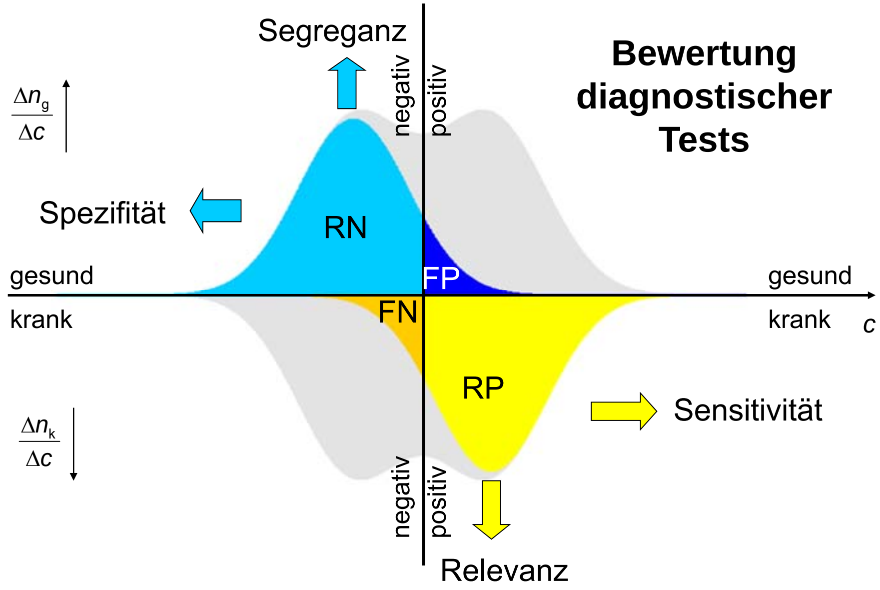
Bewertung diagnostischer Tests



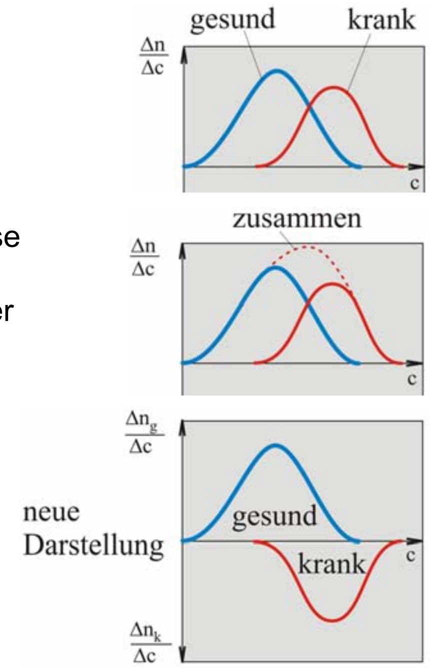
KAD 2011.11.24

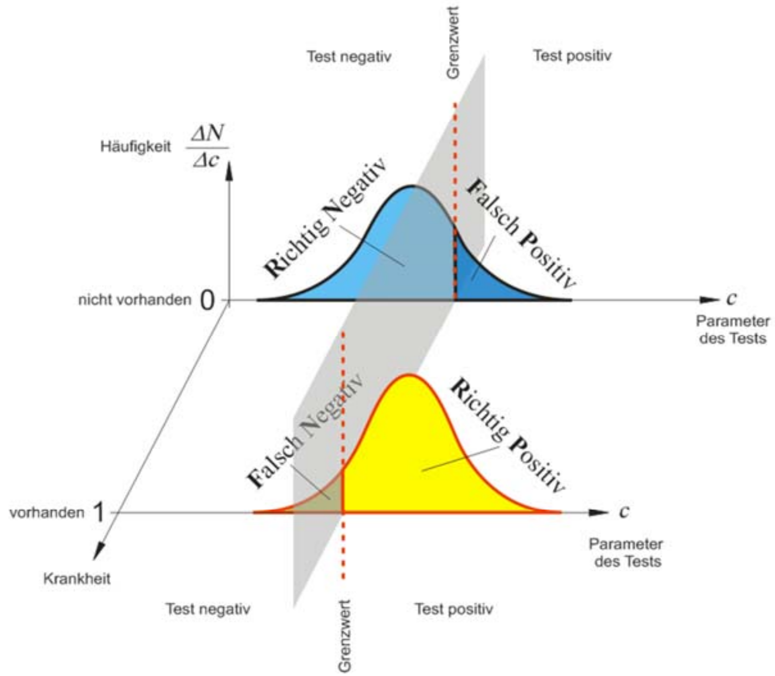
Überlappende Populationen

Annahme:

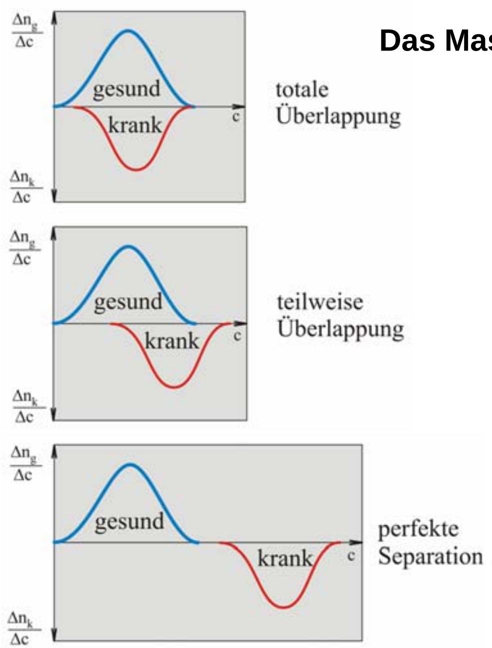
eine messbare Grösse (z.B Konzentration) vergrössert sich in der kranken Population

