

KAD 2011.11.15

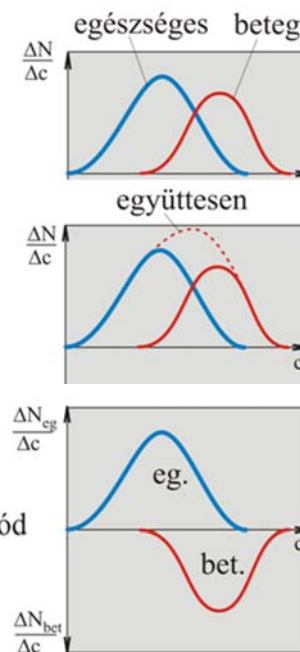
Átlapoló eloszlások

feltételezés:

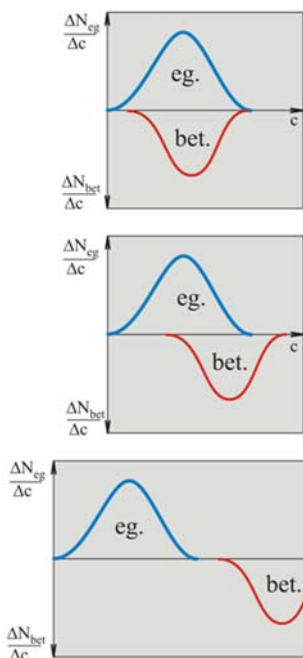
egy mérhető mennyiség
 (pl. koncentráció)
 megnövekszik a beteg
 populációban

(a megváltozás a lényeges
 és nem a megnövekedés)

új ábrázolásmód



2



Az átlapolás mértéke

teljes átfedés

használatatlan
 módszer

részleges átfedés

gyakorlati
 esetek

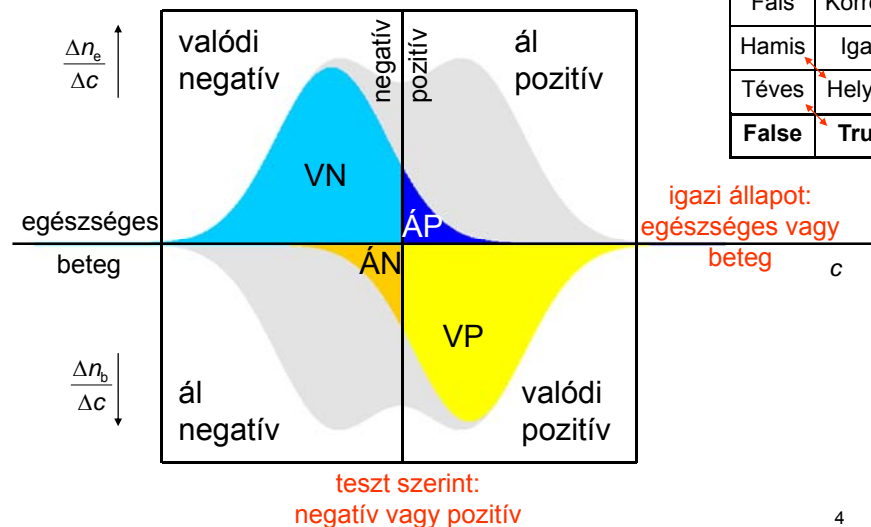
tisztá szeparáció

tökéletes
 módszer

3

Részleges átfedés

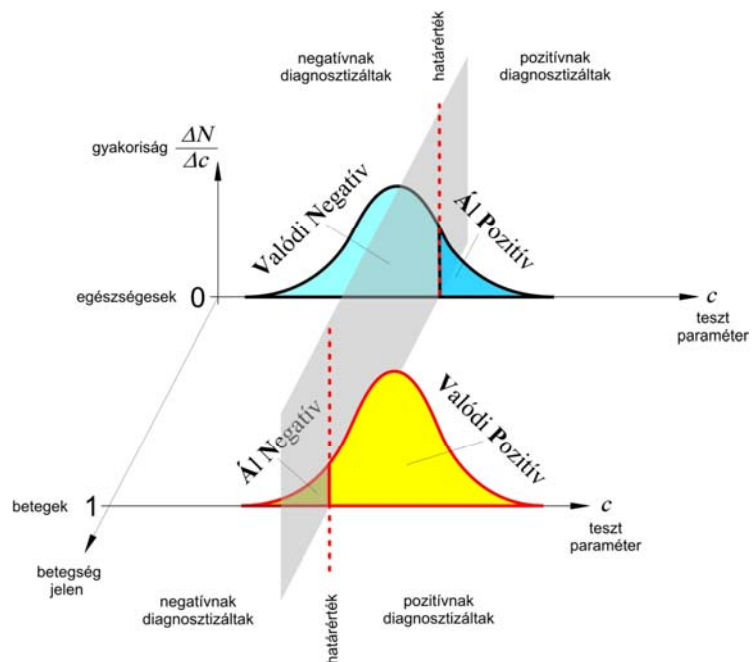
Igazságmátrix



elnevezés

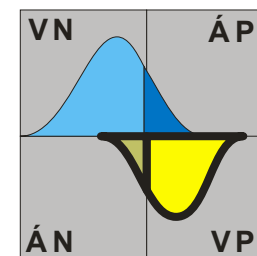
Ál	Valódi
Fals	Korrekt
Hamis	Igaz
Téves	Helyes
False	True

