf der Ebene der symp-

tomlosen Zielbevölkerung, aber auch für ein symptomloses Individuum in Unkenntnis der es erwartenden Zustände. Die individuelle und die gesellschaftliche Bilanz gehen also von dem gleichen (evidenzgestützten) Mengengerüst aus (Prävalenz bzw. Risiko, Testgüte). Auf beiden Ebenen muss die Nutzenbilanz positiv sein, um ein Screening-Programm überhaupt anbieten zu können bzw. eine Teilnahme individuell in Erwägung zu ziehen. Allerdings kann sich die durch den Regulator unterstellte „gesellschaftliche“ Bewertung der Zustände von einer individuellen Bewertung unterscheiden. Das unterstreicht die Notwendigkeit der individuell aufgeklärten Zustimmung auch bei „programmierten“ Screening-Angeboten.

Ein Individuum kann für sich selber positive und negative Effekte miteinander verrechnen. Solche individuellen Abwägungen verschieden dimensionierter Effekte finden überall im Gesundheitswesen statt, z. B. wenn die Einnahme schmerzlindernder und entzündungshemmender nichtsteroidaler Antirheumatika (NSAR) gegen eine damit verbundene erhöhte Blutungsneigung abgewogen werden muss. Damit die Abwägung auch beim Screening möglich ist, braucht das Individuum allerdings nicht die abstrakte Erwartungsnutzentheorie der Tabelle 1, sondern für ihn verständliche Quantifizierungen der zu erwartenden Versorgungspfade (z. B. Raftery u. Chorozoglou 2011; Welch u. Passow 2014): Welchen Aufwand bedeutet z. B. eine Darmspiegelung für ihn, wie viele von 1.000 Teilnehmern werden zur Abklärung auffälliger Befunde einbestellt, was kann dabei passieren, welche Behandlungsoptionen stehen zur Verfügung? Und wie viele von 1.000 Teilnehmern werden mittel- und langfristig welchen Nutzen erreichen („number needed to screen“), wie viele werden aber auf welche Weise falsch positiv versorgt oder überdiagnostiziert?

Individuelle Zustimmung zur Teilnahme löst jedoch auf der gesellschaftlichen Ebene das ethische Dilemma nicht, dass ein Programmangebot im Interesse eines positiven Nettonutzens eines Teils der Teilnehmer einem anderen

Teil der Teilnehmer einen negativen Nettonutzen, d. h. Schaden, zumutet. Regulatoren schwächen dieses Dilemma für sich zunehmend dadurch ab, dass sie in einem mittleren Risikobereich wohlstrukturierte Programmvorgaben vermeiden, sondern die Entscheidung über die Früherkennungsteilnahme explizit den Versicherten und ihren Ärzten anheimstellen, die das Vorgehen risiko- und präferenzabhängig kontextuieren, wie sie es in der klinisch-kurativen Individualmedizin kennen („discretionary care“). Bei Früherkennungsmaßnahmen im Niedrigprävalenzbereich und mit großer Unsicherheit der Nutzenbewertung müssen sie jedoch kritischer entscheiden als bei akut-dringlichen „kurativen“ Maßnahmen, d. h. Ärzte wie Patienten müssen die komplexe Evidenzlage angemessen berücksichtigen können. Ärztliche Einwilligungsaufklärung muss im Übrigen in einer Situation mit wenig Handlungsdruck noch gründlicher sein als in akuten klinischen Situationen (Windus et al. 1992).

Eine andere Folge der Entscheidungskomplexität sind „politische Ersatzvornahmen“. Beim Mammografie-Screening kommt dies besonders deutlich zum Ausdruck: Hier entschied das Parlament nach öffentlicher Anhörung von Sachverständigen einmütig, dass es einen großen Screening-Nutzen gebe und Deutschland gegenüber den europäischen Leitlinien im Verzug sei (Deutscher Bundestag 2002).

