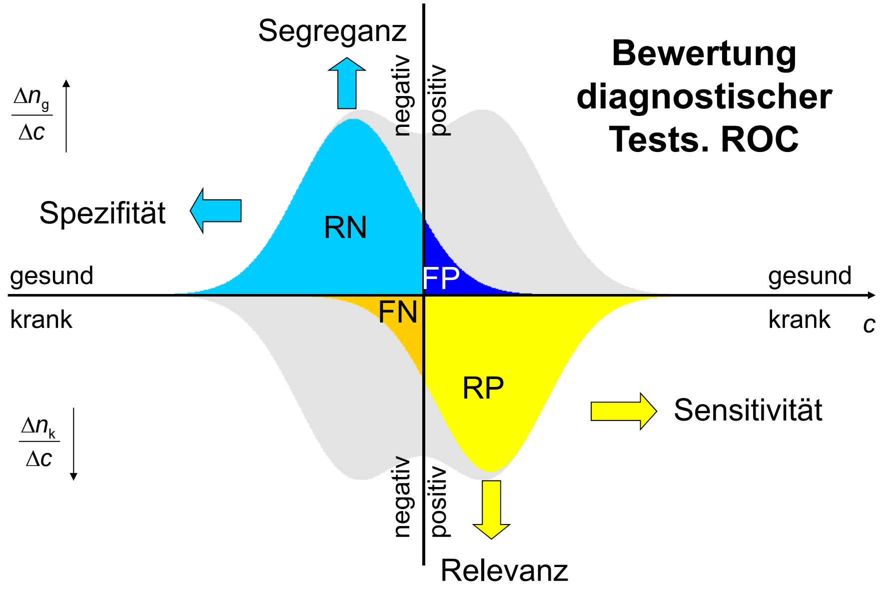


Bewertung diagnostischer Tests. ROC

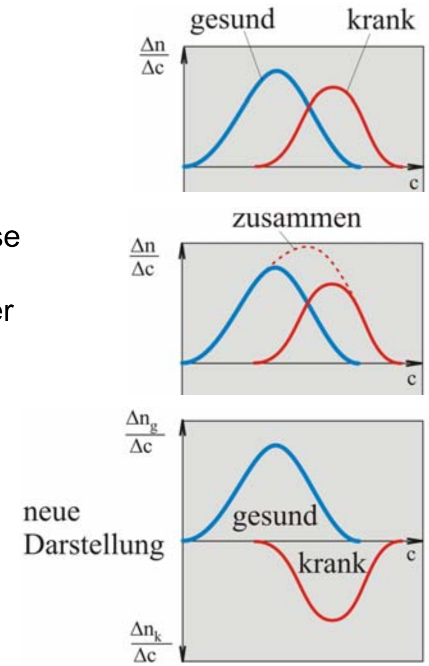
KAD 2021.03.23

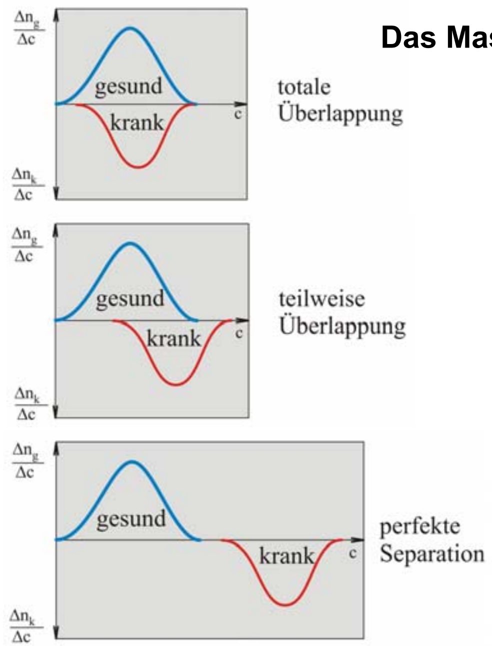
Überlappende Populationen

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

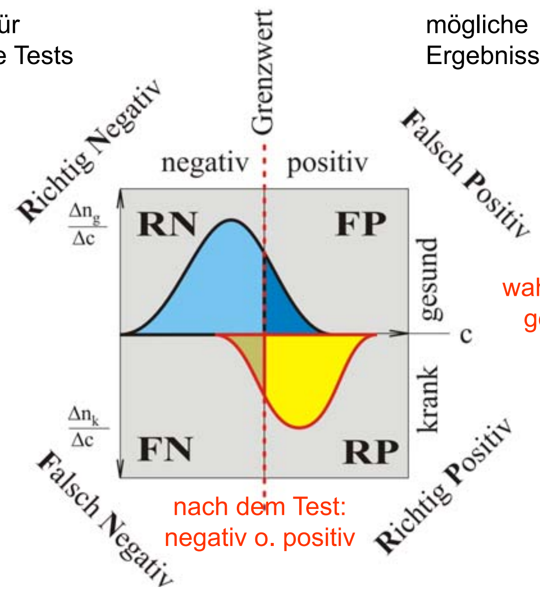
reelle Methoden

perfekte Methode

Wahrheitsmatrix

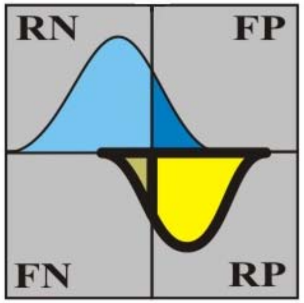
Basismatrix für diagnostische Tests

mögliche Ergebniskonstellationen

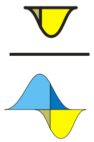


wahrer Zustand: gesund oder krank

Prävalenz

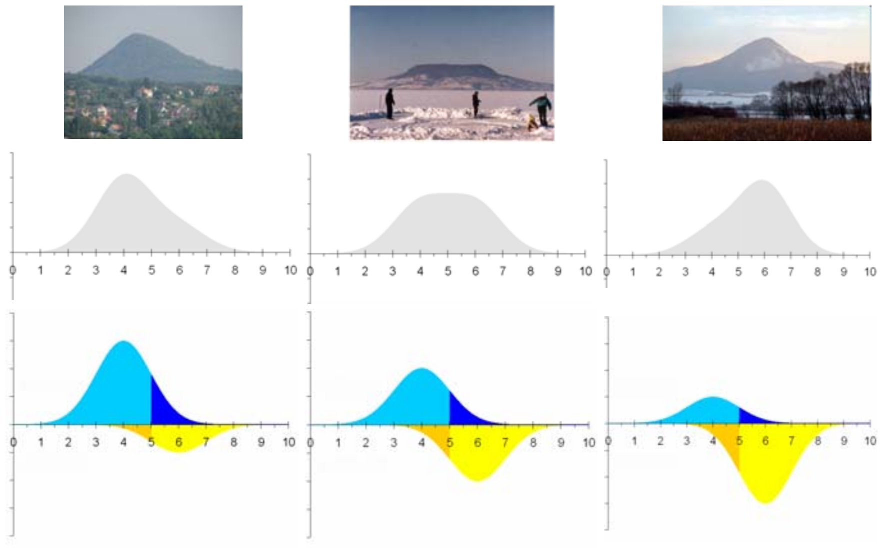


= Krankheitshäufigkeit
= Häufigkeit einer Krankheit in einer Population
= Wahrscheinlichkeit vor dem Test
= Vortestwahrscheinlichkeit
= a-priori-Wahrscheinlichkeit



$$= \boxed{w} = \frac{\text{alle Kranken}}{\text{alle Untersuchten}} = \boxed{\frac{RP + FN}{RP + RN + FN + FP}} = \frac{de - sp}{se - sp}$$

