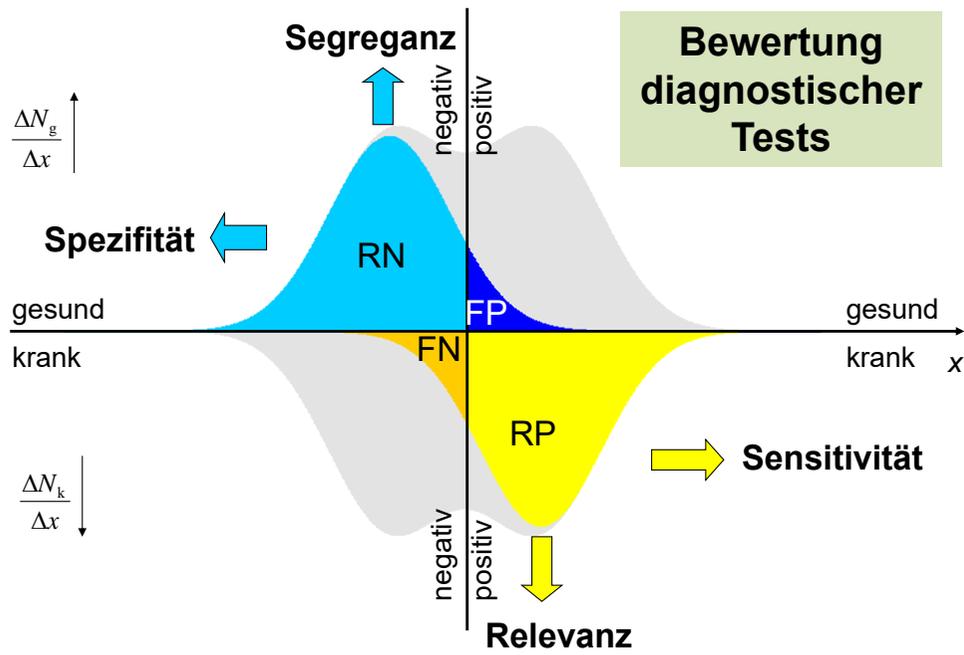
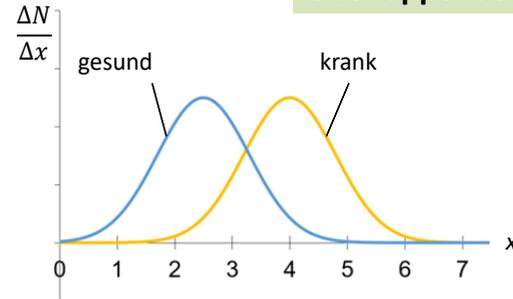


Bewertung diagnostischer Tests



KAD 2023.11.06

Überlappende Verteilungen

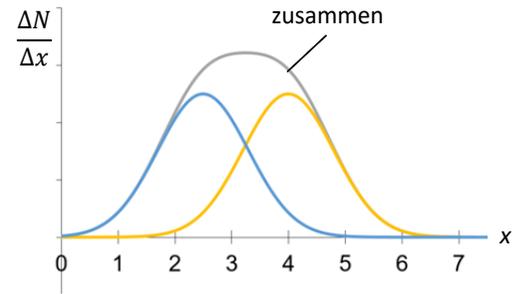


alles hat eine **Verteilung** (Augenfarbe, Größe, Cholesterinspiegel,...)

betrachten wir eine kontinuierliche Variable mit einer anderen Verteilung in der Krankenpopulation als in der gesunden

nehmen wir an, daß der messbare Parameter bei der kranken Population typischerweise **größer** ist als bei der gesunden

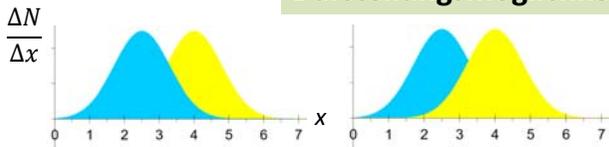
(wenn kleiner, können wir den Kehrwert des ursprünglichen Parameters verwenden)