4



Prevalencia

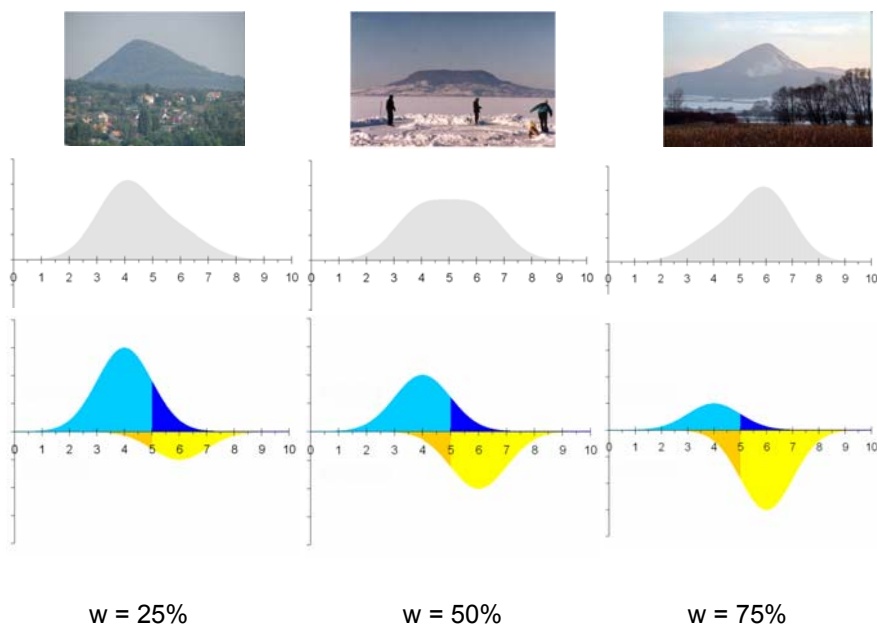
=elterjedtség
=a tesztet megelőző
valószínűség
= a-priori-
valószínűség



a betegség
gyakorisága a
vizsgált
populációban

$$\frac{\text{VN}}{\text{VN} + \text{ÁN}} = w = \frac{\text{beteg}}{\text{összes}} = \frac{\text{ÁN} + \text{VP}}{\text{VN} + \text{ÁP} + \text{ÁN} + \text{VP}} = \frac{de - sp}{se - sp}$$

6



7

A tesztek megbízhatósága a következő diagnosztikus paraméterekkel írható le:

szenzitivitás
specifitás
relevancia
szegregancia

csak 3 független!

Mindegyik teszt módszer összehasonlítható egy referencia módszerrel („Goldstandard”)

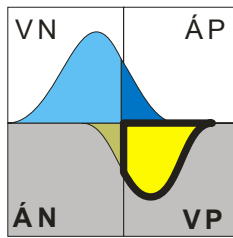


az ami biztosan jó, hiteles
(néha csak a boncolás
eredménye)

8

Diagnosztikus **szenzitivitás**

=érzékenység
=valódi pozitív
arány
=sensitivity



annak a valószínűsége, hogy
a teszt egy beteget
pozitívnak talál
pozitív a betegek
között

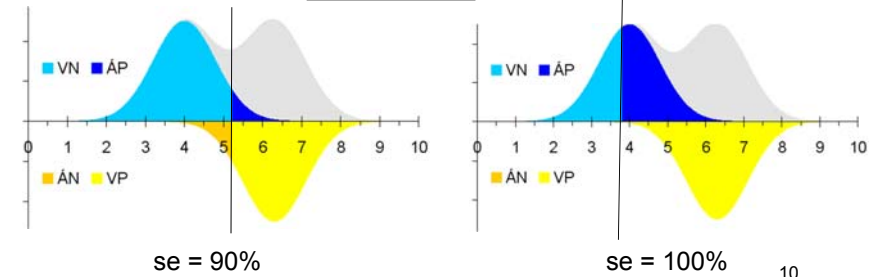
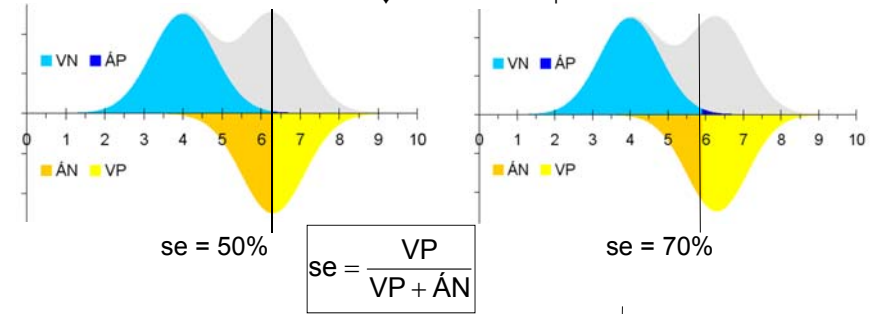
$P(\text{pozitív}|\text{beteg})$

$$\frac{\text{VP}}{\text{VP} + \text{ÁN}} = \text{se} = \frac{\text{VP}}{\text{VP} + \text{ÁN}}$$

Nagy szenzitivitású tesztek (közel 100%) a korai diagnózis
során kívánatosak (screening), ekkor kevés beteg marad
felismerés nélkül.