die Veränderung ist wichtig, nicht die Vergrösserung





Das Mass der Überlappung



nutzlose Methode

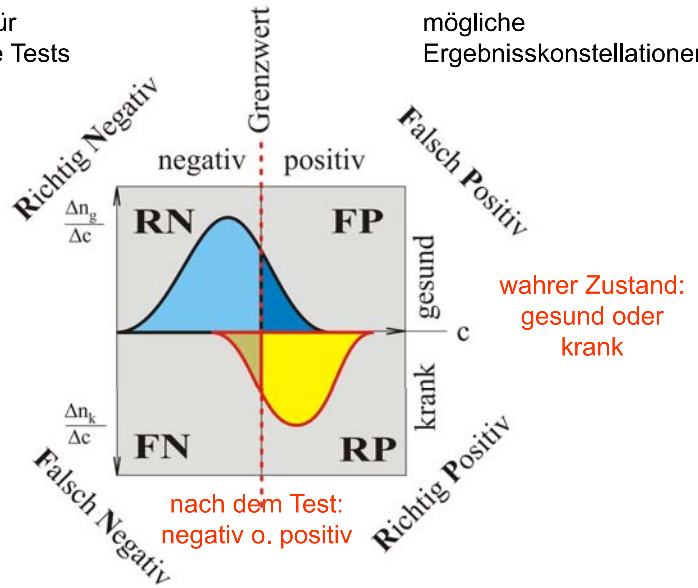
reelle Methoden

perfekte Methode

Wahrheitsmatrix

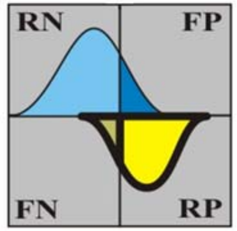
Basismatrix für diagnostische Tests

mögliche Ergebnisskonstellationen

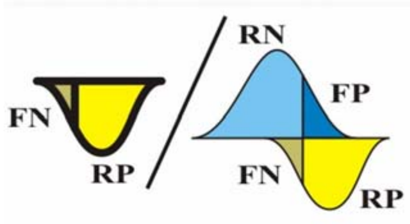


Prävalenz

$$(w = \frac{de - sp}{se - sp})$$

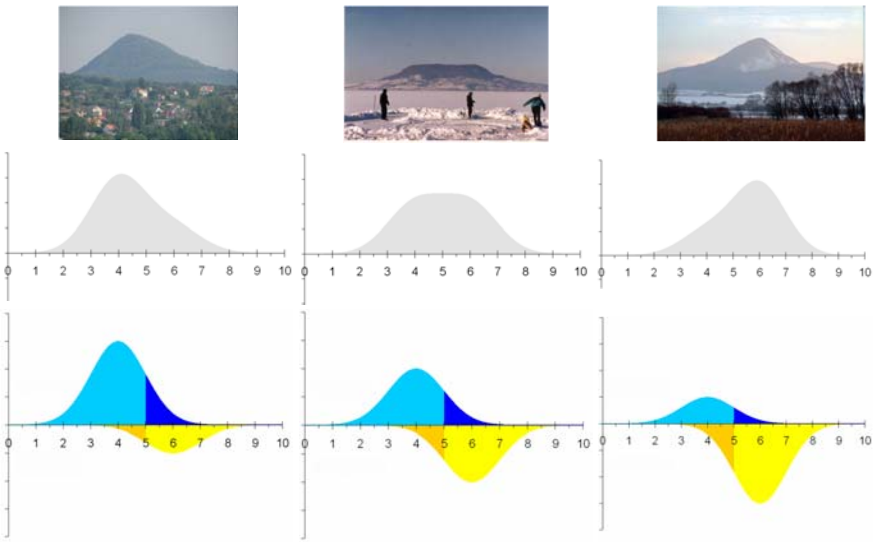


Prävalenz: Häufigkeit einer Krankheit in einer Population, Krankheitshäufigkeit, Wahrscheinlichkeit vor dem Test (Vortestwahrscheinlichkeit, a-priori-Wahrscheinlichkeit)



$$\boxed{w = \frac{RP + FN}{RP + RN + FN + FP}}$$

$$w = \frac{\text{alle Kranken}}{\text{alle Untersuchten}}$$



w = 25% w = 50% w = 75%

Die Zuverlässigkeit diagnostischer Tests wird mit den **Kennwerten** (Validätsparameter) beschrieben.

Sensitivität

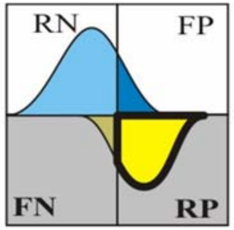
Spezifität

Relevanz

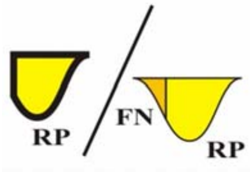
Segreganz

Jeder Test sollte an einem internationalen Standard geeicht werden, und es sollte eine **Referenzmethode** (Goldstandard) zur Erfassung des tatsächlichen Zustandes des Patienten verfügbar sein.

Diagnostische **Sensitivität** (se)



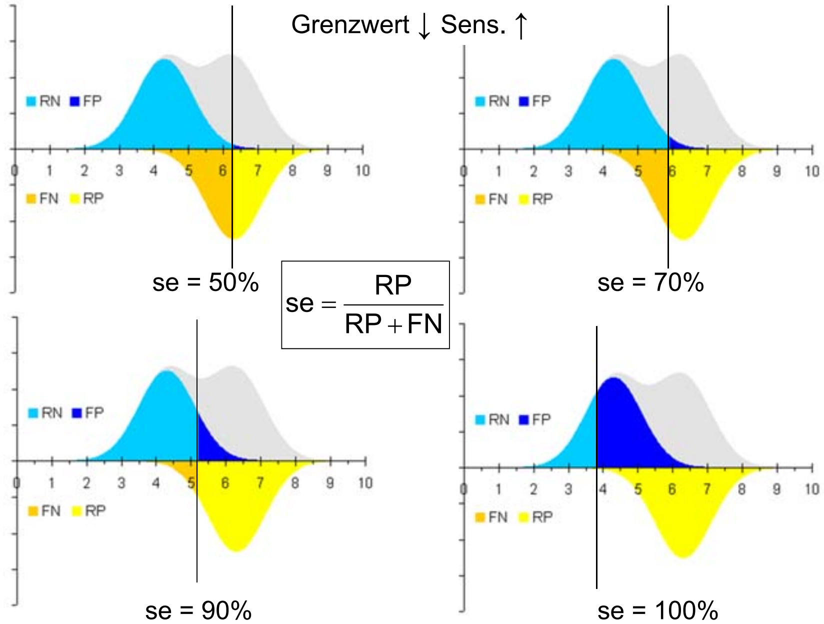
Empfindlichkeit. Wahrscheinlichkeit, einen Kranken als positiv zu erkennen



$$se = \frac{\text{richtig positiv}}{\text{krank}} = \frac{RP}{RP + FN}$$

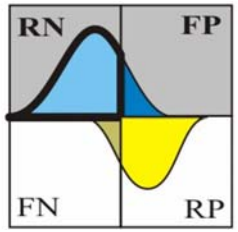
Tests mit hoher Sensitivität sind bei der **Frühdiagnostik** (screening) von Krankheiten erwünscht, und wenn es darauf ankommt, dass möglichst wenig Kranke unentdeckt bleiben.