vgl.: Inzidenz = Anzahl der Neuerkrankungen pro Jahr/Monat/…
und pro 100 000/1000/… Einwohner



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

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

Sensitivität

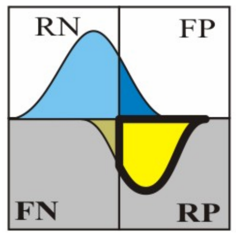
Spezifität

Relevanz

Segreganz

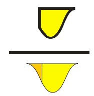
Jeder Test sollte an einem internationalen Standard geeicht werden, und es sollte eine **Referenzmethode** (Goldstandard) zur Erfassung des tatsächlichen Zustandes des Patienten verfügbar sein.

Diagnostische **Sensitivität**



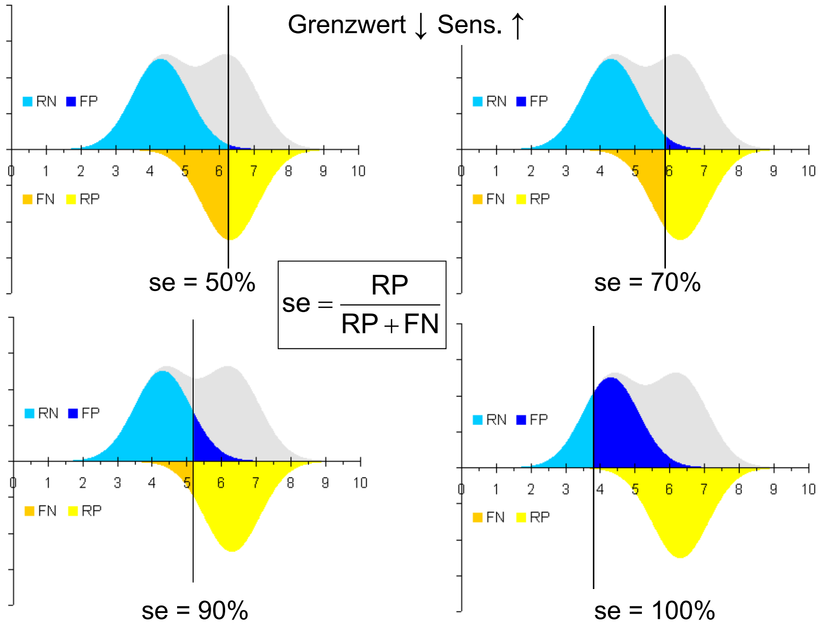
= Empfindlichkeit

= Wahrscheinlichkeit, einen Kranken als positiv zu erkennen



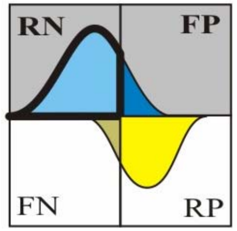
$$= \boxed{se} = \frac{\text{richtig positiv}}{\text{krank}} = \boxed{\frac{RP}{RP + FN}} = p(\text{positiv}|\text{krank})$$