So wie Mengengerüst (geringe Prävalenz der Zielkrankheiten!), Testgüte und Nutzen-Relationen von Früherkennungsprogrammen typischerweise ausfallen, ist ihr gesellschaftlicher Gesamtnutzen sensitiv gegenüber der Höhe der direkten Durchführungskosten und den unerwünschten (falsch positiven) Wirkungen für die Gesunden. Der in der Werbung für das Programm herausgestellte Nutzen für früh erkannte Fälle beeinflusst das Gesamtergebnis eher wenig, da diese Fälle zwar großen Nutzen gewinnen können, aber selten sind. Dies spricht dafür, ein Screening-Programm schonend und kostensensibel zu organisieren. Für

die Versicherten der gesetzlichen Krankenversicherung sind die direkten Kosten zudem durch das Sachleistungsprinzip auf einen Wert von Null heruntersubventioniert. Für Ärzte bedeuten Früherkennungsprogramme wenig riskante Einnahmen. Diese Aspekte tragen zu Bewertungsunterschieden zwischen Systemebene und Individualebene bei, obwohl eine Früherkennungsteilnahme auch für individuelle Teilnehmer mit Opportunitätskosten verbunden ist.

Als weiteres Problem der Nutzenbewertung können bei zunehmend perfektionierten Untersuchungsmethoden deutlich häufiger als früher Frühstadien ans Licht gebracht werden, von denen unbekannt ist, wann oder ob sie in klinisch relevante Stadien umschlagen werden und ob sie mit den für fortgeschrittene Stadien entwickelten Therapieschemata wirksam oder schonend genug behandelt werden können. Die Verschiebung im Spektrum entdeckter Stadien bringt also neue Krankheitsentitäten hervor, die aus eigenem Recht prognostisch und mit kontrollierten Therapiestudien eingeordnet werden müssen. Das traf z.B. auf das sog. ductale Carcinoma in situ (DCIS) der Brust zu, dessen Häufigkeit bei systematischer Suche per Mammografie nennenswert zunahm (Jatoi u. Baum 1995; Correa et al. 2010). Das trifft auch für das lokal begrenzte Prostatakarzinom zu, für dessen Management kontrollierte Überwachung hinreichen kann (Bill-Axelsson et al. 2014; Hamdy et al. 2016; Wilt et al. 2017; Wellek et al. 2017). Damit wird jedoch die Grundidee der Früherkennung, die Vorverlegung des Diagnosezeitpunktes, hinfällig.

Zu b) die Einladung zur Teilnahme an diesem Früherkennungsprogramm

Als Zielgruppe für die Einladung zum Test kommt nur eine Versichertengruppe in Betracht, für die der Test akzeptabel ist (Tests, die auf Schwangerschaftsunterbrechung hinauslaufen, sind z.B. nicht in allen Bevölkerungs-

gruppen akzeptabel), für die die Zielkrankheit hinreichend häufig oder belastend ist² (das kann auf eine untere Altersgrenze zum Screening oder eine andere Abgrenzung einer Gruppe hohen Risikos hinauslaufen) und für die ein möglicher Früherkennungsnutzen sich noch zu Lebzeiten manifestieren kann (das kann auf eine obere Altersgrenze zum Screening hinauslaufen, die mit Altersrationierung von Leistungen nichts zu tun hat).

Das Angebot zur Teilnahme an einem positiv vorbewerteten Früherkennungsprogramm kann mehr oder weniger nachdrücklich vermittelt werden, von einer kaum kommunizierten „Anspruchsberechtigung“ bis hin zu systematischen Einladungsmodellen, wie sie in Deutschland für das mammografische Screening auf Brustkrebs eingerichtet und für das Darmkrebs- und Gebärmutterhals-Screening beschlossen wurden. Schon die Erfahrungen mit dem Gebärmutterhals-Screening haben gezeigt, dass es wirksamer ist, viele Anspruchsberechtigte mindestens einmal zu screenen als wenige repetitiv. Dies spricht dafür, nicht allein auf die Teilnahme immer derselben selbstmotivierten Freiwilligen zu setzen, sondern nachdrücklich für breite Inanspruchnahme zu werben, gerade auch in schwer erreichbaren Risikogruppen. Allerdings befinden sich die Gestaltungsverantwortlichen dabei in einem Dilemma. Je nachdrücklicher die Werbung, desto mehr wird die Selbstbestimmung der Versicherten durch Paternalismus übersteuert, dem man zwar die gute Absicht nicht absprechen muss, der aber wie oben gezeigt der individuellen Aufklärungsnotwendigkeit nicht gerecht wird und individuelle Präferenz nicht mehr respektiert.