Grenzwert ↓ Sens. ↑

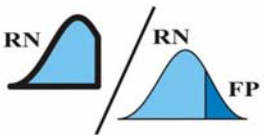


$$se = \frac{RP}{RP + FN}$$

Diagnostische **Spezifität** (sp)



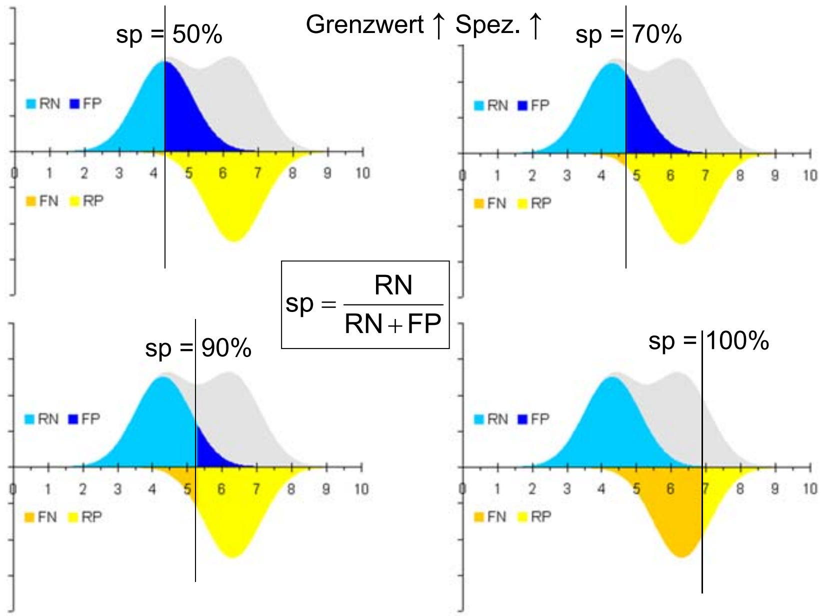
Wahrscheinlichkeit, einen Gesunden als negativ zu erkennen



$$sp = \frac{\text{richtig negativ}}{\text{gesund}} = \frac{RN}{RN + FP}$$

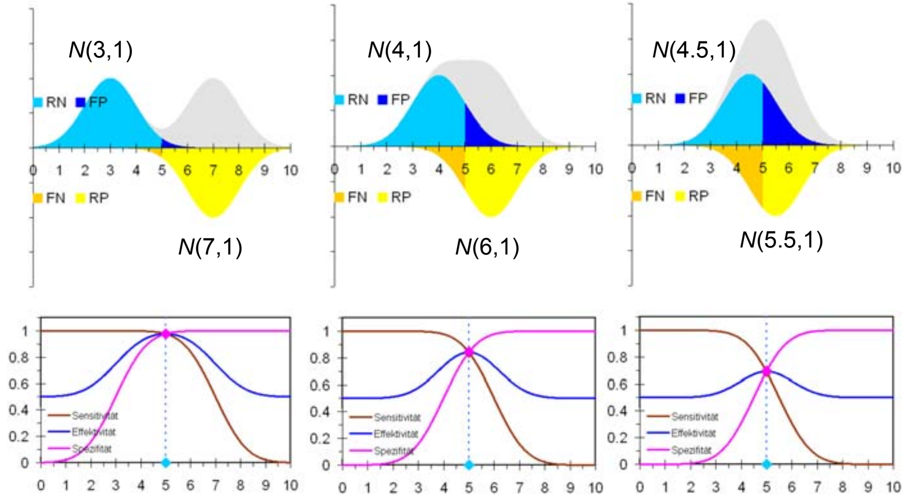
Tests mit hoher Spezifität sind als **Bestätigungstests** erwünscht und in allen Situationen, in denen eine falsch-positive Diagnose fatale Folgen hätte.

Grenzwert ↑ Spez. ↑



$$sp = \frac{RN}{RN + FP}$$

Sensitivität u. Spezifität: gegenläufige Eigenschaften von Testen

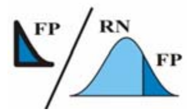


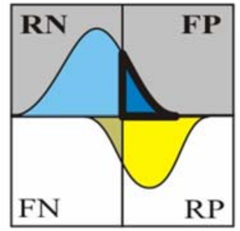
grössere Überlappung der Verteilungen →

Diagnostische Falschpositivrate

$$1-sp = \frac{FP}{RN+FP}$$

(vgl. Fehler 1. Art)

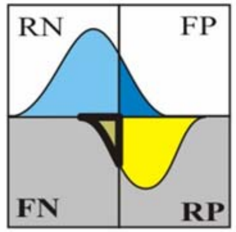




Diagnostische Falschnegativrate

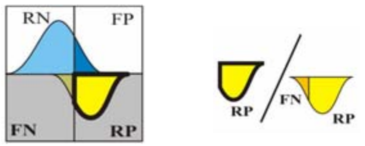
$$1-se = \frac{FN}{FN+RP}$$

(vgl. Fehler 2. Art)



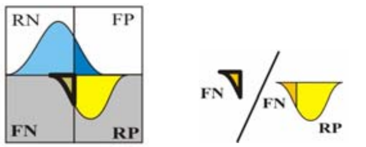
horizontale Raten hängen von der Prävalenz nicht ab

Sensitivität (se)