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

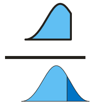


$$se = \frac{RP}{RP + FN}$$

Diagnostische **Spezifität**

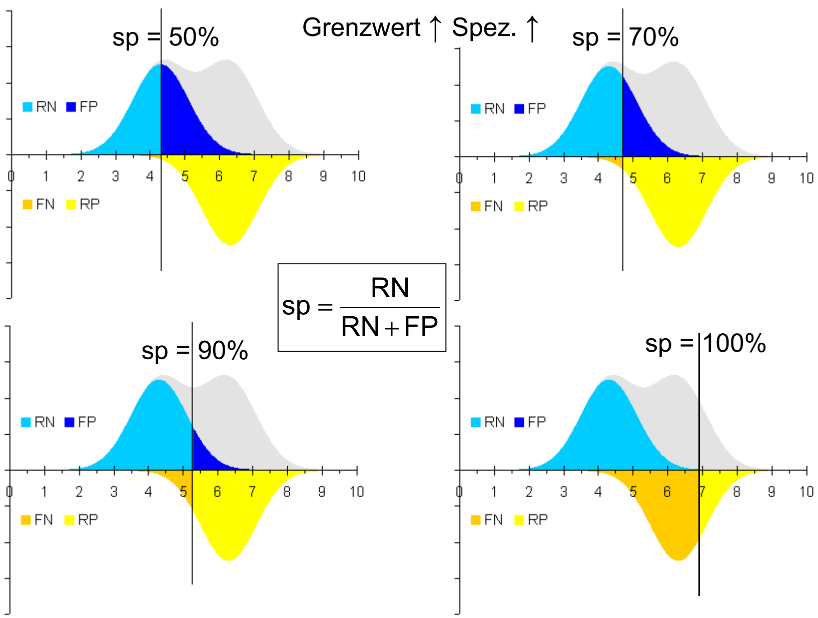


= Wahrscheinlichkeit, einen Gesunden als negativ zu erkennen



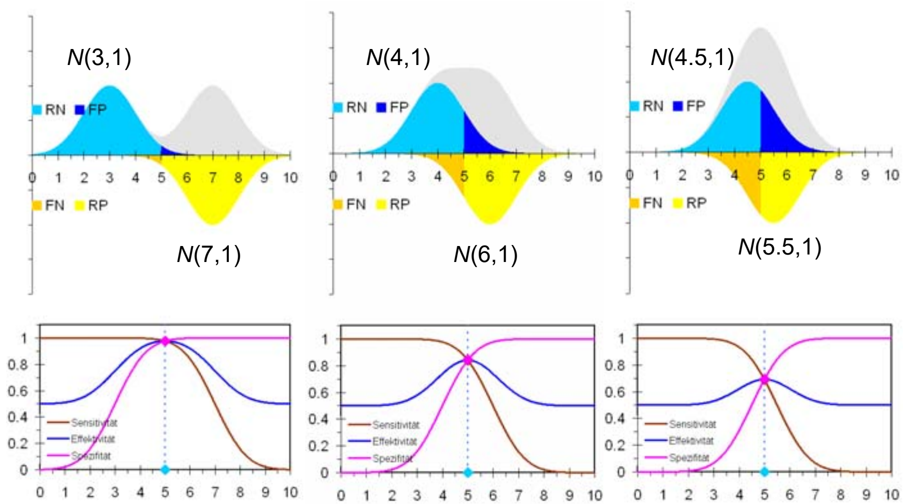
$$= sp = \frac{\text{richtig negativ}}{\text{gesund}} = \boxed{\frac{RN}{RN + FP}} = p(\text{negativ}|\text{gesund})$$

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



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

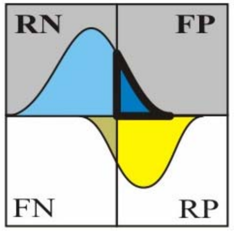
Sensitivität u. Spezifität: gegenläufige Eigenschaften von Testen

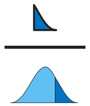


grössere Überlappung der Verteilungen ⟶

Diagnostische Falschpositivrate

(vgl. Fehler 1. Art)

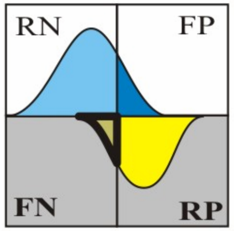


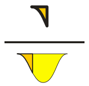


$$= 1 - \mathrm{sp} = \frac{\mathrm{FP}}{\mathrm{gesund}} = \frac{\mathrm{FP}}{\mathrm{RN} + \mathrm{FP}} = p(\mathrm{positiv}|\mathrm{gesund})$$

Diagnostische Falschnegativrate

(vgl. Fehler 2. Art)

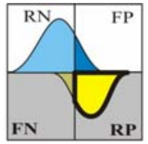


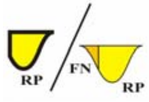


$$= 1 - \mathrm{se} = \frac{\mathrm{FN}}{\mathrm{krank}} = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{RP}} = p(\mathrm{negativ}|\mathrm{krank})$$

Horizontale Raten hängen von der Prävalenz nicht ab

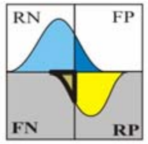
Sensitivität
(se)