Auch eine hundertprozentige Beteiligung am Screening kann die Gesamtsterblichkeit nach einer Modellrechnung nur gering senken, beim mammografischen Screening um 1,8%, bei den

² In § 25 Abs. 1 SGB V heißt es „bevölkerungsmedizinisch bedeutsame Krankheiten“.

anderen Krebslokalisationen deutlich weniger (Stang u. Jöckel 2018). Deswegen und wegen der Verschlechterung der Wirtschaftlichkeit durch ein ungünstiges Verhältnis von Vorhaltekosten zu gesundheitlichen Effekten diskutiert man eine Teilnahme-Untergrenze, ab der ein Screening-Programm für Bevölkerung und Gesundheitswesen sinnvoll wird. Bei Einführung des mammografischen Screenings war deswegen unter Bezug auf eine EU-Leitlinie eine (zweijährige) Teilnehmerate von mindestens 70% als Randbedingung für die Ein- und Fortführung des Programms angestrebt worden. Im letztverfügbaren Jahresbericht der Kooperationsgemeinschaft Mammografie lag die Einladungsrate der Anspruchsberechtigten 2015 bei praktisch 100%, die Teilnehmerate bei knapp 52% (überdurchschnittlich in den neuen Bundesländern) (Kooperationsgemeinschaft Mammografie 2017).

Die Politik hat nach breiter (fach-)öffentlicher Debatte in dieser Hinsicht einen Paradigmenwechsel vollzogen (s. Kap. 6). Vorrang vor der Verbesserung der Teilnehmeraten als intermediärer Zielgröße des Programms hat die verbesserte Information der Bürgerinnen und Bürger über Nutzen und Risiken der Krebsfrüherkennung mit dem Ziel einer informierten Entscheidung. Damit ist das bislang weithin konsenterte „Werben“ für die Teilnahme im Kern obsolet. Wird eine informierte Entscheidung im Alltag gelebt, können die Teilnehmeraten sogar sinken, was unbedingt zu respektieren wäre. Es steht zu vermuten, dass dieser Wandel noch längere Zeiträume in Anspruch nehmen wird.

Ein geplantes Screening-Programm gibt Teilnehmintervalle vor (siehe z.B. die komplexen Vorgaben zum Darmkrebs-Screening in Teil II, § 3 der o.g. oKFE-RL). Allgemein gesagt ist die Wiederholungsfrequenz des Screenings abhängig von der Verteilung der Wachstumsgeschwindigkeiten der Zielläsionen und der methodenabhängigen (durchschnittlichen) Vorverlegung des Diagnosezeitpunkts in der Früherkennung („lead time“, Hill et al. 1984). Das in Deutschland zu Beginn der organisierten Früherkennung für alle Zielläsionen und Maßnah-

men einheitlich festgelegte jährliche Intervall ist offensichtlich zu schematisch und wird für jüngere Screening-Verfahren wie die Mammografie und die Koloskopie, zukünftig auch für das Screening auf Gebärmutterhalskrebs und den Check-up 35, ausdifferenziert. Internationale Empfehlungen halten seit Jahren beim Cervix-Screening ein 3- bis 5-jähriges Intervall für ausreichend, wenn die Qualität der Abstriche gut ist; bei mehrfach negativ gescreenten Teilnehmerinnen kann auf weitere Untersuchungen verzichtet werden (Hakama et al. 1986; Curry et al. 2018). Die Überarbeitung des Früherkennungsprogramms auf Cervixkarzinome hinsichtlich Altersgrenzen und Intervallen wird zudem wegen der Möglichkeit, HPV-Infektionen nach Risiko zu differenzieren, und der zusätzlich verfügbaren, gerade auf Jungen ausgetesteten HPV-Impfung dringlich. Ein Konzept des G-BA für eine neu organisierte Früherkennung auf Gebärmutterhalskrebs³ und ein Evaluationskonzept des IQTiG (2017) liegen vor.

Zu c) die aufgeklärte Entscheidung einer symptomlosen Person zur Teilnahme

Diese Entscheidung können Versichertengemeinschaft und Individuum nur treffen, wenn gute wissenschaftliche Evidenz zu Organisation, Mengengerüsten, richtig positiven und falsch positiven Folgemaßnahmen und positiven wie negativen gesundheitlichen Langzeiteffekten erarbeitet wurde und mit Hinweisen auf ihre Aussagefähigkeit verständlich aufbereitet zur Verfügung steht. Die methodischen Anforderungen an vertrauenswürdige Screening-Programme sind also hoch. Allgemein sind die Qualität von Informationsmaterialien und Aufklärungsgesprächen und das Wissen der Bürger jedoch unzureichend (Dierks u. Schmacke 2014; Seidel et al. 2014; s. Kap. 5). Da

