

Datum: 20.07.2023  
Aktualisierung: März 2026

# Problem der Mehrfach-Diag- nostik

am Beispiel der Brustkrebsfrüher-  
kennung



**Dr. Joachim C. Zink**

Facharzt für Innere Medizin  
Facharzt für Allgemeinmedizin  
Medizinische Informatik

# 1 Einleitung

Die folgenden Überlegungen befassen sich nicht mit der grundsätzlichen Bewertung des Mammographie-Screenings, sondern mit den statistischen Konsequenzen mehrstufiger diagnostischer Verfahren. Ziel ist es, anhand eines vereinfachten Modells zu zeigen, welche Effekte sich aus der Kombination mehrerer diagnostischer Tests unter Berücksichtigung ihrer Sensitivität, Spezifität und der zugrunde liegenden Prävalenz ergeben können.

Die Analyse beschränkt sich auf die Kombination von Mammographie und Magnetresonanztomographie (MRT).

Das Mammakarzinom ist mit jährlich etwa 75 000 Neuerkrankungen (2023: 75 090 Frauen, 770 Männer) die häufigste Krebserkrankung bei Frauen in Deutschland. Auf Basis aktueller Inzidenzdaten wird geschätzt, dass etwa eine von acht Frauen im Laufe ihres Lebens an Brustkrebs erkranken wird. Trotz erheblicher Fortschritte in Diagnostik und Therapie sterben jährlich weiterhin rund 18 000 Frauen an den Folgen dieser Erkrankung.

Zur bildgebenden Diagnostik des Mammakarzinoms stehen im Wesentlichen drei Verfahren zur Verfügung:

1. Mammographie
2. Sonographie der Brust
3. Magnetresonanztomographie (MRT)

Im Jahr 2005 wurde in Deutschland ein populationsbasiertes Mammographie-Screeningprogramm eingeführt, das sich zunächst an Frauen im Alter von 50 bis 69 Jahren richtete. Ziel dieses Programms ist die frühzeitige Detektion maligner Tumoren in einer asymptomatischen Population.

Seit dem 1. Juli 2024 wurde die obere Altersgrenze angehoben. Frauen können nun bis zum 75. Lebensjahr am Screening teilnehmen. Bei regelmäßiger Teilnahme erfolgt die Untersuchung im zweijährigen Intervall,

sodass im Verlauf des Screeningzeitraums maximal 13 Mammographien durchgeführt werden.

Die Teilnahmequote im deutschen Mammographie-Screening liegt derzeit bei etwa 50 % der anspruchsberechtigten Frauen. Verlässliche bundesweite Daten darüber, wie viele Frauen das Screening nach einer oder mehreren Teilnahmen dauerhaft abbrechen, liegen bislang nicht vor.

Die Mammographie gilt insbesondere bei postmenopausalen Frauen als primäres bildgebendes Screeningverfahren. Ultraschall und MRT werden überwiegend als ergänzende Diagnostik eingesetzt, beispielsweise zur Abklärung suspekter Mammographiebefunde oder bei Patientinnen mit erhöhtem Risiko.

Vor diesem Hintergrund stellen sich zwei grundlegende diagnostische Fragen:

1. Wie hoch ist die Wahrscheinlichkeit eines tatsächlich vorliegenden Mammakarzinoms, wenn erstmals ein positiver Befund in der Mammographie (oder im MRT) erhoben wird?
2. Führt die Kombination mehrerer diagnostischer Verfahren zwangsläufig zu einer höheren diagnostischen Sicherheit?

Die folgenden Berechnungen basieren auf vereinfachten Modellannahmen zu Prävalenz, Sensitivität und Spezifität. Ziel ist es nicht, reale klinische Parameter exakt abzubilden, sondern die statistischen Implikationen zu verdeutlichen, die bei der Kombination mehrerer diagnostischer Tests wirksam werden können.