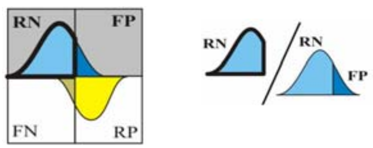
$$se = \frac{RP}{RP+FN}$$

Falschnegativrate (1-se)



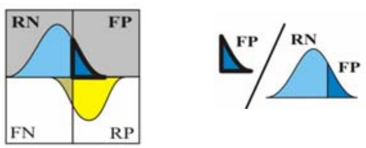
$$1-se = \frac{FN}{FN+RP}$$

Spezifität (sp)



$$sp = \frac{RN}{RN+FP}$$

Falschpositivrate (1-sp)



$$1-sp = \frac{FP}{RN+FP}$$

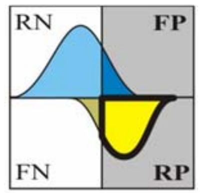
Vorhersagewerten (prädiktive Werte, vertikale Raten)

Wahrscheinlichkeiten nach dem Test, Nachtestwahrscheinlichkeiten, a-posteriori-Wahrscheinlichkeiten

Diagnostische **Relevanz**

positiv prädiktiver Wert,
positiver Vorhersagewert,
positive predictive value

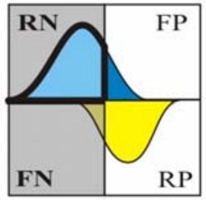
$$PPV = \frac{se \cdot w}{se \cdot w + (1-sp) \cdot (1-w)}$$



Wahrscheinlichkeit eines Test-Positiven, krank zu sein

$$\boxed{PPV = \frac{\text{richtig positiv}}{\text{positiv}} = \frac{RP}{RP+FP}}$$

Diagnostische **Segreganz**



negativ prädiktiver Wert,
negativer Vorhersagewert,
negative predictive value

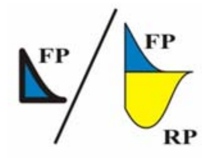
$$NPV = \frac{sp \cdot (1-w)}{sp \cdot (1-w) + (1-se) \cdot w}$$

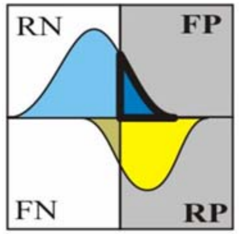
Wahrscheinlichkeit eines Test-Negativen, gesund zu sein

$$\boxed{PVN = \frac{\text{richtig negativ}}{\text{negativ}} = \frac{RN}{RN + FN}}$$

Falschalarm(rate)

$$1\text{-}PPV = \frac{FP}{FP + RP}$$

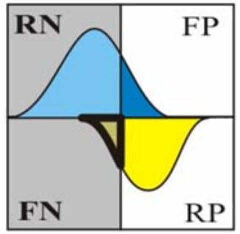




---

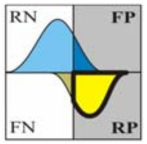
falsche Beruhigung(srate)

$$1\text{-}NPV = \frac{FN}{FN + RN}$$



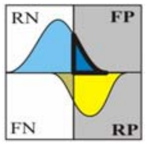
vertikale Raten hängen von der Prävalenz ab

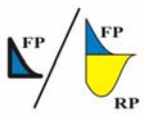
Relevanz (PPV)



$$PPV = \frac{RP}{FP + RP}$$

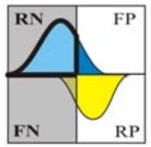
Falschalarmrate (1-PPV)





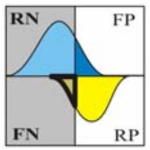
$$1\text{-}PPV = \frac{FP}{FP + RP}$$

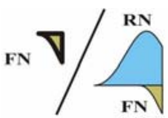
Segreganz (NPV)



$$NPV = \frac{RN}{RN + FN}$$

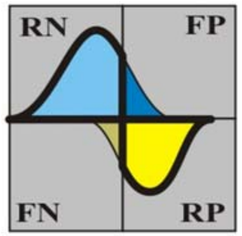
falsche Beruhigung (1-NPV)





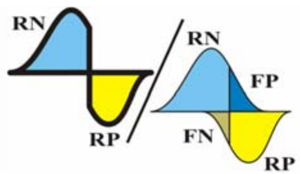
$$1\text{-}NPV = \frac{FN}{RN + FN}$$

Diagnostische **Effektivität**



richtige Klassifikationsrate

$$de = se \cdot w + sp \cdot (1-w)$$



$$de = \frac{\text{richtig positiv} + \text{richtig negativ}}{\text{gesund} + \text{krank}}$$

$$\boxed{de = \frac{RP + RN}{RN + FP + FN + RP}}$$

oft: Grenzwert ist so gewählt, dass Effektivität maximal ist

Effekt der Prävalenz

Beispiel *A*: ***w* = 50%**

PVN = 90%

			Test	
			negativ	positiv
sp = 90%	Gold-standard	gesund	90	10
		krank	10	90	se = 90%

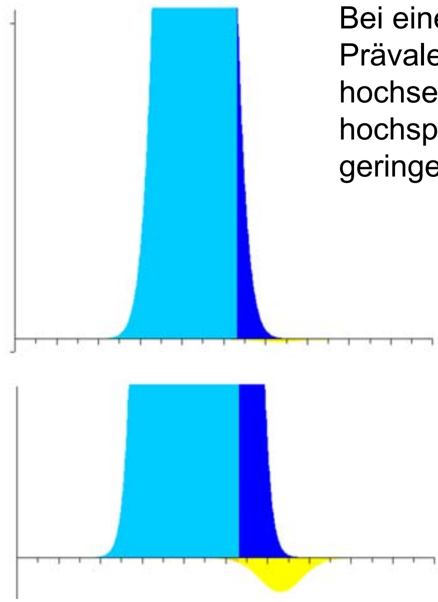
(de = 90%) PVP = 90%

Beispiel *B*: ***w* = 10%**

PVN = 99%

			Test	
			negativ	positiv
sp = 90%	Gold-standard	gesund	810	90
		krank	10	90	se = 90%

(de = 90%) PVP = 50%



Bei einer sehr kleineren Prävalenz können die hochsensitive und gleichzeitig hochspezifische Tests sehr geringe Relevanz haben.