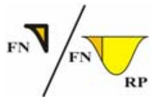




$$\mathrm{se} = \frac{\mathrm{RP}}{\mathrm{RP} + \mathrm{FN}}$$

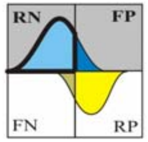
Falschnegativrate
(1-se)

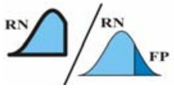




$$1 - \mathrm{se} = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{RP}}$$

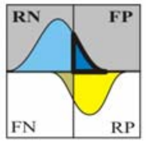
Spezifität
(sp)

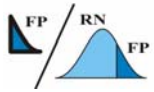




$$\mathrm{sp} = \frac{\mathrm{RN}}{\mathrm{RN} + \mathrm{FP}}$$

Falschpositivrate
(1-sp)





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

Vorhersagewerten (prädiktive Werte, vertikale Raten)

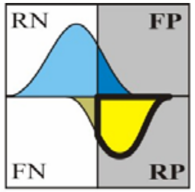
Wahrscheinlichkeiten nach dem Test, Nachtestwahrscheinlichkeiten,
a-posteriori-Wahrscheinlichkeiten

Diagnostische **Relevanz**

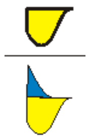
= positiv prädiktiver Wert

= positiver Vorhersagewert

= positive predictive value, PPV



Wahrscheinlichkeit eines Test-Positiven, krank zu sein



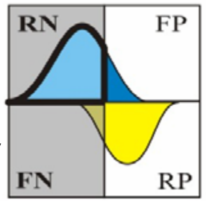
$$= \boxed{\mathrm{PPV}} = \frac{\mathrm{RP}}{\mathrm{positiv}} = \boxed{\frac{\mathrm{RP}}{\mathrm{RP} + \mathrm{FP}}} = p(\mathrm{krank}|\mathrm{positiv}) = \frac{\mathrm{se} \cdot \mathrm{w}}{\mathrm{se} \cdot \mathrm{w} + (1 - \mathrm{sp}) \cdot (1 - \mathrm{w})}$$

Diagnostische **Segreganz**

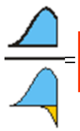
= negativ prädiktiver Wert

= negativer Vorhersagewert

= negative predictive value, NPV

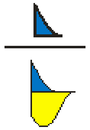


Wahrscheinlichkeit eines Test-Negativen, gesund zu sein

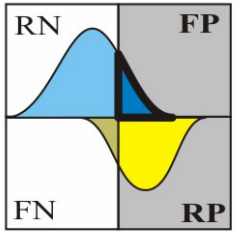


$$= \boxed{\mathrm{NPV}} = \frac{\mathrm{RN}}{\mathrm{negative}} = \boxed{\frac{\mathrm{RN}}{\mathrm{RN} + \mathrm{FN}}} = p(\mathrm{gesund}|\mathrm{negative}) = \frac{\mathrm{sp} \cdot (1 - \mathrm{w})}{\mathrm{sp} \cdot (1 - \mathrm{w}) + (1 - \mathrm{se}) \cdot \mathrm{w}}$$

Falschalarm(rate)

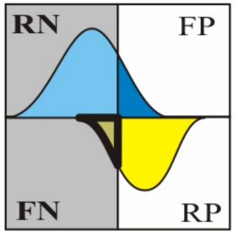


$$= 1 - \text{PPV} = \frac{\text{FP}}{\text{positiv}} = \frac{\text{FP}}{\text{FP} + \text{RP}} = p(\text{gesund}|\text{positiv})$$



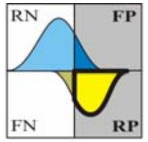
Falsche Beruhigung(srate)

$$= 1 - \text{NPV} = \frac{\text{FN}}{\text{negativ}} = \frac{\text{FN}}{\text{FN} + \text{RN}} = p(\text{krank}|\text{negativ})$$



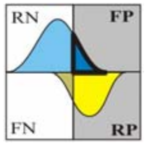
vertikale Raten hängen von der Prävalenz ab

Relevanz (PPV)



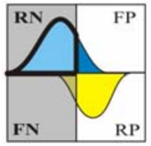
$$\text{PPV} = \frac{\text{RP}}{\text{FP} + \text{RP}}$$

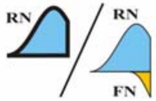
Falschalarmrate (1-PPV)



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

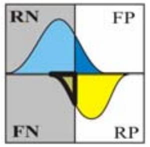
Segreganz (NPV)





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

falsche Beruhigung (1-NPV)

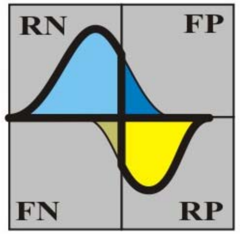


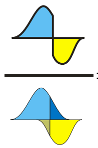
$$1\text{-NPV} = \frac{\text{FN}}{\text{RN} + \text{FN}}$$

Diagnostische **Effektivität**

= richtige Klassifikationsrate

= accuracy





$$= \boxed{\text{de}} = \frac{\text{RP} + \text{RN}}{\text{alle Untersuchten}} = \boxed{\frac{\text{RP} + \text{RN}}{\text{RN} + \text{FP} + \text{FN} + \text{RP}}} = \text{se} \cdot \text{w} + \text{sp} \cdot (1 - \text{w})$$

oft: Grenzwert ist so gewählt, dass Effektivität maximal ist

Effekt der Prävalenz

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

NPV = 90%

sp = 90%

		Test: negativ	Test: positiv
Goldstandard	gesund	90	10
	krank	10	90

se = 90%

(de = 90%)