9

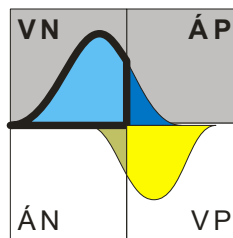
határérték ↓ szenzitivitás ↑



10

Diagnosztikus **specifitás**

=fajlagosság
=valódi negatív
arány
=specificity



annak a valószínűsége,
hogy a teszt egy
egészségeset
negatívnak talál
negatív az
egészségesek között

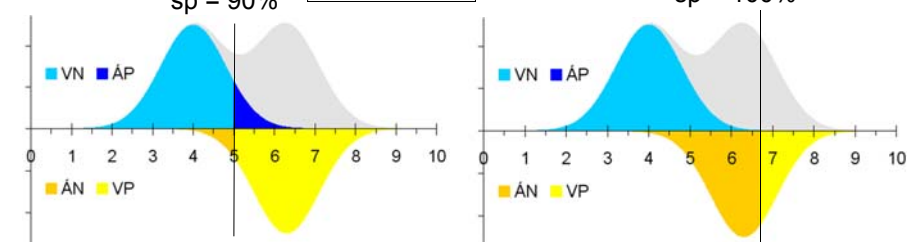
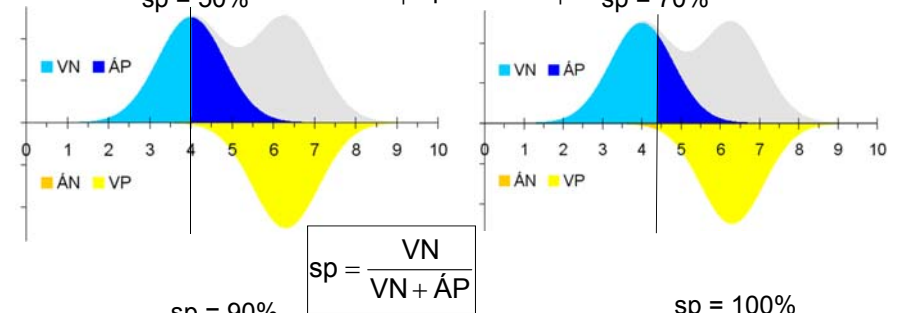
$P(\text{negatív}|\text{egészséges})$

$$\frac{\text{VN}}{\text{VN} + \text{ÁP}} = \text{sp} = \frac{\text{VN}}{\text{VN} + \text{ÁP}}$$

Magas specifitású tesztek (közel 100%) akkor fontosak, ha
az álpozitív értékek súlyos következménnyel járnak.

11

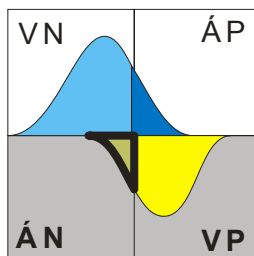
sp = 50% határérték ↑ specifitás ↑ sp = 70%



12

Álnegatív arány

=másodfajú hiba
=false-negative
rate/fraction



annak a
valószínűsége, hogy a
teszt egy beteget
negatívnak talál

negatív a betegek
között

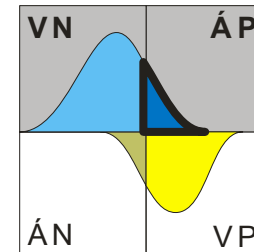
$$P(\text{negatív}|\text{beteg})$$

$$\frac{\text{ÁN}}{\text{ÁN} + \text{VP}} = 1 - \text{se} = \frac{\text{ÁN}}{\text{ÁN} + \text{VP}}$$

13

Álpozitív arány

=elsőfajú hiba
=false-positive
rate/fraction



annak a
valószínűsége, hogy
a teszt egy
egészségeset
pozitívnak talál

pozitív az
egészségesek között

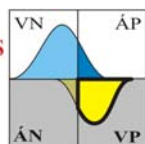
$$P(\text{pozitív}|\text{egészséges})$$

$$\frac{\text{ÁP}}{\text{VN} + \text{ÁP}} = 1 - \text{sp} = \frac{\text{ÁP}}{\text{VN} + \text{ÁP}}$$

14

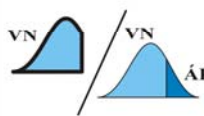
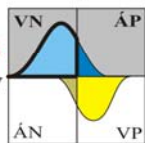
Arányok „vízszintesen” (prevalenciától függetlenek)

SZENZITIVITÁS
VALÓDI POZITÍV
ARÁNY
(se)



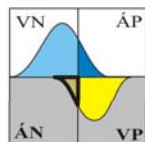
$$= \frac{\text{VN}}{\text{VN} + \text{VP}}$$

SPECIFICITÁS
VALÓDI NEGATÍV
ARÁNY
(sp)



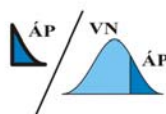
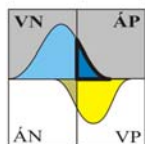
$$= \frac{\text{VN}}{\text{VN} + \text{ÁP}}$$

**ÁLNEGATÍV
ARÁNY**
másodfajú hiba



$$= \frac{\text{VP}}{\text{VN} + \text{VP}}$$

**ÁLPOZITÍV
ARÁNY**
elsőfajú hiba



$$= \frac{\text{ÁP}}{\text{VN} + \text{ÁP}}$$

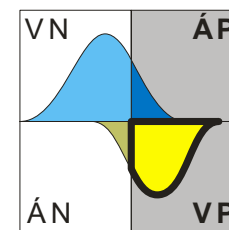
teszt utáni (a-posteriori) valószínűségek, a prevalenciától erősen függenek