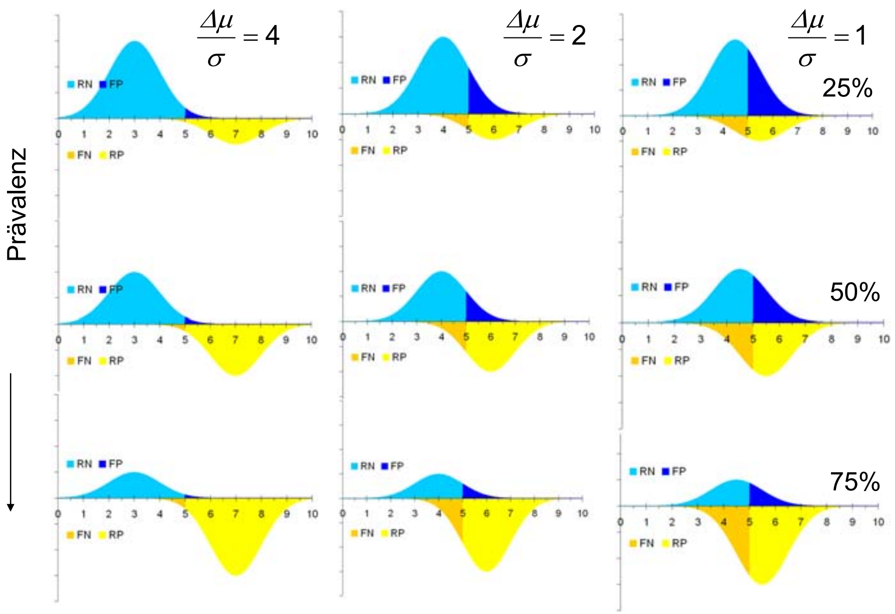
Prävalenz = 0.1 %

Sensitivität = 98 %

Spezifität = 98 %

↓

Relevanz = 4 %



grössere Überlappung der Verteilungen →

Übersichtstabelle

← bedingte wahrscheinlichkeit (Bayes)

Sensitivität	se	$\frac{RP}{RP+FN}$	$p(P \mid K)$	Testpositiven zw. den Kranken	Richtigpositivrate, Empfindlichkeit
Spezifität	sp	$\frac{RN}{RN+FP}$	$p(N \mid G)$	Testnegativen zw. den Gesunden	Richtignegativ-rate
Falschnegativrate	1-se	$\frac{FN}{RP+FN}$	$p(N \mid K)$	Testnegativen zw. den Kranken	
Falschpositivrate	1-sp	$\frac{FP}{RN+FP}$	$p(P \mid G)$	Testpositiven zw. den Gesunden	
Relevanz; positiver prädiktiver Wert	PVP	$\frac{RP}{RP+FP}$	$p(K \mid P)$	Kranken zw. den Testpositiven	positiver Vorhersagewert
Segreganz; negativer prädiktiver Wert	PVN	$\frac{RN}{RN+FN}$	$p(G \mid N)$	Gesunden zw.den Testnegativen	negativer Vorhersagewert
Falschalarmrate	1-PVP	$\frac{FP}{RP+FP}$	$p(G \mid P)$	Gesunden zw. den Testpositiven	Fehlalarmrate
falsche Beruhigungsrate	1-PVN	$\frac{FN}{RN+FN}$	$p(K \mid N)$	Kranken zw. den Tesnegativen	

Wiederholung

Bedingte Wahrscheinlichkeit, $p(A|B)$

Die Wahrscheinlichkeit dass A zutrifft unter der Voraussetzung dass B eingetreten ist.

$$\boxed{p(A|B) = \frac{p(A \text{ und } B)}{p(B)}}$$

zB: beim Würfelexperiment:

$$p(>4|\text{Gerade}) = \frac{p(>4 \text{ und Gerade})}{p(\text{Gerade})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Ergänzungsmaterial

$$p(\text{positiv} \mid \text{krank}) = \frac{p(\text{positiv und krank})}{p(\text{krank})} = \frac{\frac{RP}{n}}{\frac{RP+FN}{n}} = \frac{RP}{RP+FN}$$

Sensitivität

Die totale Wahrscheinlichkeit:

Es seien $B_1, B_2, \ldots, B_n$ Ereignisse eines vollständigen Ereignissystems, dann gilt für ein beliebiges Zufallereignis A:

$$\boxed{p(A) = \sum_{i=1}^{n} p(A|B_i)p(B_i)}$$

Beispiel

B_1: krank
B_2: gesund
A: positiv

$$p(\text{positiv}) = p(A) = p(A|B_1)\cdot p(B_1) + p(A|B_2)\cdot p(B_2) =$$
$$= p(\text{positiv}|\text{krank})\cdot p(\text{krank}) + p(\text{positiv}|\text{gesund})\cdot p(\text{gesund}) =$$
$$= se \cdot w + (1-sp)\cdot(1-w) =$$
$$= \frac{RP}{RP+FN}\cdot\frac{RP+FN}{RP+FN+RN+FP} + \frac{FP}{RN+FP}\cdot\frac{RN+FP}{RP+FN+RN+FP} =$$
$$= \frac{RP+FP}{RP+FN+RN+FP}$$

Ergänzungsmaterial

Satz von Bayes:

B_1: krank
B_2: gesund
A: positiv

Wir suchen die Wahrscheinlichkeit B_k unter der Voraussetzung A gilt

$$p(B_k|A) = \frac{p(A|B_k)p(B_k)}{\sum_{i=1}^{n} p(A|B_i)p(B_i)} = \frac{p(A|B_k)p(B_k)}{p(A)}$$

Beispiel

$$p(B_1|A) = \frac{p(A|B_1)\cdot p(B_1)}{p(A|B_1)\cdot p(B_1) + p(A|B_2)\cdot p(B_2)} = PPV$$