PPV = 90%

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

NPV = 99%

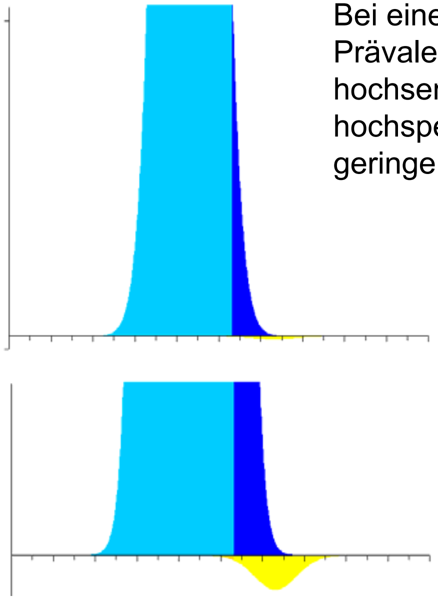
sp = 90%

		Test: negativ	Test: positiv
Goldstandard	gesund	810	90
	krank	10	90

se = 90%

(de = 90%)

PPV = 50%



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

Prävalenz = 0.1 %

Sensitivität = 98 %

Spezifität = 98 %

↓

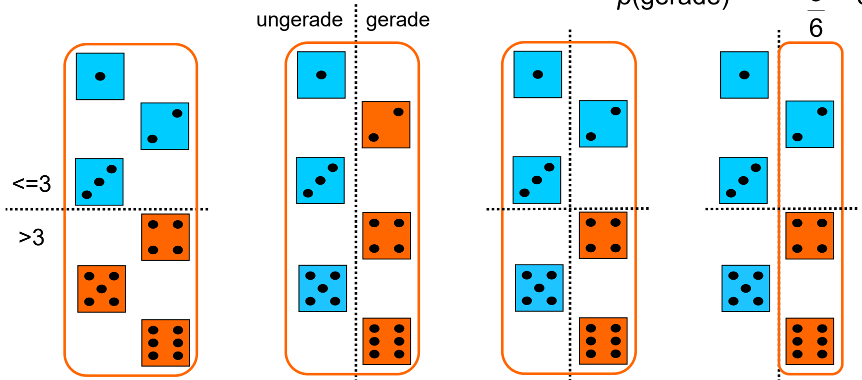
Relevanz = 4 %

Bedingte Wahrscheinlichkeit

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

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

zB: bei Würfelexperiment: $p(>3|\text{gerade}) = \dfrac{p(>3 \text{ und gerade})}{p(\text{gerade})} = \dfrac{\frac{2}{6}}{\frac{3}{6}} = \dfrac{2}{3}$



$p(>3) = \frac{3}{6} = \frac{1}{2}$ $\quad p(\text{gerade}) = \frac{3}{6} = \frac{1}{2}$ $\quad p(>3 \text{ und gerade}) = \frac{2}{6}$ $\quad p(>3 \mid \text{gerade}) = \frac{2}{3} = \frac{4}{6}$

*oder: *p*(*A* gegeben *B*), *p* von *A* vorausgesetzt *B*

Übersichtstabelle

← bedingte Wahrscheinlichkeit (Bayes)

Sensitivität	se	$\frac{RP}{RP+FN}$	$p(P\|K)$	Testpositiven zw. den Kranken	Richtigpositivrate, Empfindlichkeit	hängen von der Prävalenz nicht ab
Spezifität	sp	$\frac{RN}{RN+FP}$	$p(N\|G)$	Testnegativen zw. den Gesunden	Richtignegativ-rate	
Falschnegativrate	1-se	$\frac{FN}{RP+FN}$	$p(N\|K)$	Testnegativen zw. den Kranken		
Falschpositivrate	1-sp	$\frac{FP}{RN+FP}$	$p(P\|G)$	Testpositiven zw. den Gesunden		
Relevanz; positiver prädiktiver Wert	PPV	$\frac{RP}{RP+FP}$	$p(K\|P)$	Kranken zw. den Testpositiven	positiver Vorhersagewert	Prävalenzabhängigkeit
Segreganz; negativer prädiktiver Wert	NPV	$\frac{RN}{RN+FN}$	$p(G\|N)$	Gesunden zw.den Testnegativen	negativer Vorhersagewert	
Falschalarmrate	1-PPV	$\frac{FP}{RP+FP}$	$p(G\|P)$	Gesunden zw. den Testpositiven	Fehlalarmrate	
falsche Beruhigungsrate	1-NPV	$\frac{FN}{RN+FN}$	$p(K\|N)$	Kranken zw. den Tesnegativen		

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

Sensitivität

(1) Der Satz der totalen Wahrscheinlichkeit:

Wenn die Ereignisse $B_1, B_2, \ldots, B_n$ ein vollständiges Ereignissystem bilden $p(B_i) > 0$, $i = 1, 2, \ldots, n$, dann

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

B_1: krank
B_2: nicht krank
A: positiv

z.B.

$$p(A) = p(A|B_1) \cdot p(B_1) + p(A|B_2) \cdot p(B_2) = se \cdot w + (1-sp) \cdot (1-w) =$$

$$= \frac{RP}{RP+FN} \cdot \frac{RP+FN}{RP+FN+RN+FP} + \frac{FP}{RN+FP} \cdot \frac{RN+FP}{RP+FN+RN+FP} = \frac{RP+FP}{RP+FN+RN+FP}$$

(2) Satz von Bayes:

Umgekehrt suchen wir unter der Annahme von Ereignis A nach der Wahrscheinlichkeit von Ereignis B_k.