³ Pressemitteilung vom 14. April 2018: https://www.g-ba.de/downloads/34-215-742/14_2018-04-17_KFE-RL_Zervixkarzinom-Screening.pdf

jedes neue diagnostische oder therapeutische Verfahren die Risiko-Nutzen-Balance der Früherkennung verschiebt, müssen Informationen periodisch aktualisiert werden.⁴

Wenn ein Individuum wie gesehen für seine Entscheidung, eine Früherkennungsmaßnahme in Anspruch zu nehmen, vor dem Schleier der Ungewissheit über die für ihn folgende Ereigniskette gerade dieselbe quantitative Evidenz wie die Versichertengemeinschaft als Ganze braucht, jedoch die nachfolgenden Maßnahmen und gesundheitlichen Resultate anders bewerten kann als die Versichertengemeinschaft, braucht es zusätzlich Unterstützung zur Klärung seiner individuellen Präferenzen, jedenfalls solange es Gründe sehen kann, sich nicht dem gesellschaftlichen Hauptstrom anzuschließen.

Ein solcher Grund könnte z.B. eine einschlägige genetische Belastung in der Familie sein, die oft mit entsprechenden Krankheitserfahrungen bei Angehörigen verbunden ist. Personen mit solchen Zusatzrisiken können präferenzbedingt leichter zu Früherkennungsmaßnahmen und intensiveren Interventionen motivierbar sein als Versicherte mit durchschnittlichem Risiko. Ob sie daraus allerdings mehr Nutzen realisieren können als eine Person durchschnittlichen Risikos, ist eine andere, gesondert zu untersuchende Frage.

Zu d) die Durchführung notwendiger Folgeuntersuchungen zur Abklärung eines auffälligen Früherkennungstests

Auf welche Weise auffällige Screening-Tests abgeklärt werden, ist untrennbarer Bestandteil jeder Früherkennungsstrategie. Ob z.B. die Spiegeleinstellung des Gebärmuttermundes schon als Teil des zytologischen Screenings unter Sicht

eines Mikroskops (Kolposkop) durchgeführt wird (Beller et al. 1982) oder ob eine Kolposkopie erst nach zytologischer Auffälligkeit zur Kontrolluntersuchung gehört, führt zu unterschiedlichen Ergebnissen. Bei welchen Befunden im Brustkrebs-Screening zusätzlich eine Ultraschalluntersuchung angezeigt ist, ist Gegenstand von Kosten-Effektivitäts-Analysen (z.B. Sprague et al. 2015). Mit welcher Technik der Prostata nach einem auffälligen PSA-Test Gewebe entnommen wird, hat Auswirkungen auf Entdeckungsraten (s. Kap. 10). Mit welcher Verzögerung eine empfohlene Biopsie durchgeführt wird, ist Parameter der Qualitätssicherung der Früherkennung (Meyer et al. 1997). Systematisch eingeholte Zweitmeinungen für auffällige Pathologiebefunde können die Behandlungssindikation präzisieren und sogar Kosten sparen (Tosteson et al. 2018). Da ständig diagnostische Fortschritte möglich werden, ist der Diskurs um die beste (wirksamste, wirtschaftlichste) Abklärungsstrategie im Screening grundsätzlich offen. Fortschritte der Abklärungsstrategie müssen zu einer Neubewertung vorangehender Schritte der Früherkennung führen.

Da in der GKV Abklärungsmaßnahmen normalerweise „kurativ“ abgerechnet werden, ist bei einer evaluativen Auswertung von Abrechnungstatistiken nicht immer klar, ob eine Maßnahme Folge einer Früherkennungsuntersuchung oder einer Abklärung von Symptomen gewesen ist. Der Bewertungsausschuss hat daher für Früherkennungsleistungen je spezielle Gebührenordnungspositionen im Einheitlichen Bewertungsmaßstab festgelegt (z.B. GOP 01733 für zytologische Untersuchung nach KFE-RL, 01741 für Koloskopie nach KFE-RL oder 01748 für Sonografie nach US-BAA-RL). Krebsregister bemühen sich um eine genaue Zuordnung der Maßnahmen.

Zu e) die indizierte Frühintervention

Seit langem ist klar, dass Erfolgsstatistiken der klinisch-kurativen Individualmedizin für die

⁴ Neben den Krankenkassen hat das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) den Auftrag, über Vorteile und Nachteile von Untersuchungs- und Behandlungsverfahren zu informieren (www.iqwig.de). Auch die DEGAM stellt Entscheidungshilfen zur Verfügung (DEGAM 2018; Kötter 2016).