Diagnosztikus relevancia

=korrekt pozitivitás

=pozitív prediktív
érték

=positive predictive
value =PPV



a betegség
jelenlétének a
valószínűsége,
ha a teszt pozitív

beteg a pozitívak
között

$$P(\text{beteg}|\text{pozitív})$$

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

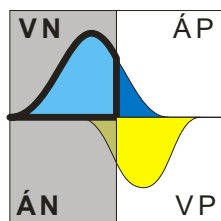
16

Diagnosztikus szegregancia

=korrekt negativitás

=negatív prediktív érték

=negative predictive value =NPV



a betegség hiányának a valószínűsége, ha a teszt negatív

egészséges a negatívak között

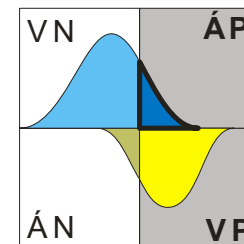
$P(\text{egészséges}|\text{negatív})$

$$\frac{\text{VN}}{\text{VN} + \text{ÁN}} = \text{NPV} = \frac{\text{VN}}{\text{összes negatív}} = \frac{\text{VN}}{\text{ÁN} + \text{VN}} = \frac{\text{sp} \cdot (1-w)}{\text{sp} \cdot (1-w) + (1-\text{se}) \cdot w}$$

17

Téves figyelemfelkeltő arány

=false alarm rate



a betegség hiányának a valószínűsége, ha a teszt pozitív

egészséges a pozitívak között

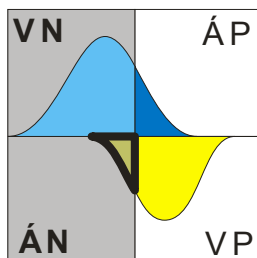
$P(\text{egészséges}|\text{pozitív})$

$$\frac{\text{ÁP}}{\text{ÁP} + \text{VP}} = 1 - \text{PPV} = \frac{\text{ÁP}}{\text{összes pozitív}} = \frac{\text{ÁP}}{\text{ÁP} + \text{VP}}$$

18

Téves megnyugtató arány

=false reassurance rate



a betegség jelenlétének a valószínűsége, ha a teszt negatív

beteg a negatívak között

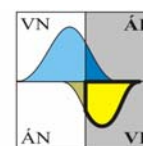
$P(\text{beteg}|\text{negatív})$

$$\frac{\text{ÁN}}{\text{ÁN} + \text{VN}} = 1 - \text{NPV} = \frac{\text{ÁN}}{\text{összes negatív}} = \frac{\text{ÁN}}{\text{ÁN} + \text{VN}}$$

19

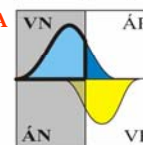
Arányok „fügőlegesen”

RELEVANCIA
KORREKT POZITIVITÁS pozitív prediktív érték



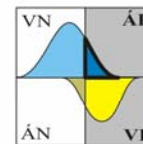
$$\frac{\text{VP}}{\text{ÁP} + \text{VP}} = \text{PPV} = \frac{\text{se} \cdot w}{\text{se} \cdot w + (1-\text{sp}) \cdot (1-w)}$$

SZEGREGANCIA
KORREKT NEGATIVITÁS negatív prediktív érték



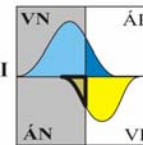
$$\frac{\text{VN}}{\text{VN} + \text{ÁN}} = \text{NPV} = \frac{\text{sp} \cdot (1-w)}{\text{sp} \cdot (1-w) + (1-\text{se}) \cdot w}$$

TÉVES FIGYELEM-FELKELTŐ ARÁNY



$$\frac{\text{ÁP}}{\text{VP} + \text{ÁP}} = 1 - \text{PPV}$$

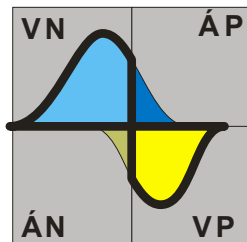
TÉVES MEGNYUGTATÁSI ARÁNY



$$\frac{\text{ÁN}}{\text{VN} + \text{ÁN}} = 1 - \text{NPV}$$

Diagnosztikus effektivitás

=accuracy



helyes besorolás aránya
(korrekt klasszifikáció)

$$de = \frac{VP + VN}{\text{összes}} = \frac{VP + VN}{VP + \text{ÁN} + VN + \text{ÁP}} = se \cdot w + sp \cdot (1 - w)$$

gyakran a határértéket úgy választjuk meg, hogy az effektivitás maximális legyen

21

Áttekintő tábla

szenzitivitás	se	$\frac{VP}{VP + \text{ÁN}}$	$p(P B)$	pozitív a betegek között	VPa (valódi pozitív arány)	prevalenciától függetlenek
specifitás	sp	$\frac{VN}{VN + \text{ÁP}}$	$p(N E)$	negatív az egészségesek között	VNa (valódi negatív arány)	
álnegatív arány	1-se	$\frac{\text{ÁN}}{VP + \text{ÁN}}$	$p(N B)$	negatív a betegek között	ÁNa	
álpozitív arány	1-sp	$\frac{\text{ÁP}}{VN + \text{ÁP}}$	$p(P E)$	pozitív az egészségesek között	ÁPa	
relevancia	PPV	$\frac{VP}{VP + \text{ÁP}}$	$p(B P)$	beteg a pozitívak között		a-posteriori valószínűségek
szegregancia	NPV	$\frac{VN}{VN + \text{ÁN}}$	$p(E N)$	egészséges a negatívak között		
téves figyelemfelkeltő arány	1-PPV	$\frac{\text{ÁP}}{VP + \text{ÁP}}$	$p(E P)$	egészséges a pozitívak között		
téves megnyugtató arány	1-NPV	$\frac{\text{ÁN}}{VN + \text{ÁN}}$	$p(B N)$	beteg a negatívak között		