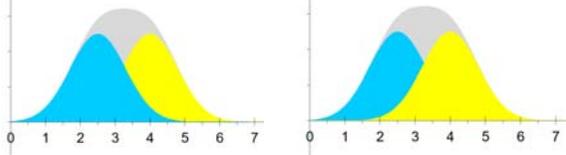
die Abbildung zeigt zwei solcher Dichtefunktionen; die Fläche unter der Kurve entspricht der Anzahl der Individuen

im vorliegenden Beispiel ist die Anzahl der Gesunden und Kranken gleich (die Standardabweichung des Parameters ist auch gleich)

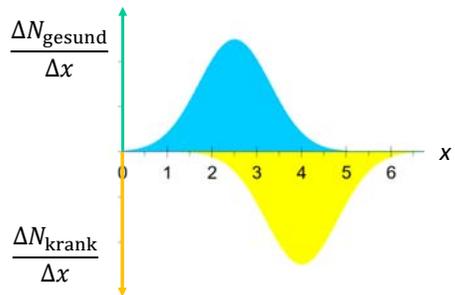
Darstellungsmöglichkeiten



aufgrund der großen Bedeutung der Bereiche unter der Kurve bevorzugen wir ein Bild, das die Bereiche einfärbt, anstelle einer Strichzeichnung



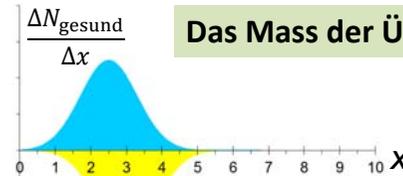
wegen der Überlappungen ist eine korrekte Farbgebung in der traditionellen Darstellung schwierig (oder unmöglich)



vorgeschlagene neue Version: anstatt der negativen Achse eine weitere positive Achse für die Kranken

Das Mass der Überlappung

egészs.
avg: 2,5
sd: 0,8
beteg
avg: 3,0
sd: 0,8
egészs.
avg: 2,5
sd: 0,8
beteg
avg: 5,5
sd: 0,8
egészs.
avg: 2,5
sd: 0,8
beteg
avg: 7,5
sd: 0,8



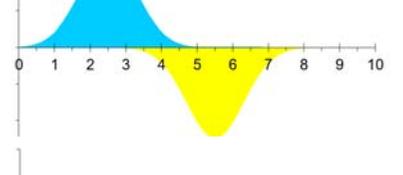
totale Überlappung

nutzlose Methode



teilweise Überlappung

(möglicherweise) verwendbare Methode



keine Überlappung

perfekte Methode

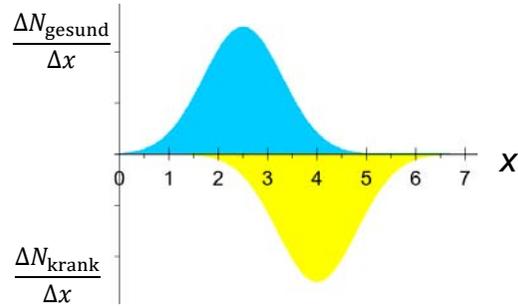
Häufigkeit einer Krankheit in einer Population