Insbesondere soll gezeigt werden, dass intuitive Erwartungen hinsichtlich diagnostischer Sicherheit unter bestimmten Bedingungen täuschen können, wenn mehrere Untersuchungsverfahren auf eine Population mit niedriger Prävalenz der Erkrankung angewendet werden.

## 2 Terminologie

Zur analytischen Beschreibung diagnostischer Testverfahren werden im Folgenden etablierte epidemiologische Kennzahlen verwendet. Für jede Stufe eines diagnostischen Prozesses lassen sich aus Prävalenz sowie den Testgüteparametern Sensitivität und Spezifität bedingte Wahrscheinlichkeiten ableiten, insbesondere der positive Vorhersagewert (positive predictive value, PPV).

Zur besseren Nachvollziehbarkeit werden Prozentangaben in absolute Fallzahlen überführt und anschließend in einem Baumdiagramm dargestellt. Im Folgenden werden die verwendeten Begriffe definiert.

### 2.1 Prävalenz (Pr)

Synonyme: Vortest-Wahrscheinlichkeit, a-priori-Wahrscheinlichkeit.  
Die Prävalenz beschreibt den Anteil von Personen in einer Population, die zu einem bestimmten Zeitpunkt an einer definierten Erkrankung leiden. Sie entspricht damit der Wahrscheinlichkeit, dass eine zufällig ausgewählte Person aus der betrachteten Population tatsächlich erkrankt ist.

$$Pr = \frac{\text{Anzahl Erkrankte}}{\text{Größe der Population}}$$

Die Prävalenz bestimmt somit die Ausgangswahrscheinlichkeit vor Durchführung irgendwelcher diagnostischer Tests und stellt damit eine zentrale Größe für die Interpretation von Untersuchungsergebnissen dar.

Beispiel Mammakarzinom:

In einer Population asymptomatischer Frauen im Alter von 50 – 69 Jahren finden sich im Mittel etwa 8 Erkrankte pro 1 000 Frauen [5]. Daraus ergibt sich eine Prävalenz von

$$Pr = 0.008$$

Diese Größe entspricht der Vortest-Wahrscheinlichkeit, mit der ein Mammakarzinom in der betrachteten Population vorliegt.

## 2.2 Sensitivität (Se)

Die Sensitivität beschreibt die Fähigkeit eines diagnostischen Tests, tatsächlich erkrankte Personen korrekt als krank zu identifizieren. Sie entspricht dem Anteil richtig-positiver Testergebnisse unter allen Erkrankten, also der Erkennungs- oder Detektionsrate.

$$Se = \frac{\text{Anzahl richtig Positive}}{\text{Anzahl Erkrankte}}$$

Die Sensitivität wird bestimmt, indem der Test bei einer Population angewendet wird, deren Krankheitsstatus bereits durch ein Referenzverfahren („Goldstandard“) gesichert ist.

Die Komplementärgröße („Versagerquote“) ist die Falsch-Negativ-Rate:

$$FNR = 1 - Se$$

Sie beschreibt den Anteil der Erkrankten, bei denen die Erkrankung durch den Test fälschlicherweise nicht erkannt wird.

Für die Mammographie wird in der Literatur (uneinheitlich) eine Sensitivität von etwa 90% angegeben, was bedeutet, dass etwa 10 % der tatsächlich erkrankten Frauen durch dieses Verfahren nicht erkannt werden.

Die Sensitivität der kontrastmittelverstärkten MRT liegt höher und wird mit ca. 97% angegeben.

Für die folgenden Modellrechnungen gelten folgende Annahmen:

Se Mammographie = 0.90

Se MRT=0.97

## 2.3 Spezifität (Sp)

Die Spezifität beschreibt die Fähigkeit eines diagnostischen Tests, gesunde Personen korrekt als gesund zu klassifizieren. Sie entspricht dem Anteil richtig-negativer Testergebnisse unter allen tatsächlich Gesunden.

$$Sp = \frac{\text{Anzahl richtig Negative}}{\text{Anzahl Gesunde}}$$

Die Komplementärgröße ist die Falsch-Positiv-Rate:

$$FPR = 1 - Sp$$

Sie gibt den Anteil gesunder Personen an, die durch den Test fälschlicherweise als krank eingestuft werden.