A prevalencia hatása

Pl. A: **w = 50%**

NPV = 90%

sp = 90%

		teszt		
		negatív	pozitív	
Gold-standard	egészs.	90	10	se = 90%
	beteg	10	90	

(de = 90%) PPV = 90%

NPV = 99%

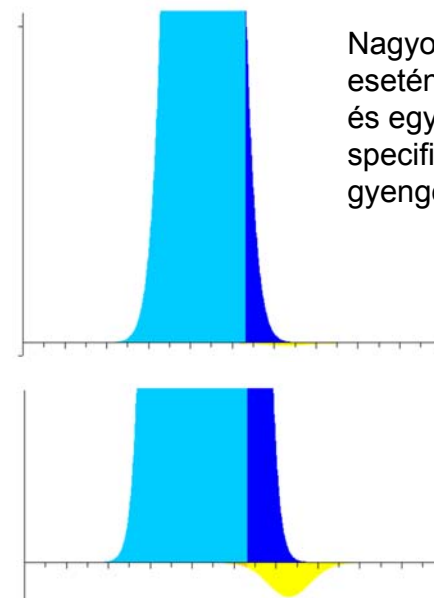
Pl. B: **w = 10%**

sp = 90%

		teszt		
		negatív	pozitív	
Gold-standard	egészs.	810	90	se = 90%
	beteg	10	90	

(de = 90%) PPV = 50%

23



Nagyon kicsi prevalencia esetén egy magasan szenzitív és egyidejűleg magasan specifikus teszt nagyon gyenge relevanciájú lehet

prevalencia = 0.1 %
szenzitivitás = 98 %
specifitás = 98 %
↓
relevancia = 4 %

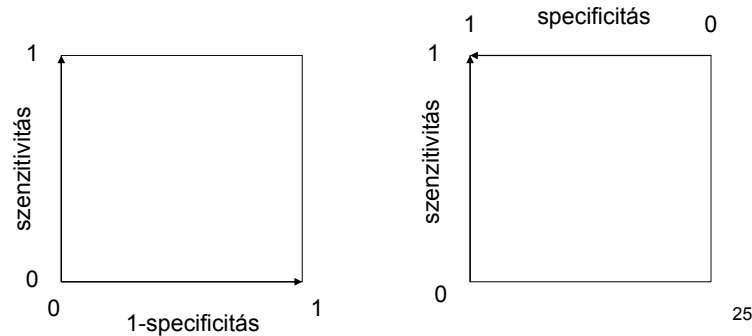
24

Különböző diagnosztikus módszerek összehasonlítása. ROC görbék

ROC: receiver-operator (operating) characteristic

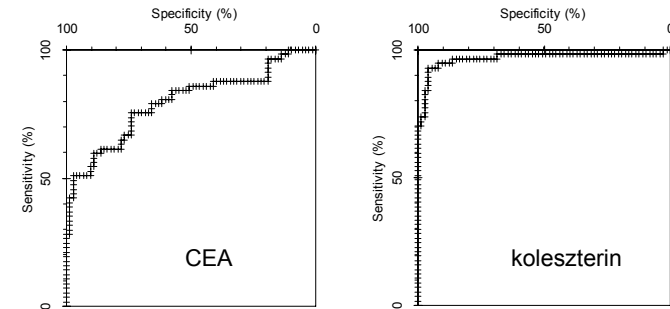
az első ROC az 1950-es években (receiver: radar vevő)

1970 körül az első orvosi alkalmazások



Hasüregi folyadékgyülemek daganatdiagnosztikája

a CEA és a koleszterin koncentráció megemelkedése
karcinózással kísért rákkal hozható összefüggésbe



Melyik módszer jobb? Hogyan célszerű kiválasztani a legjobb határértéket?

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

26

$$de = se \cdot w + sp \cdot (1 - w) \quad (3 \text{ független adat})$$

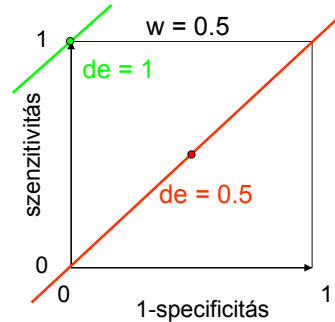
$$\frac{de}{1-w} = \frac{w}{1-w} se + (sp-1) + 1$$

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

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

meredekség

tengelymetszet



ha $w = 0.5$: $se = 1 \cdot (1-sp) + 2 \cdot de - 1$

Azok a pontok, amelyek azonos diagnosztikus effektivitásúak, olyan egyenesen vannak, amelynek a meredeksége egy.

Ha a $de = 0.5$, akkor a tengelymetszet: 0;

Ha a $de = 1$, akkor a tengelymetszet: 1.