= Krankheitshäufigkeit

= Vortestwahrscheinlichkeit

= a-priori-Wahrscheinlichkeit

Prävalenz

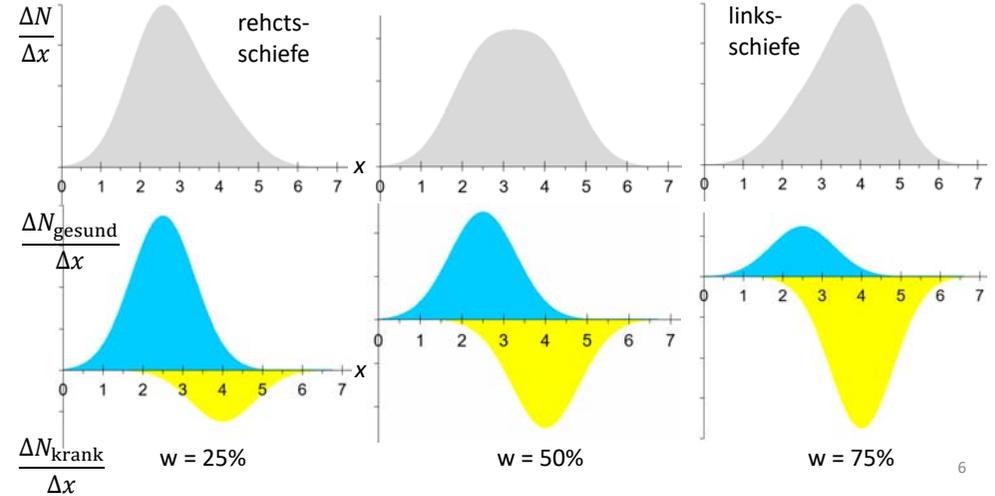


$$w = \frac{\text{alle Kranken}}{\text{alle Untersuchten}} = \frac{\text{alle Kranken}}{\text{alle Kranken} + \text{alle Gesunden}} = \frac{de - sp}{se - sp}$$

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

5

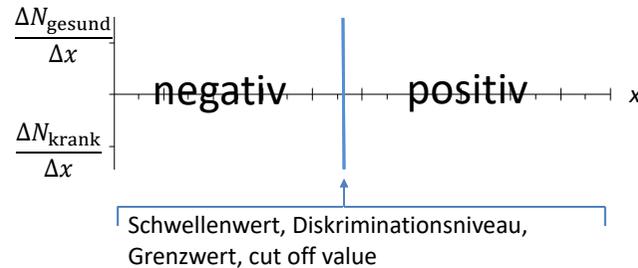
Auswirkung der Prävalenz auf die gemeinsame Verteilung



6

Ein negatives Testergebnis unterhalb der Schwelle und ein positives Testergebnis darüber

durch die Wahl eines **Schwellenwertes** unter den möglichen Messparameterwerten entscheiden wir, welche positiven Werte und welche negativ laut Testmethode



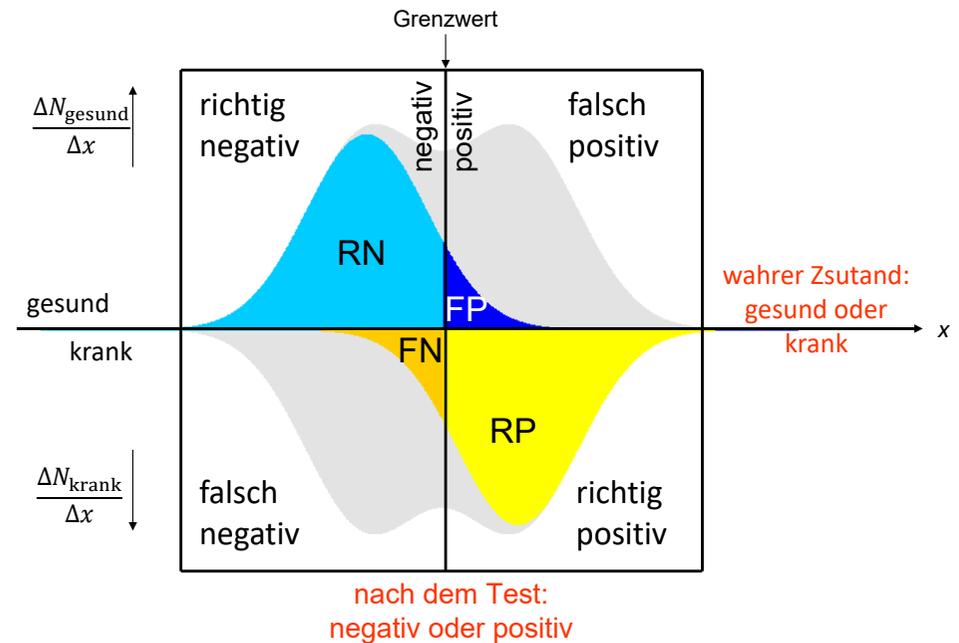
die **Anforderung**, dass krank und die positiv bzw. die gesund und die negativ passen so gut wie möglich zusammen

die **Klassifizierung** ist jedoch fast nie perfekt:

- solche Kranken, die positiv getestet wurden: richtig positiv, RP ✓
- solche Kranken, die negativ getestet wurden: falsch negativ, FN ☹️
- solche Gesunden, die negativ getestet wurden: richtig negativ, RN ✓
- solche Gesunden, die positiv getestet wurden: falsch positiv, FP ☹️