Für die Mammographie wird in der Literatur eine Spezifität von etwa 0.95 angegeben. Für die MRT liegt sie typischerweise niedriger; häufig werden Werte um 0.91 berichtet.

In der S3-Leitlinie zum Mammakarzinom werden für die kontrastmittelverstärkte MRT teilweise noch geringere Spezifitäten angegeben, abhängig von Indikation und Studienpopulation.

Niedrige Spezifitätswerte führen zu einer erhöhten Zahl falsch-positiver Befunde und damit zu zusätzlichen diagnostischen Abklärungen, einschließlich weiterer Bildgebung oder Biopsien.

Annahmen für die Modellrechnung:

$$Sp \text{ Mammographie} = 0.95$$

$$Sp \text{ MRT} = 0.91$$

## 2.4 Positiver Vorhersagewert (*positive predictive value*, PPV)

Der positive Vorhersagewert beschreibt die Wahrscheinlichkeit, dass eine Person mit positivem Testergebnis tatsächlich erkrankt ist. Er entspricht dem Anteil richtig-positiver Ergebnisse unter allen positiven Testergebnissen.

Formal:

$$Sp = \frac{\text{Anzahl richtig Positive}}{\text{Gesamtzahl positiv Geteste}}$$

$$PPV = P(K|T+)$$

Der PPV ist eine bedingte Wahrscheinlichkeit und hängt neben Sensitivität und Spezifität entscheidend von der Prävalenz der Erkrankung in der untersuchten Population ab.

Ein diagnostischer Test erhöht den Informationsgehalt einer Untersuchung dann, wenn er zu einer deutlichen Steigerung der Nachtest-Wahrscheinlichkeit im Vergleich zur Vortest-Wahrscheinlichkeit führt.

Die Zusammenhänge zwischen Krankheitsstatus und Testergebnis lassen sich in einer Vierfeldertafel (Kontingenztafel) übersichtlich darstellen. Im diagnostischen Kontext können zwei grundsätzliche Fehlklassifikationen auftreten:

1. Falsch-negative Befunde  
(Erkrankung vorhanden, Test negativ)
2. Falsch-positive Befunde  
(Erkrankung nicht vorhanden, Test positiv)

## Abbildung 1:

Mögliche Ergebnisse eines diagnostischen Tests

**Diagnostische Maßnahme**  
Mögliche Ergebnisse

Diagnoseverfahren		Testergebnis	Goldstandard		$\Sigma$	
			Zielgröße: Krankheit			
			liegt vor	liegt nicht vor		
Positiv	<b>a</b>	<b>b</b>	$n_1$	Positive Predictive Value (PPV) $\frac{a}{n_1}$		
	<b>c</b>	<b>d</b>	$n_2$	Negative Predictive Value (NPV) $\frac{d}{n_2}$		
$\Sigma$		<b><math>n_3</math></b>	<b><math>n_4</math></b>	<b>n</b>		

  

Sensitivität $\frac{a}{n_1}$	Spezifität $\frac{d}{n_4}$
------------------------------	----------------------------

  

$n_1 = a + b$	$n_3 = a + c$	$c \Rightarrow$ falsch Negative
$n_2 = c + d$	$n_4 = b + d$	$b \Rightarrow$ falsch Positive

Häufig wird in der Medizin der Fehler mit den schwerwiegenderen Folgen als Fehler erster Art oder Alpha-Fehler bezeichnet, der weniger gravierende Fehler als Fehler zweiter Art oder Beta-Fehler.

Im Falle von bösartigen Tumoren gilt das Nichterkennen der Malignität konventionsgemäß wegen der potentiell tödlichen Folgen als Fehler erster Art [3]. Dies entspricht dem fälschlichen Beibehalten der Nullhypothese, obwohl die Alternativhypothese zutrifft.