27

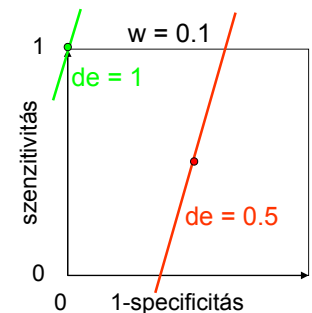
$$se = \frac{1-w}{w} (1-sp) + \frac{1}{w} de + \frac{w-1}{w}$$

meredekség tengelymetszet

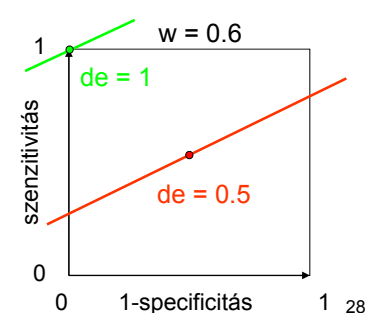
ha $w < 0.5$: azonos diagnosztikus effektivitás esetén az egyenes meredeksége nagyobb mint 1.

ha $w > 0.5$: azonos diagnosztikus effektivitás esetén az egyenes meredeksége kisebb mint 1.

pl. $w = 0.1$, meredekség: 9

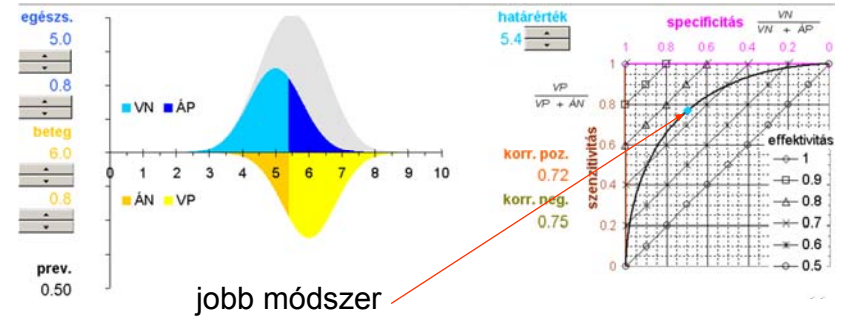
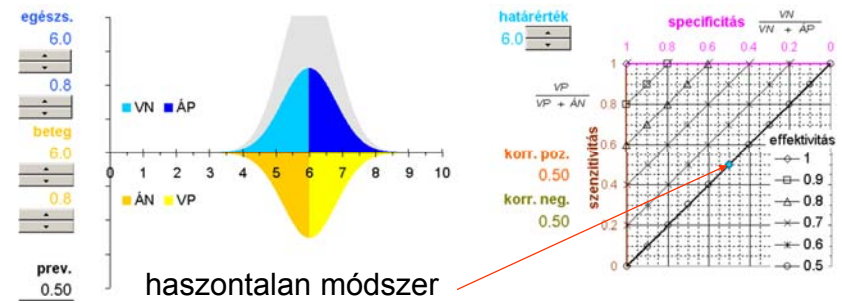
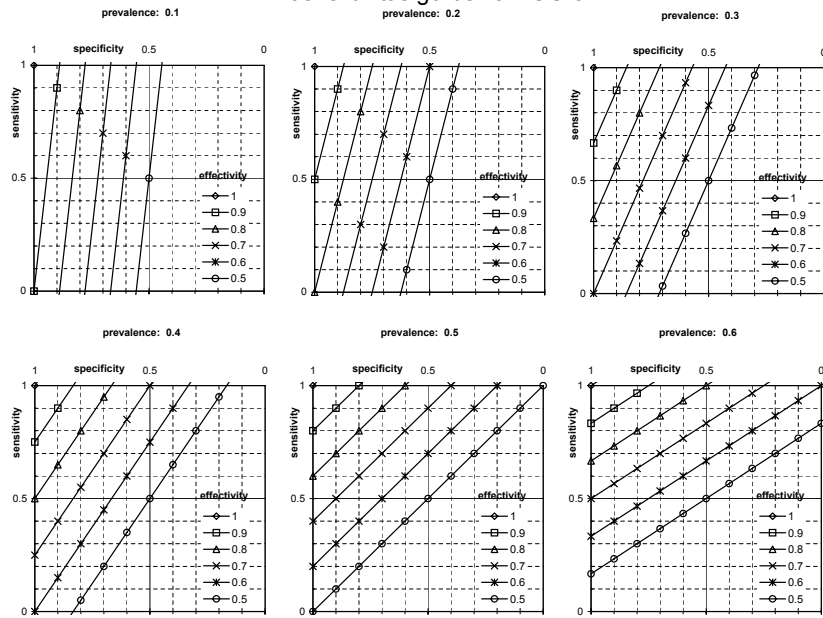