7

Wahrheitsmatrix



8

Die "Güte" eines diagnostischen Tests

diagnostische Tests unterteilen die Untersuchten in (Test)**positive** und (Test)**negative** Gruppen basierend auf einem (oder mehreren) gemessenen Parametern

die „Güte“ der Teilung kann **nicht durch eine einzige Zahl** charakterisiert werden

(a) inwieweit es diejenigen fangen, die **gefangen werden müssen**?

z.B. die Wahrscheinlichkeit, für eine Person, die mit einem COVID infiziert ist, zu behaupten, positiv zu sein

(b) inwieweit es diejenigen lässt, die **in Ruhe gelassen werden sollten**?

z.B. die Wahrscheinlichkeit, für eine Person, die nicht mit einem COVID infiziert ist, zu behaupten, negativ zu sein

(c) wie zuverlässig das **positive Testergebnis** ist?

bei positivem Testergebnis, wie sicher ist der Untersuchte krank

z.B. bei positivem COVID-Test, wie sicher ist, dass der Untersuchte mit COVID infiziert ist

(d) wie zuverlässig das **negative Testergebnis** ist?

bei negativem Testergebnis, wie sicher ist der Untersuchte gesund

z.B. bei negativem COVID-Test, wie sicher ist, dass der Untersuchte nicht mit COVID infiziert ist

9

Die Zuverlässigkeit der Tests kann durch folgende Kennwerten (Validitätsparameter) beschrieben werden:

Sensitivität

Spezifität

Relevanz

Segreganz

nur 3 unabhängige!

Jeder Test sollte an einem internationalen Standard geeicht werden, und es sollte eine Referenzmethode (Goldstandard) zur Erfassung des tatsächlichen Zustandes des Patienten verfügbar sein (manchmal nur das Ergebnis einer Autopsie)



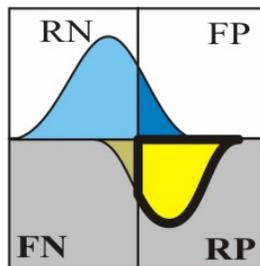
10

Diagnostische Sensitivität

= Empfindlichkeit

= richtig-positiv Rate

= sensitivity



Wahrscheinlichkeit, einen Kranken als positiv zu erkennen

positiv bei Kranken

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

Grenzwert ↓ Sensitivität ↑

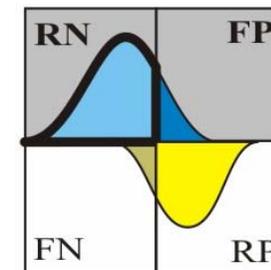
Tests mit hoher Sensitivität sind bei der Frühdiagnostik (screening) von Krankheiten erwünscht, und wenn es darauf ankommt, dass möglichst wenig Kranke unentdeckt bleiben.

11

Diagnostische Spezifität

= richtig-negativ Rate

= specificity



Wahrscheinlichkeit, einen Gesunden als negativ zu erkennen

negativ bei Gesunden

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

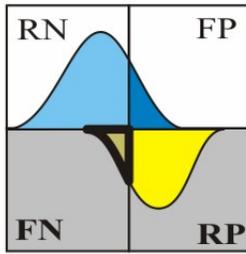
Grenzwert ↑ Spez. ↑

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

12

Falschnegativrate

= Fehler 2. Art
= false-negative rate/fraction



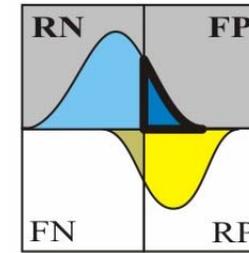
Wahrscheinlichkeit, einen Kranken als negativ zu erkennen
negativ bei Kranken

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

13

Falschpositivrate

= Ausfallquote, Ausfallrate
= Fehler 1. Art
= false-positive rate/fraction



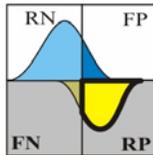
Wahrscheinlichkeit, einen Gesunden als positiv zu erkennen
positiv bei Gesunden