B_1: krank
B_2: nicht krank
A: positiv

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

z.B.

$$p(\text{krank}|\text{positiv}) = \frac{p(\text{positiv}|\text{krank})\cdot p(\text{krank})}{p(\text{positiv}|\text{krank})\cdot p(\text{krank}) + p(\text{positiv}|\text{gesund})\cdot p(\text{gesund})} =$$

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

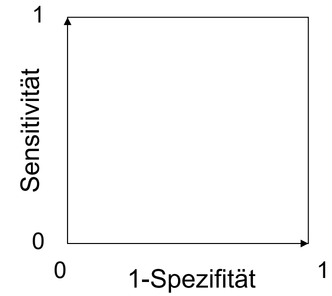
$$p(B_2|1-A) = \frac{p(1-A|B_2)\cdot p(B_2)}{p(1-A|B_1)\cdot p(B_1) + p(1-A|B_2)\cdot p(B_2)} = \boxed{\frac{sp\cdot(1-w)}{(1-se)\cdot w + sp\cdot(1-w)} = NPV}$$

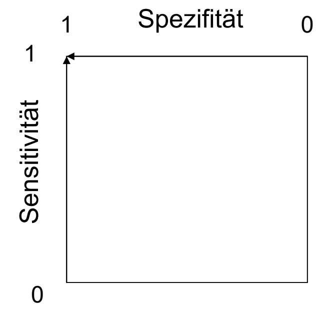
Vergleichung verschiedener diagnostischer Methode. ROC Kurven (Ergänzungsmaterial)

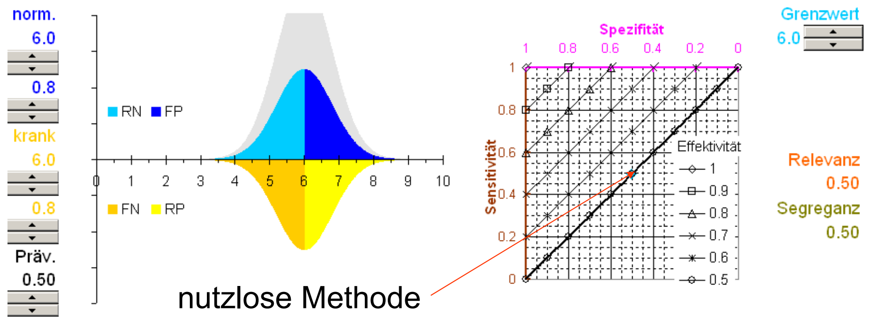
ROC: receiver-operator (operating) characteristic

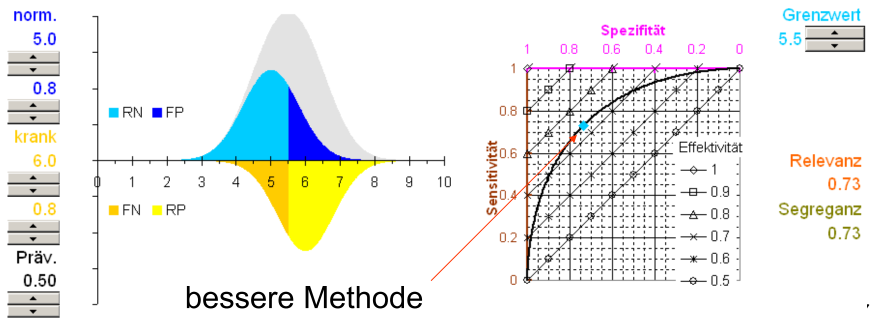
ca. 1950: erste ROC Analyse (receiver: Radar Empfänger)

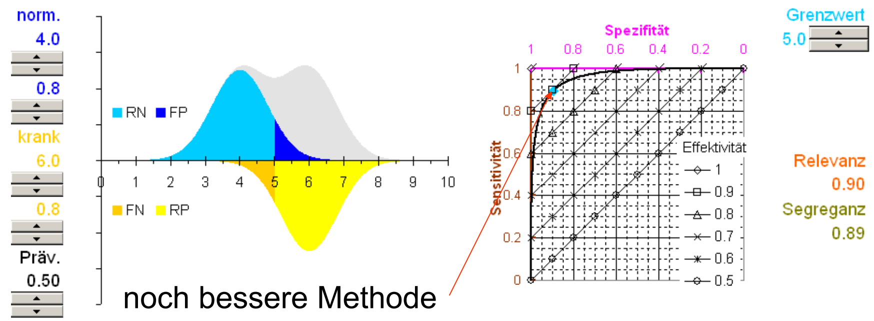
ca. 1970: die erste medizinische Anwendungen