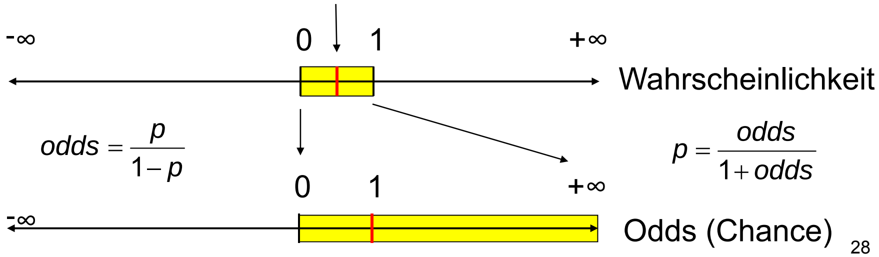
$$p(\text{krank}|\text{positiv}) = \frac{p(\text{positiv}|\text{krank})\cdot p(\text{krank})}{p(\text{positiv}|\text{krank})\cdot p(\text{krank}) + p(\text{positiv}|\text{gesund})\cdot p(\text{gesund})} =$$
$$= \boxed{\frac{se\cdot w}{se\cdot w + (1-sp)\cdot(1-w)} = PPV}$$

$$p(B_2|1-A) = \frac{p(1-A|B_2)\cdot p(B_2)}{p(1-A|B_1)\cdot p(B_1) + p(1-A|B_2)\cdot p(B_2)} = \boxed{\frac{sp\cdot(1-w)}{(1-se)\cdot w + sp\cdot(1-w)} = NPV}$$

Charakterisierungsmöglichkeiten des Eintretens von Ereignissen

Ereignis A	Wahrscheinlichkeit $p(A)$	odds
unmögliches Ereignis	0	0
Ereignis A und Gegenereignis von A haben die gleichen Chancen	0.5	1
sicheres Ereignis	1	∞

Ereignis A und Gegenereignis von A haben die gleiche Chance



Likelihood ratio

Faktor, der angibt in welchem Ausmass eine Untersuchung die Vortestchance verändert.

likelihood ratio eines positiven Testresultates:

$$\boxed{LR_{pos}} = \frac{\frac{RP}{FP}}{\frac{krank}{gesund}} = \frac{\frac{RP}{RP+FN}}{\frac{FP}{RN+FP}} = \frac{RP}{RP+FN}\,\frac{1}{\frac{FP}{RN+FP}} = \boxed{\frac{se}{1-sp}}$$

likelihood ratio eines negativen Testresultates:

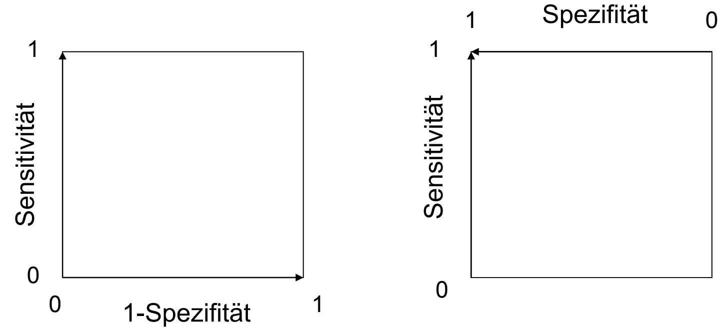
$$\boxed{LR_{neg}} = \frac{\frac{FN}{RN}}{\frac{krank}{gesund}} = \frac{\frac{FN}{RP+FN}}{\frac{RN}{RN+FP}} = \frac{FN}{RP+FN}\,\frac{1}{\frac{RN}{RN+FP}} = \boxed{\frac{1-se}{sp}}$$

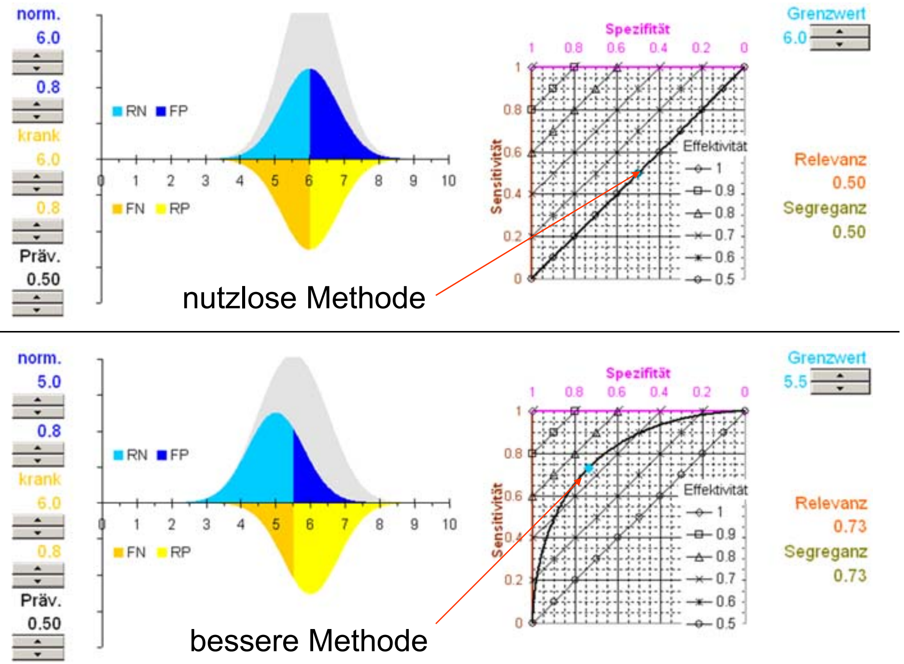
Vergleichung verschiedener diagnostischer Methode. ROC Kurven