pl. $w = 0.6$, meredekség: 0.66

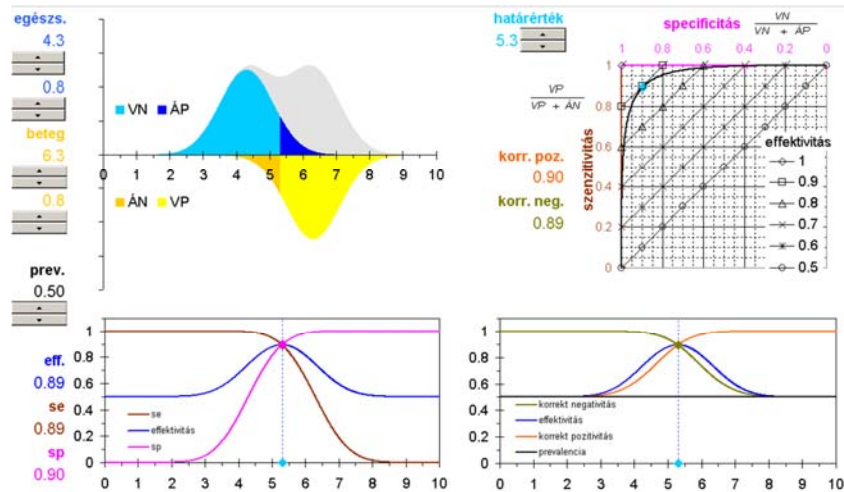


28

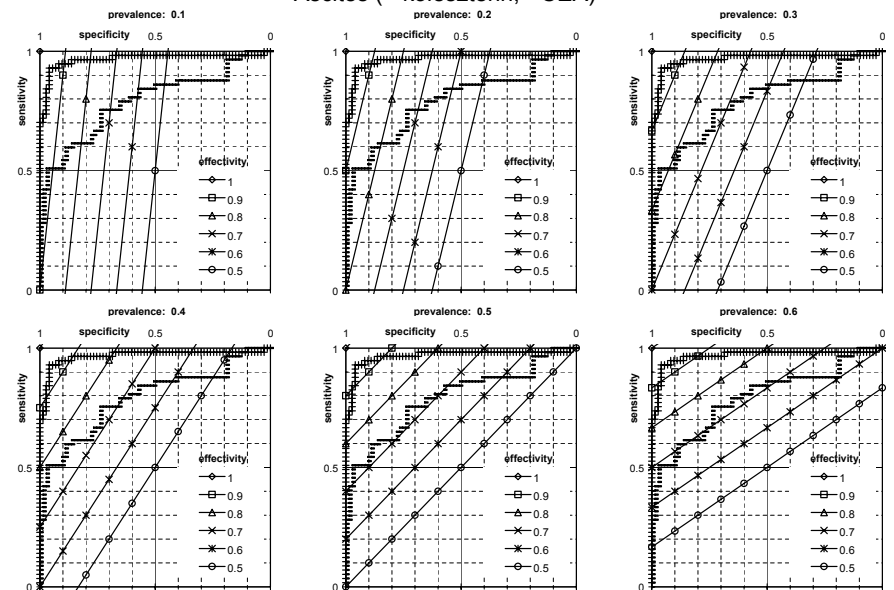
Izoeffektivitás görbék a ROC-on



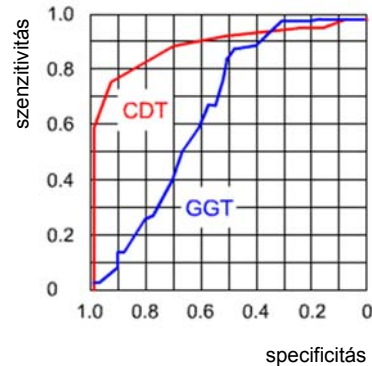
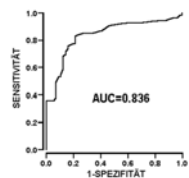
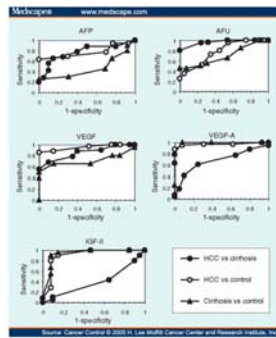
ROC analízis



Ascites (+ koleszterin, - CEA)



További példák



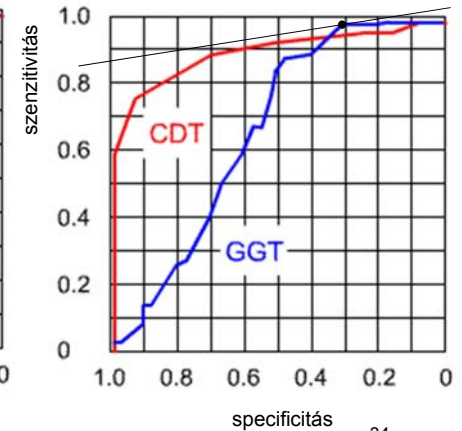
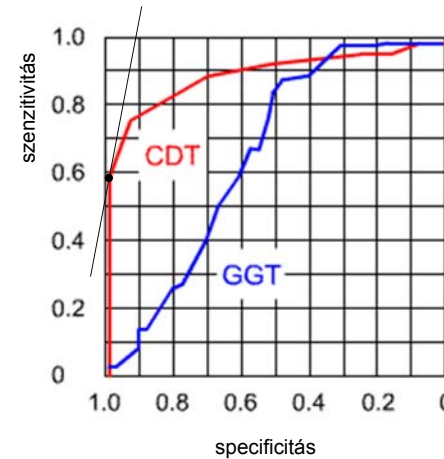
A CDT (carbohydrate deficient transferrin) és GGT (gamma-Glutamyltransferase) módszerek ROC görbéi az alkoholizmus kimutatására. Mivel a CDT görbe gyakorlatilag mindig balra van a GGT-től, ezért a CDT-t jobb módszernek tartják.

33

Példa: maximalizáljuk a diganosztikus effektivitást!

alacsony prevalencia esetén a CDT módszer jobb

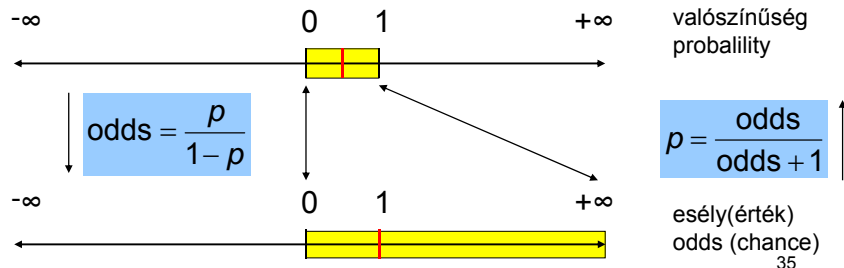
magasabb prevalencia esetén a GGT módszer jobb



34

Események bekövetkezési esélyeinek számszerűsítésére szolgáló mérőszámok

esemény E	valószínűség probability, $p(E)$	esély(érték) odds
lehetetlen	0	0
E bekövetkezése és nem bekövetkezése azonos esélyű	0.5	1
biztos	1	∞



35

Eset-kontroll vizsgálat (case-control study)

Kérdés: Lehet-e egy betegség kialakulásában szerepet játszó tényező egy adott rizikófaktor?