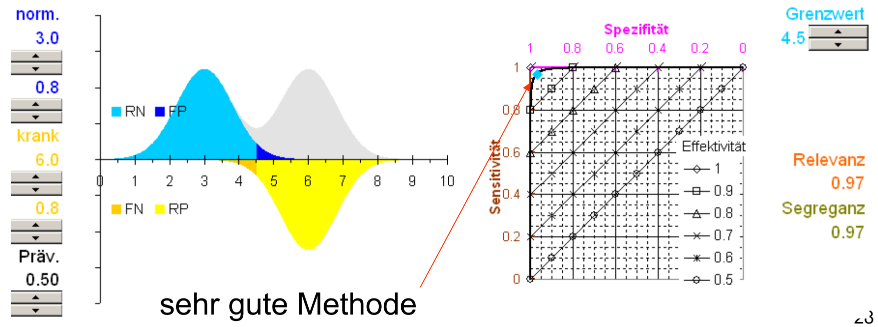




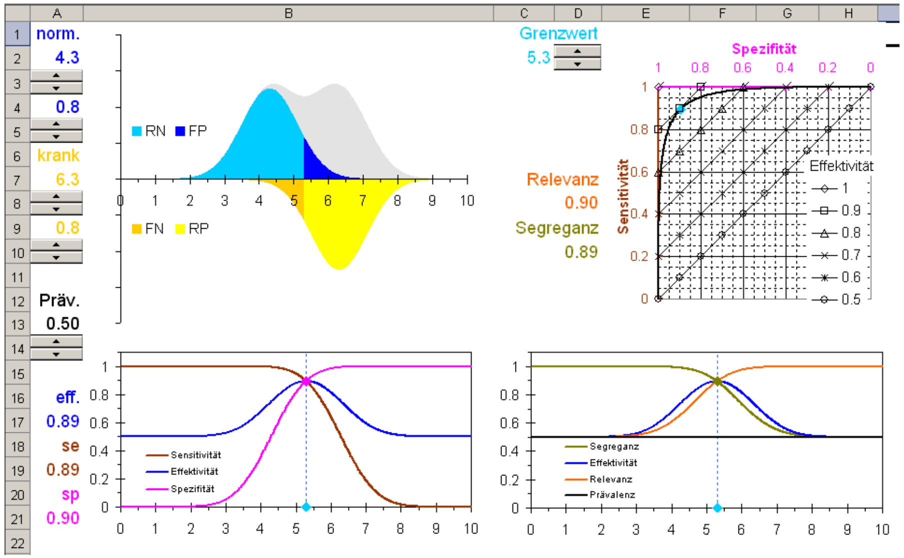






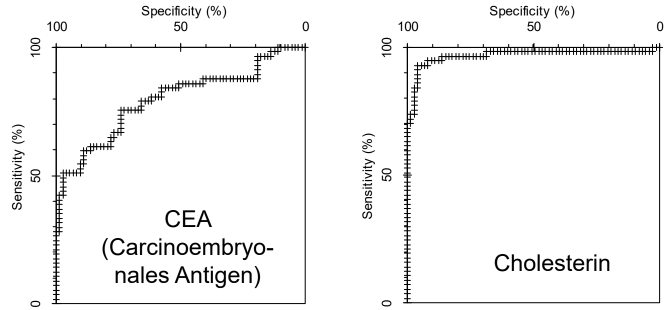


ROC Analyse



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

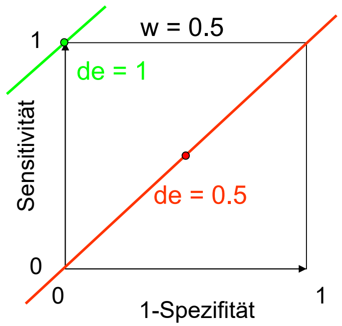
$$de = se \cdot w + sp \cdot (1-w)$$

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

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

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

Steigung 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.

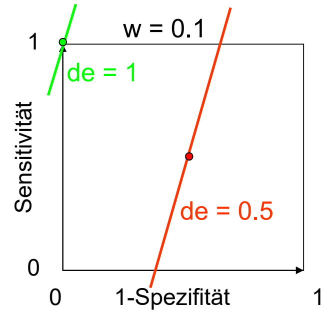
$$\boxed{se = \frac{1-w}{w}(1-sp) + \frac{1}{w} de + \frac{w-1}{w}}$$

Steigung Achsenabschnitt

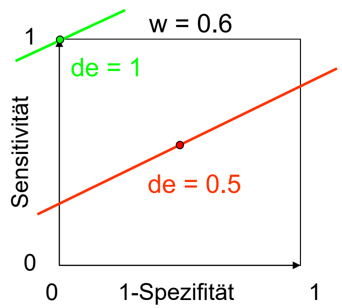
wenn w < 0.5: Die Steigung der Geraden mit gleicher diagnostischen Effektivität ist grösser als 1.