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

14

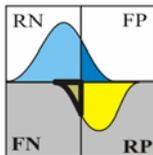
Horizontale Raten hängen von der Prävalenz nicht ab

Sensitivität
(se)



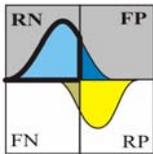
$$\text{se} = \frac{\text{RP}}{\text{RP} + \text{FN}}$$

Falschnegativrate
(1-se)



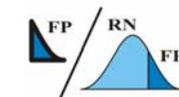
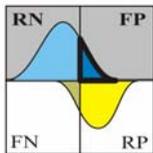
$$1 - \text{se} = \frac{\text{FN}}{\text{FN} + \text{RP}}$$

Spezifität
(sp)



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

Falschpositivrate
(1-sp)



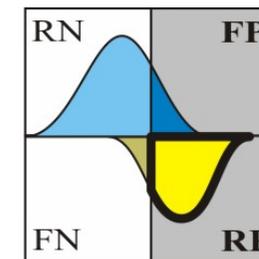
$$1 - \text{sp} = \frac{\text{FP}}{\text{RN} + \text{FP}}$$

15

Post-Test (a-posteriori) Wahrscheinlichkeiten

Diagnostische Relevanz

= positiv prädiktiver Wert
= positiver Vorhersagewert
= positive predictive value, PPV



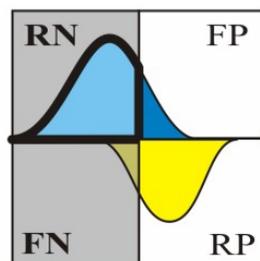
Wahrscheinlichkeit eines Test-Positiven, krank zu sein
krank zwischen den Positiven

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

16

Diagnostische Segreganz

= negativ prädiktiver Wert
 = negativer Vorhersagewert
 = negative predictive value, NPV



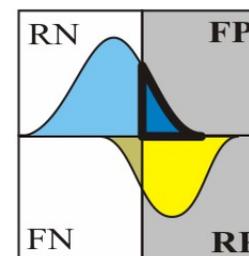
Wahrscheinlichkeit eines Test-Negativen, gesund zu sein
 gesund zwischen den Negativen

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

17

Falschalarm(rate)

= 1-PPV
 = false alarm rate



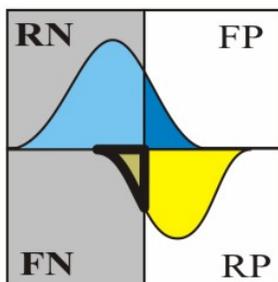
Wahrscheinlichkeit eines Test-Positiven, gesund zu sein
 gesund zwischen den Positiven

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

18

Falsche Beruhigung(rate)

= 1-NPV
 = false reassurance rate



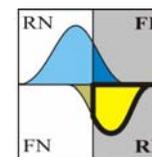
Wahrscheinlichkeit eines Test-Negativen, krank zu sein
 krank zwischen den Negativen

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

19

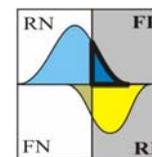
Vertikale Raten hängen von der Prävalenz ab

Relevanz (PPV)



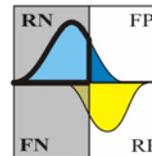
$$\text{PPV} = \frac{\text{RP}}{\text{FP} + \text{RP}}$$

Falschalarmrate (1-PPV)



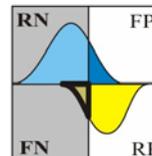
$$1 - \text{PPV} = \frac{\text{FP}}{\text{FP} + \text{RP}}$$

Segreganz (NPV)



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

falsche Beruhigung (1-NPV)