In den statistischen Wissenschaften wird im Gegensatz hierzu üblicherweise die fälschliche Ablehnung der Nullhypothese (der Fehlalarm) als Alpha-Fehler bezeichnet.

## 3 Darstellung am konkreten Beispiel

### 3.1 Mammographie/MRT als Einzeluntersuchung

Aus klinischer Perspektive besteht die zentrale Anforderung an ein Screeningverfahren darin, eine tatsächlich vorliegende Erkrankung möglichst zuverlässig zu erkennen. Gleichzeitig ist eine geringe Rate falsch-positiver Befunde von großer Bedeutung, da die Fehl-Klassifikation gesunder Personen zu unnötigen Folgeuntersuchungen, risikobehafteten Biopsien, psychischer Belastung und vermeidbaren Kosten führen kann.

Die diagnostische Leistungsfähigkeit eines Tests wird nicht allein durch Sensitivität und Spezifität bestimmt. Für die klinische Interpretation eines positiven Befundes hat sich der positive Vorhersagewert (PPV) bewährt. Seine Berechnung entspricht der praktischen Anwendung des Bayes-Theorems.

Mehr Screening mit breiter Indikationsstellung führt nicht automatisch zu mehr Sicherheit in der diagnostischen Aussage. Mit jeder zusätzlichen Untersuchung im Zielkollektiv der unter Risiko stehenden Frauen steigt nicht nur der diagnostische Gewinn für Einzelne – es kommen auch weitere Falsch-Positive hinzu.

Und da sich in einer (unselektierten) Kohorte normalerweise mehr Gesunde als Kranke befinden, zumindest bei Brustkrebs, werden bei großem Screening-Volumen zwangsläufig auch mehr Falsch-Positive detektiert als wirklich Erkrankte.

Diese Überlegung begründet, warum die Prävalenz auf alle theoretischen Berechnungen, die man durchführen kann, einen erheblichen Einfluss hat.

Die folgende Abbildung verdeutlicht die Zusammenhänge (gerundete Werte):



bei 7 (krank und erkannt) von insgesamt 57 (positiv Getesteten), also bei rund 12.3% (positive predictive value:  $PPV = 7/57$ ).

Oder anders formuliert: nahezu 88% der Frauen mit einem erstmaligen, positiven Mammographie-Screening-Befund haben keinen Brustkrebs.

Die Situation ändert sich, wenn als zweites Untersuchungsverfahren beispielsweise ein MRT angeschlossen wird, um die Ergebnisse durch ergänzende Bildgebung evtl. „weiter zu verbessern“.

Theoretische Annahme:

Würde man *alle* 1 000 Frauen mit dem MRT nachuntersuchen (wird man aus medizinischen und Kostengründen nicht tun, das Beispiel soll nur die statistischen Implikationen verdeutlichen), so würde auf der zweiten Stufe auch die eine, mit der Mammographie nicht entdeckte Tumorpatientin durch das MRT nahezu sicher diagnostiziert werden. Insgesamt ergäben sich damit acht tatsächlich Erkrankte, aber rund 89 Falsch-Positive nach konsekutiver Durchführung des MRT-Verfahrens.

Zur Verdeutlichung:

Nach der alleinigen Mammographie gab es 50 Falsch-Positive. Auf der zweiten Stufe (Mammographie plus MRT) ist diese Zahl auf 89 angewachsen, sofern man tatsächlich alle Frauen mit dem MRT nachuntersuchen würde.

Durch die große Zahl neu hinzugekommener Falsch-Positiver (im Vergleich zu nur einer hinzugekommenen Richtig-Positiven) würde der Vorhersagewert von 12.3% nach der ersten Stufe jetzt plötzlich auf rund 8.2% absinken ( $PPV = 8/97$ ).