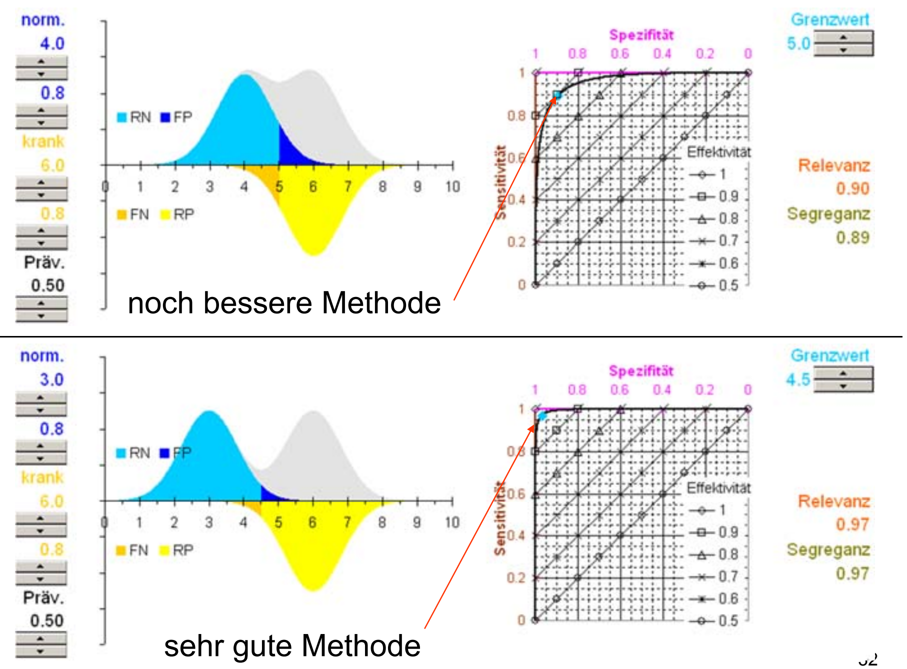
ROC: receiver-operator (operating) characteristic

ca. 1950: erste ROC Analyse (receiver: Radar Empfänger)

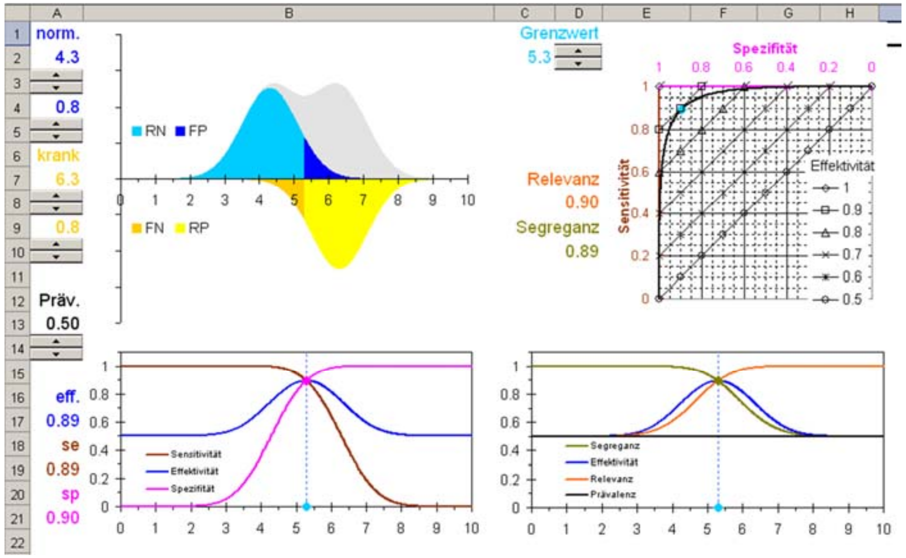
ca. 1970: die erste medizinische Anwendungen





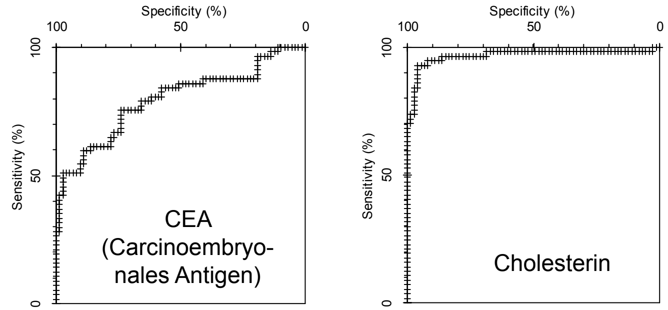


ROC Analyse



Beispiel: Tumormarker im Bauchwasser (Ascites)

Die Erhöhung von CEA (und/oder Cholesterin) Konzentration im Bauchwasser kann mit Karzinose in Zusammenhang bringen.



Welche Methode ist besser? Wie kann man den optimalen Grenzwert auswählen?

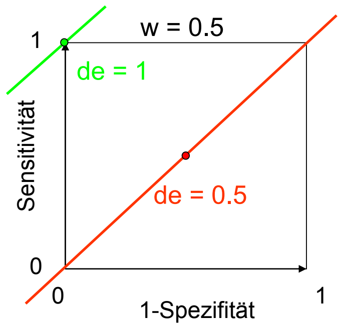
Gulyás M, Kaposi AD, Elek G, Szollár LG, Hjerpe A, Value of carcinoembryonic antigen (CEA) and cholesterol assays of ascitic fluid in cases of inconclusive cytology, J Clinical Pathology 2001 (54) 831-835

$$de = se \cdot w + sp \cdot (1-w)$$

$$\frac{de}{1-w} = \frac{w}{1-w} se + (sp-1) + 1$$

$$(1-sp) + \frac{de}{1-w} - 1 = \frac{w}{1-w} se$$

$$se = \underbrace{\frac{1-w}{w}}_{\text{Steigung}}(1-sp) + \underbrace{\frac{1}{w}de + \frac{w-1}{w}}_{\text{Achsenabschnitt}}$$



wenn $w = 0.5$: $\boxed{se = 1 \cdot (1-sp) + 2 \cdot de - 1}$

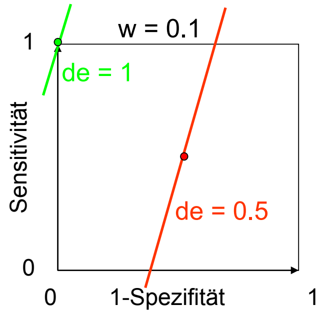
Die Punkte, die gleiche diagnostische Effektivität haben, sind auf den Geraden mit einer Steigung von 1.