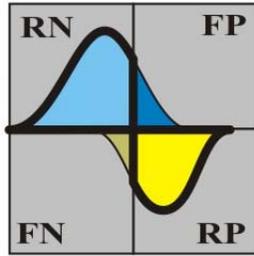


$$1 - \text{NPV} = \frac{\text{FN}}{\text{RN} + \text{FN}}$$

20

Diagnostische Effektivität

= richtige
Klassifikationsrate
= accuracy



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

oft: Grenzwert ist so gewählt, dass Effektivität maximal ist

21

Effekt der Prävalenz

NPV = 94,7%

Beispiel A: $w = 50\%$

sp = 90%

Gold- standard	gesund	90	10
	krank	5	95

(de = 92,5%)

se = 95%

PPV = 90,5%

NPV = 99,4%

Beispiel B: $w = 10\%$

sp = 90%

Gold- standard	gesund	810	90
	krank	5	95

(de = 90,5%)

se = 95%

PPV = 51,4%

22

Übersichtstabelle

Sensitivität	se	$\frac{RP}{RP + FN}$	$p(P K)$	Testpositiven zw. den Kranken	Richtigpositivrate, Empfindlichkeit
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
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 Testnegativen	

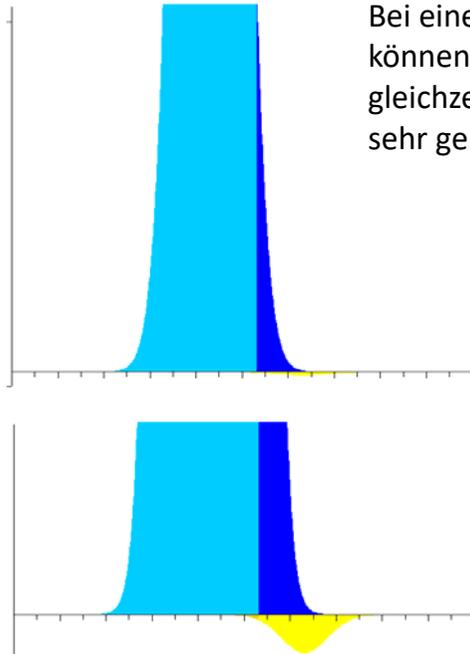
hängen von der Prävalenz nicht ab

Prävalenzabhängigkeit

bedingte Wahrscheinlichkeit (Bayes)