In diesem hypothetischen Szenario nähme trotz zusätzlicher Bildgebung die diagnostische Sicherheit bei einem positiven Befund ab, da die falsch-positiven Befunde überwiegen

Abbildung 2 zeigt, dass das MRT letztlich nur dazu dienen würde, die eine, der Mammographie entgangene Patientin zu identifizieren. Und diese gehört möglicherweise (mit einer 10 bis 20%-igen

Wahrscheinlichkeit) in die Gruppe der „Überdiagnosen“, die von einer Entdeckung gar nicht profitiert.

### 3.2 Mammographie als Serien-Untersuchung

Die wiederholte Anwendung eines diagnostischen Tests führt zwangsläufig zu einer kumulativen Wahrscheinlichkeit diagnostischer Fehlklassifikationen. Zur Illustration kann das vereinfachte probabilistische Modell der Bernoulli-Kette herangezogen werden.

Bei einer einmaligen Mammographie beträgt die Wahrscheinlichkeit, dass eine gesunde Frau korrekt als gesund klassifiziert wird, entsprechend der Spezifität:  $p = 0.95$

Unter der vereinfachenden Annahme stochastischer Unabhängigkeit der einzelnen Untersuchungen ergibt sich für zehn aufeinanderfolgende Screeninguntersuchungen die Wahrscheinlichkeit, dass eine gesunde Frau in keiner der Untersuchungen ein falsch-positives Ergebnis erhält, zu

$$p_{10} = (0.95)^{10} \approx 0.599$$

Die Wahrscheinlichkeit, nach zehn Untersuchungen immer noch korrekt als gesund klassifiziert zu werden, beträgt demnach etwa 59.9 %.

Die Gegenwahrscheinlichkeit beschreibt das Risiko, mindestens einmal ein falsch-positives Testergebnis zu erhalten:

$$1 - (0.95)^{10} \approx 0.401$$

Damit ergibt sich unter diesen Annahmen eine Wahrscheinlichkeit von etwa 40 %, dass eine gesunde Frau im Verlauf von zehn Screeninguntersuchungen mindestens einmal ein falsch-positives Ergebnis erhält.

Diese Berechnung basiert auf der Annahme stochastischer Unabhängigkeit der einzelnen Untersuchungen, die in der klinischen Realität nur näherungsweise erfüllt ist, da neben der individuellen Brustdicke, technische und andere Faktoren korreliert sein können.

Dennoch verdeutlicht dieses Modell einen grundlegenden statistischen Sachverhalt:

Mit zunehmender Zahl wiederholter Untersuchungen steigt die kumulative Wahrscheinlichkeit falsch-positiver Befunde.

Für Screeningprogramme bedeutet dies, dass mit der Wiederholung diagnostischer Tests nicht nur die Wahrscheinlichkeit der Tumordetektion steigt, sondern eben auch die Wahrscheinlichkeit diagnostischer Fehlalarme.

Dieses Verhältnis zwischen diagnostischem Nutzen und potenziellen Fehlklassifikationen stellt einen zentralen Aspekt bei der Bewertung populationsbasierter Screeningstrategien dar.

## 4 Zusammenfassung

In einer Population asymptomatischer Frauen im Alter zwischen 50 und 69 Jahren beträgt die Prävalenz des Mammakarzinoms etwa 0.8 %. Für eine zufällig ausgewählte Frau aus dieser Altersgruppe entspricht dies der Vortest-Wahrscheinlichkeit, mit der eine Brustkrebserkrankung vor Durchführung diagnostischer Maßnahmen vorliegt.

Wird eine Mammographie als primäres Screeningverfahren eingesetzt, erhöht sich bei einem positiven Befund die Wahrscheinlichkeit einer tatsächlich vorliegenden Erkrankung auf etwa 12.3 %. Dieser Wert entspricht dem positiven Vorhersagewert (PPV) des Tests unter den im Modell angenommenen Parametern für Prävalenz, Sensitivität und Spezifität.