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

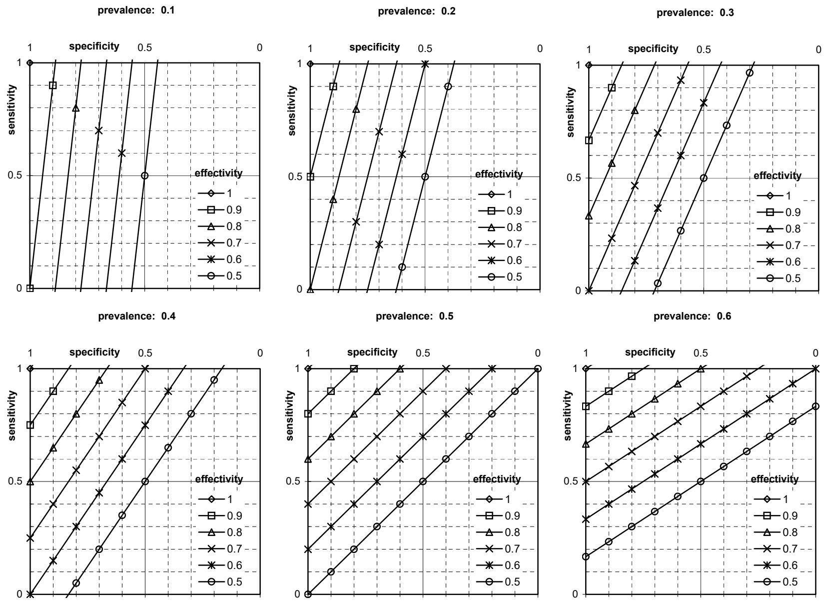
z.B. *w* = 0.1, die Steigung: 9



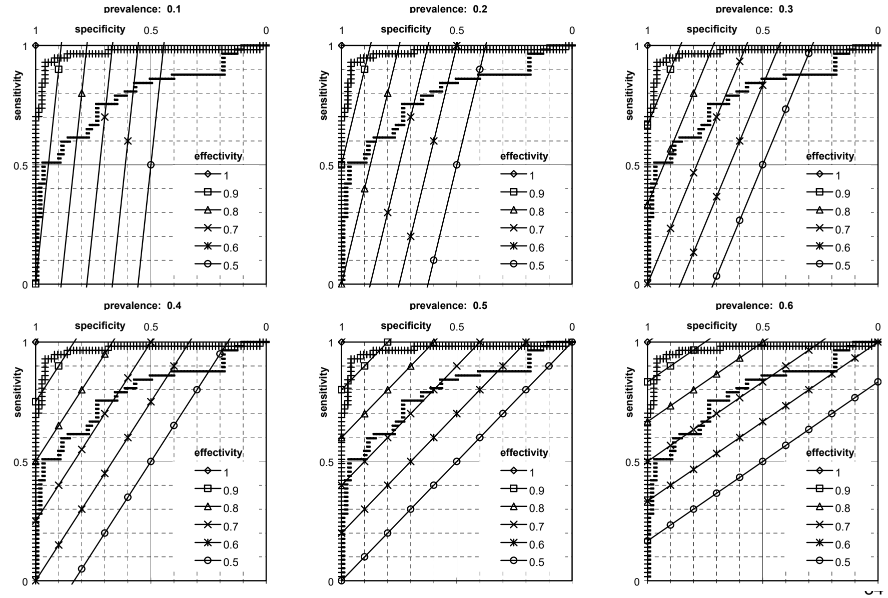
z.B. *w* = 0.6, Steigung: 0.66



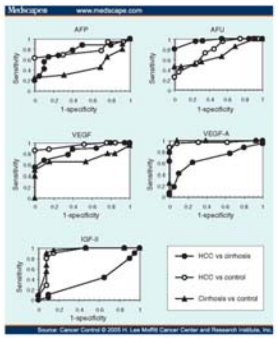
Isoeffektive Kurven auf ROC

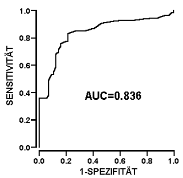


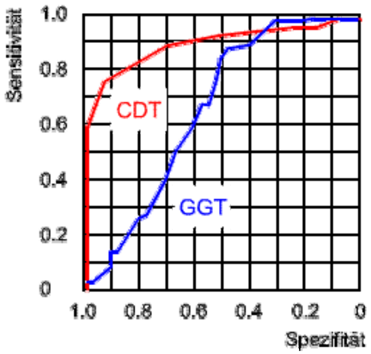
Ascites (+ Cholesterin, – CEA)



Weitere Beispiele



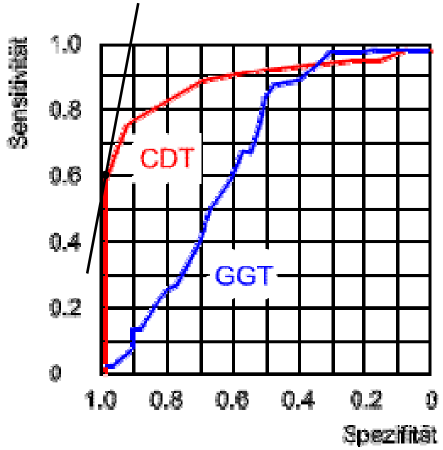




ROC für CDT (carbohydrate deficient transferrin) und GGT (gamma-Glutamyltransferase) in Bezug auf Alkoholismus. Da CDT praktisch immer auf der linken, oberen Seite der GGT liegt, ist CDT ein wesentlich besser Test für Alkoholkonsum als GGT

Beispiel: maximalisieren wir die diagnostische Effektivität!

bei einem kleineren Prävalenzwert ist die CDT Methode besser



bei einem höheren Prävalenzwert ist die GGT Methode besser