24

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



Prävalenz = 0.1 %

Sensitivität = 98 %

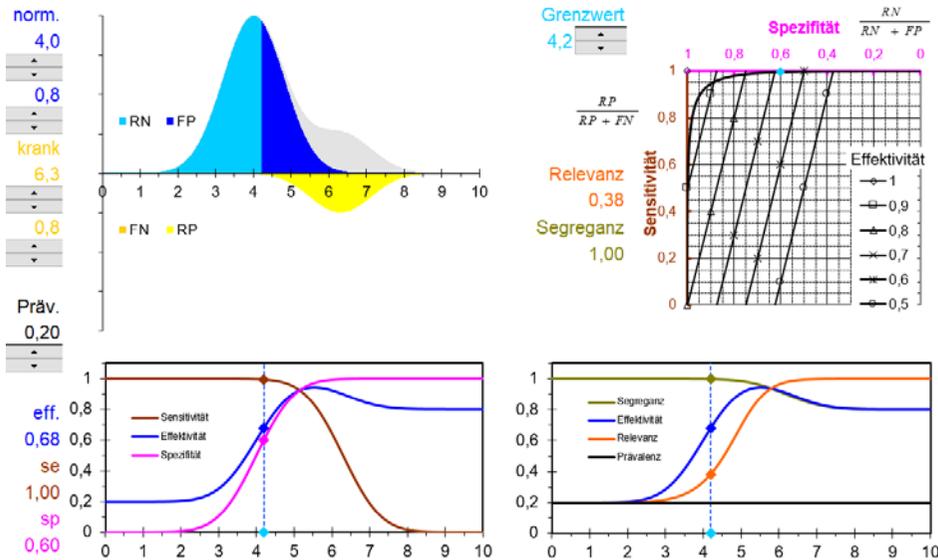
Spezifität = 98 %

↓
Relevanz = 4 %

23

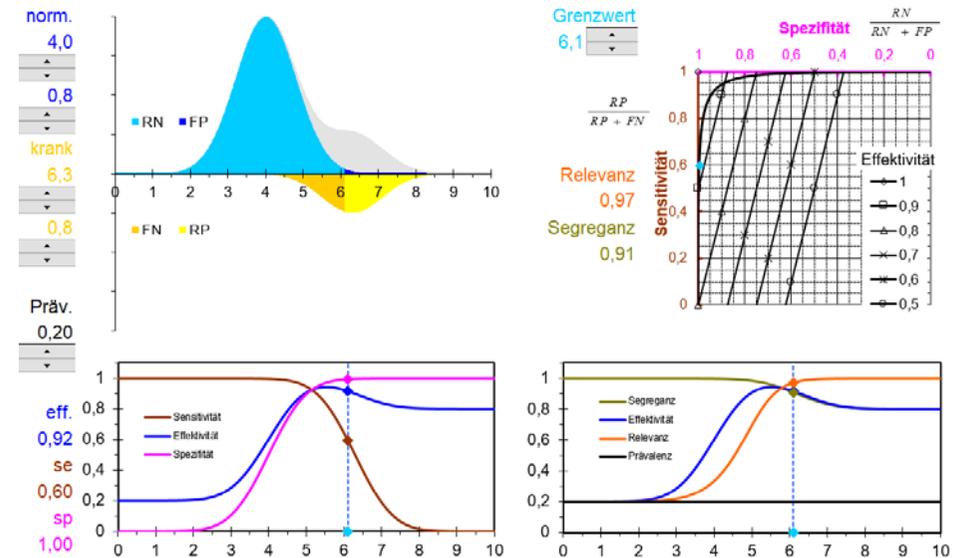
inwieweit es diejenigen fangen, die **gefangen werden müssen**?

Maximierung der diagnostischen Sensitivität



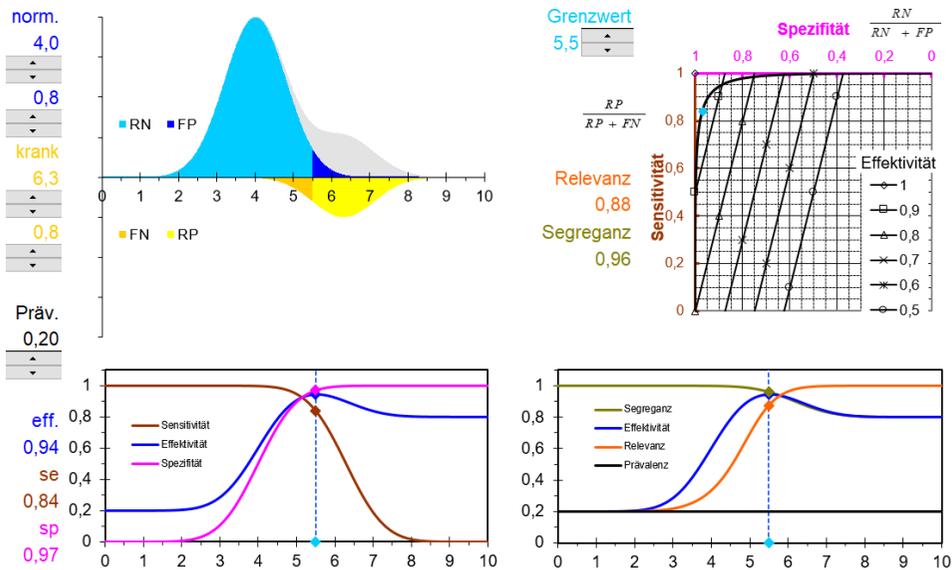
inwieweit es diejenigen lässt, die **in Ruhe gelassen werden sollten**?

Maximierung der diagnostischen Spezifität



genauso wichtig ist es: diejenigen fangen, die **gefangen werden müssen** und diejenigen lassen, die **in Ruhe gelassen werden sollten**

Maximierung der diagnostischen Effektivität



alles hat eine **Verteilung**
die Verteilung von Kranken und Gesunden **überlappen** sich
ob es möglich ist zu entscheiden, was wichtiger ist:

Botschaft zum Mitnehmen

die Krankheit bei so vielen Patienten wie möglich zu erkennen, damit sie behandelt werden können (**Maximierung der Sensitivität**) oder

bei möglichst wenigen Gesunden einen falsch-positiven Wert anzunehmen (Minimalisierung der Falschpositivrate oder **Maximierung der Spezifität**), damit sie keine unnötige Therapie erhalten

wenn sie nicht entschieden werden können, sind sie gleichermaßen wichtig: **Maximierung der Effektivität**