Die Durchführung zusätzlicher bildgebender Diagnostik führt unter bestimmten Bedingungen jedoch nicht automatisch zu einer weiteren Erhöhung der diagnostischen Sicherheit. Würde beispielsweise hypothetisch bei allen Frauen einer Screeningkohorte nach einer Mammographie zusätzlich eine MRT-Untersuchung durchgeführt, so führte die vergleichsweise niedrige Spezifität dieses Verfahrens zu einer erheblichen Zunahme falsch-positiver Befunde. In dem dargestellten Modell läge der positive Vorhersagewert dann dadurch nur noch bei etwa 8.2%.

Dieses Ergebnis verdeutlicht, dass der Prävalenz einer Erkrankung in der untersuchten Population eine zentrale Rolle zukommt.

Ein weiterer Aspekt betrifft die wiederholte Anwendung diagnostischer Tests im Rahmen eines Screeningprogramms. Bei einer Spezifität der Mammographie von etwa 95 % ergibt sich unter der vereinfachenden Annahme stochastischer Unabhängigkeit der einzelnen Untersuchungen bei zehn Screeningrunden eine Wahrscheinlichkeit von rund 40 %, dass eine tatsächlich gesunde Frau im Verlauf dieses Zeitraums mindestens einmal ein falsch-positives Testergebnis erhält.

Mit zunehmender Zahl diagnostischer Untersuchungen steigt somit nicht nur die Wahrscheinlichkeit der Tumordetektion, sondern auch die kumulative Wahrscheinlichkeit diagnostischer Fehlklassifikationen.

Für Screeningprogramme ergibt sich daraus ein grundlegendes Spannungsfeld zwischen dem Nutzen einer frühen Tumorerkennung und den potenziellen Belastungen durch falsch-positive Befunde und den daraus resultierenden Folgeuntersuchungen.

## 5 Quellenverzeichnis

1. Braun B, Khil L, Tio J, et al.  
Differences in Breast Cancer Characteristics by Mammography Screening Participation or Non-Participation: A Retrospective Observational Study.  
Dtsch Arztebl Int. 2018;115(31-32):520–527.
2. Buschmann L, Bonberg N, Baurecht H, et al.  
Participation in the German Mammography Screening Program: An Analysis of Data From the NAKO Health Study. Dtsch Arztebl Int. 2025;122(24):655-662.
3. Fassl H.  
Einführung in die Medizinische Statistik.  
Heidelberg: Barth; 1999
4. Fochler S, Weitmann K, Domin M, Hoffmann W.  
Reasons for Non-Attendance in the German National Mammography Screening Program: Which Barriers Can Be Overcome Using Telephone Counseling?— A Randomized Controlled Trial.  
Healthcare. 2022;11(1):17.
5. Gigerenzer G.  
Das Einmaleins der Skepsis: Über den richtigen Umgang mit Zahlen und Risiken.  
Berlin: Berliner Taschenbuch Verlag; 2004.
6. Gigerenzer G, Kuoni J, Ritschard R.  
Was Ärzte wissen müssen: Die Kunst der Risikokommunikation.  
Basel: Schweizerische Ärztezeitung; 2013.

7. Heidinger O, Batzler WU, Krieg V, et al.  
The incidence of interval cancers in the German mammography screening program: results from the population-based cancer registry in North Rhine-Westphalia.  
Dtsch Arztebl Int. 2012;109(46):781-7.
8. Heinig M, Schäfer W, Langner I, Zeeb H, Haug U.  
German mammography screening program: adherence, characteristics of (non-)participants and utilization of non-screening mammography—a longitudinal analysis.  
BMC Public Health. 2023; 23:1678.
9. Leitlinienprogramm Onkologie (AWMF, DKG, DKH).  
S3-Leitlinie Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms.  
Version 5.0. Berlin: 2025.
10. Robert Koch-Institut, Gesellschaft der epidemiologischen Krebsregister in Deutschland e. V. (Hrsg).  
Krebs in Deutschland für 2021 – 2023.  
15. Ausgabe. Berlin: RKI; 2025.
11. Starker A, Kraywinkel K, Kuhnert R.  
Early detection of breast cancer: the utilization of mammography in Germany.  
Journal of Health Monitoring. 2017;2(4):69–75.