H_0 : nincs összefüggés a rizikófaktor jelenléte és a betegség kialakulása között.

Egy bizonyos betegséggel rendelkező embereket választunk ki (**eset**-ek), majd pedig két csoportba osztjuk őket egy feltételezett **rizikófaktor megléte ill. hiánya** szerint.

Pl. krónikus hörghurutban szenvedőknél feltételezzük rizikófaktoroként a dohányzást.

Választunk melléjük egy olyan csoportot (**kontroll**), amely tagjainál a vizsgált betegség nincs jelen. Őket is két csoportba osztjuk a feltételezett **rizikófaktor megléte ill. hiánya** szerint.

A kontroll csoport egyedeinek számát úgy szokás megválasztani, hogy az nagyjából egyezzen meg az esetek számával. Ez a mintában **mesterségesen** egy 50 % körüli **prevalenciát** hoz létre.

prevalenciát hoz létre.		eset	kontroll
		krónikus hörghurut	nincs betegség
rizikófaktor jelen van	dohányzik	a=40	b=20
rizikófaktor nincs jelen	nem dohányzik	c=60	d=80

Itt nincs értelme kiszámítani a dohányzók között a krónikus hörghurut valószínűségét!

	eset	kontroll
	krónikus hörghurut	nincs betegség
rizikófaktor jelen van	a=40	b=20
rizikófaktor nincs jelen	c=60	d=80

$$\text{esélyhányados} = \frac{\text{a betegség esélye a faktor megléte esetén}}{\text{a betegség esélye a faktor hiánya esetén}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

$$OR = \frac{\frac{40}{20}}{\frac{60}{80}} = \frac{40 \cdot 80}{60 \cdot 20} = 2.7$$

A betegség esélye a rizikófaktor megléte esetén 2.7-szer nagyobb, mint a kontroll csoportban, ahol a rizikófaktor nincs jelen. *(általánosan)*
 A krónikus hörghurut esélye a dohányzók között 2.7-szer nagyobb, mint a nemdohányzók között. *(konkrétan)*

Meghatározható az OR 95 %-os konfidencia intervalluma. Ha az intervallum magában foglalja az 1 értéket, akkor megtartjuk a nullhipotézist (a feltételezett rizikófaktor nem növeli a betegség esélyét).

$$SE(\ln OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

37

Követéses (megfigyeléses, kohort) vizsgálat

Kérdés: Mekkora kockázatot jelent a dohányzás a szívinfarktus kialakulása szempontjából?

H_0 : nincs összefüggés a dohányzás és a szívinfarktus kialakulása között.

Válogatás szempontja: dohányzik – nem dohányzik.

Később (pl. 10 év múlva) megvizsgáljuk az infarktus előfordulási gyakoriságát a két csoportban. Az adatokat egy 2x2-es táblázatban rendezzük el.

	beteg (infarktus)		
rizikófaktor (dohányzik)	igen	nem	összesen
igen	a	b	a+b
nem	c	d	c+d
összesen	a+c	b+d	n=a+b+c+d

38

A kockázat mértéke a dohányosok között: $a/(a+b)$.

A kockázat mértéke a nem dohányzók között: $c/(c+d)$.

Ezekből kiszámítjuk az ún. **relatív kockázatot** (RR – relative risk):

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a \cdot (c+d)}{c \cdot (a+b)}$$

ami megadja, hogy hányszor gyakoribb az adott betegség kockázata adott rizikófaktor megléte esetében. RR várható értéke 1, tehát ha H_0 igaz, akkor $RR = 1$.

Meghatározható az RR 95 %-os konfidencia intervalluma ($RR-1.96 \cdot SE$; $RR+1.96 \cdot SE$). Ha az intervallum magában foglalja az 1 értéket, akkor megtartjuk a nullhipotézist: nincs összefüggés a rizikófaktor (dohányzás) és a betegség (szívinfarktus) kialakulása között.

$$SE(\ln RR) = \sqrt{\frac{1-a/(a+b)}{a} + \frac{1-c/(c+d)}{c}}$$

39

Valószínűségi hányados (likelihood ratio)

Olyan arány, amely megadja, hogy egy teszt módszer milyen mértékben változtatja meg a betegségi esélyeket.

Pozitív teszt eredmény likelihood ratio-ja
(poszttest odds/pretest odds):

$$LR_{\text{pos}} = \frac{\frac{VP}{\overline{AP}}}{\frac{VP}{VN + \overline{AP}}} = \frac{VP}{VN + \overline{AP}} \cdot \frac{1}{\overline{AP}} = \frac{se}{1-se}$$

Negatív teszt eredmény likelihood ratio-ja:

$$LR_{\text{neg}} = \frac{\frac{\overline{AN}}{VN}}{\frac{\overline{AN}}{VN + \overline{AP}}} = \frac{\overline{AN}}{VN + \overline{AP}} \cdot \frac{1}{VN} = \frac{1-se}{sp}$$

40