Wenn *de* = 0.5 ist, dann beträgt der Achsenabschnitt 0.

$$\boxed{se = \underbrace{\frac{1-w}{w}}_{\text{Steigung}}(1-sp) + \underbrace{\frac{1}{w}de + \frac{w-1}{w}}_{\text{Achsenabschnitt}}}$$

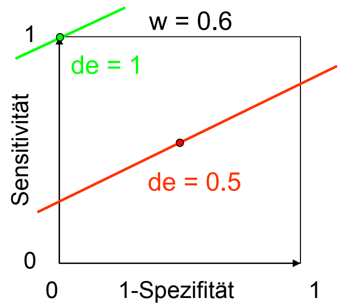
wenn w < 0.5: Die Steigung der Geraden mit gleicher diagnostischen Effektivität ist grösser als 1.

z.B. *w* = 0.1, die Steigung: 9

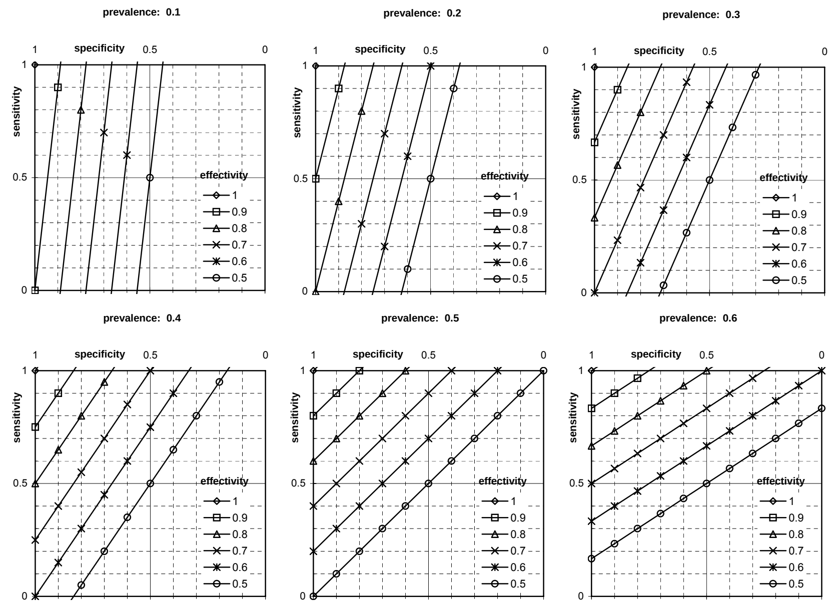


wenn w > 0.5: Die Steigung der Geraden mit gleicher diagnostischen Effektivität ist kleiner als 1.

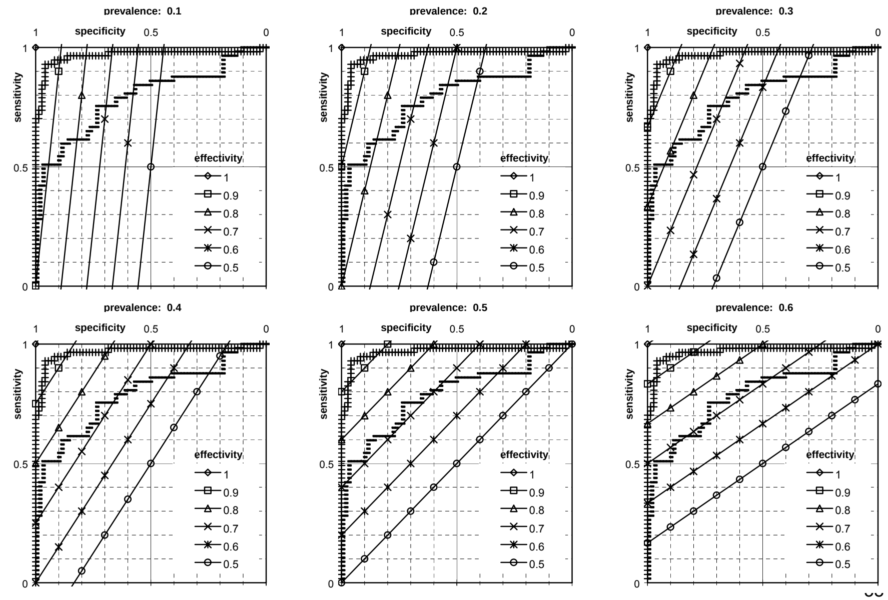
z.B. *w* = 0.6, Steigung: 0.66



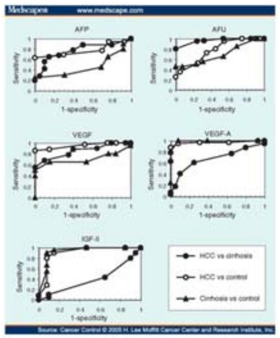
Isoeffektive Kurven auf ROC

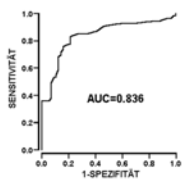


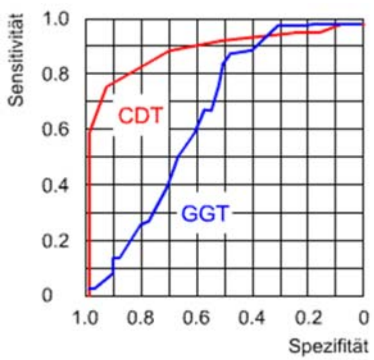
Ascites (+ Cholesterin, – CEA)



Weitere Beispiele



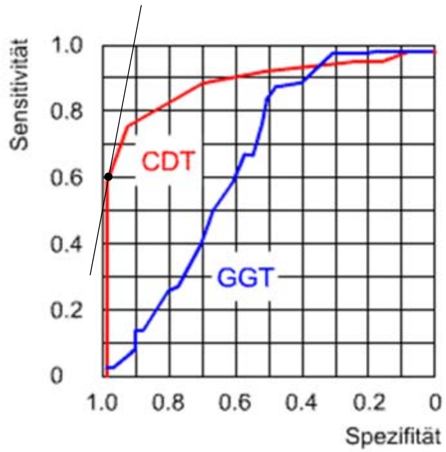




ROC für CDT (carbohydrate deficient transferrin) und GGT (gamma-Glutamyltransferase) in Bezug auf Alkoholismus. Da CDT praktisch immer auf der linken, oberen Seite der GGT liegt, ist CDT ein wesentlich besser Test für Alkoholkonsum als GGT

Beispiel: maximalisieren wir die diagnostische Effektivität!

bei einem kleineren Prävalenzwert ist die CDT Methode besser



bei einem höheren Prävalenzwert ist die GGT Methode besser