

## A TESZTEK DIAGNOSZTIKAI ÉRTÉKÉNEK ELEMZÉSE — A KLINIKUS SZEMSZÖGÉBŐL

Olosz E., Koválszki P., Monoki I., Kiss-Olosz Hajna

1977-ben döbrentünk reá, hogy amit mi a tesztek (tünetek) fajlagosságának nevezünk, azt a szerzők túlnyomó többsége ki sem számítja és „specificitás“ (fajlagosság) alatt azt érti, amit mi a negatív teszt érzékenységének nevezünk. Látszólagos tévedésünkre Barbara *McNeyl* egy, a *N. Engl. J. Med.*-ben megjelent összefoglaló referátumának (1) tanulmányozása közben jöttünk reá. A folyóirat szerkesztőségének küldött levelünkre a szerző válaszolt, elismerte, hogy az általunk javasolt elemzési módszer számtanilag helyes, de ő azért marad meg az általa használt módszernél, mert ez megegyezik a WHO terminológiájával. A választ követően, a magunk véleményét még abban az évben közöltük (2).

Ez év tavaszán, a *Lancet* szerkesztőségi cikkben foglalkozik a tesztek érték-elemzésének témájával (3), kételyeit fejezi ki a WHO terminológiájának helyességét illetően és néhány olyan kérdést vet fel, amelyekre véleményünk szerint az általunk, hozzávetőlegesen 10 éve használt módszer helyes választ ad.

Mindenekelőtt idézünk két gondolatot a *Lancet* szerkesztőségi cikkből:

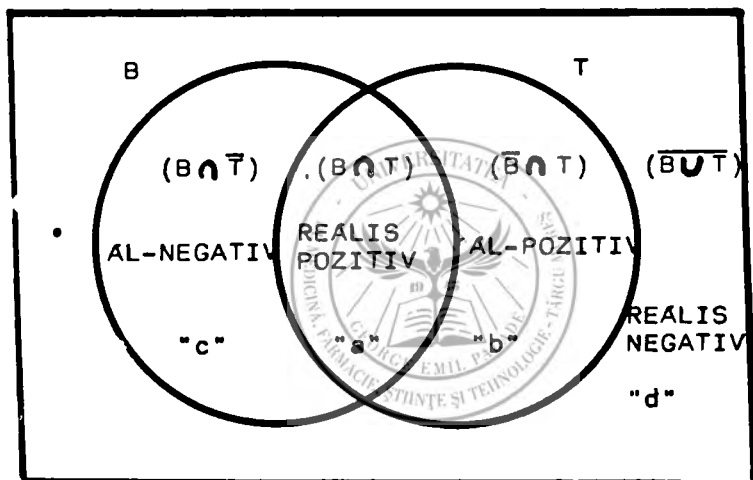
„A terminológiával foglalkozó viták mindig unalmasak, de ebben az esetben szükségesek. A tesztek vagy tünetek diagnosztikai értékét elemző közlemények rendszerint közlik az illető teszt ál-pozitív és ál-negatív hányadosát (false positive and false negative rate), de ezek a fogalmak rendkívül kétértelműek. Például az ál-pozitív hányados alatt azt is lehet érteni, hogy az egészségesek hány százalékában pozitív a teszt, de azt is, hogy az összes pozitív tesztek hány százaléka ál-pozitív.“

„Ha az ál-pozitív hányados 5 %-nyi, akkor a teszt specificitása 95 %-os. A fajlagosság (specificitás) fogalma is kétértelmű. A 95 %-os specificitást a klinikus így értelmezi: 100 olyan esetből, amelyben a teszt pozitív, 95-ben fennáll a betegség, amelynek a kórismzésére használják az illető tesztet. Csakhogy az epidemiológusok (akiknek a javaslatára fogadta el a WHO a jelenlegi terminológiát) egyebet értenek a specificitás alatt, mégpedig azt, hogy a teszt negativitása milyen százalékban osztályozza helyesen az egészségeseket. Minden esetre, ajánlatos, hogy mindazok, akik használják ezeket a kifejezéseket, közöljék, mit értenek alattuk. miként szá-

mitották ki, és hogy a nyert értékek hogyan alkalmazandók a klinikai gyakorlatban.“

Mivel a fajlagosság („specificity“) és az érzékenység („sensitivity“) — bárhogy is értelmezzük jelentésüket — bizonyos feltételes valószínűségeket fejeznek ki, mi elejétől fogva a halmazelmélet és a valószínűség-számítás módszereivel elemeztük a tesztek diagnosztikai értékét. Úgy tudjuk, hogy mi szerkesztettünk elsőként ilyen célból erre alkalmas Venn diagramokat. Az 1. ábrán látható Venn diagramon a négyszög jelképezi a

## A TESZTEK DIAGNOSZTIKAI ERTEKENEK ELEMZESERE ALKALMAS VENN DIAGRAM



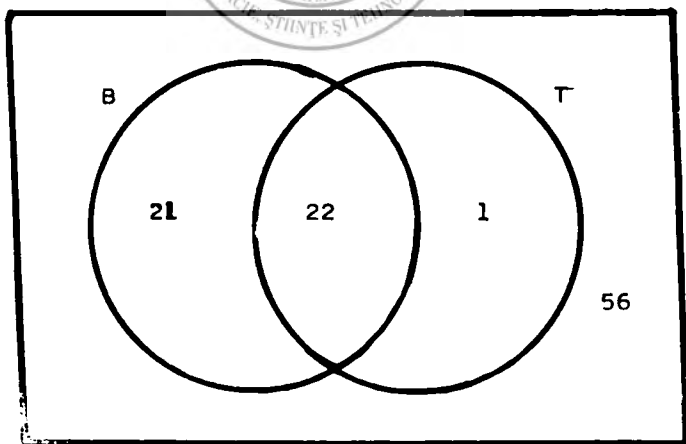
1. ábra

tanulmány tárgyát képező esetek összességét. A négyszögön belül elhelyezkedő B-vel jelölt baloldali kör tartalmazza a „B“ betegségben szenvedő betegek halmazát. A jobboldali, T-vel jelölt kör jelképezi azon esetek halmazát, amelyekben a teszt pozitív volt. Mivel a két kör szeli egymást, ez a keresztmetszet  $(B \cap T)$  azoknak az eseteknek a halmazát jelképezi, amelyekben a teszt reálistan pozitív volt. Ezt a halmazt a „contingentia táblázat“ (ún. „decisional matrix“-ot) használó szerzők „a“ betűvel jelölik. A T körnek a jobboldali, hold alakú része  $(\bar{B} \cap T)$  azoknak az eseteknek a halmazát jelképezi, amelyekben a teszt ál-pozitív volt, vagyis pozitív volt (T), noha a betegség nem volt jelen ( $\bar{B}$ ). A contingentia táblázatokon a szerzők túlnyomó többsége (pl. 4, 5) az ál-pozitív esetek halmazát „b“-vel jelöli, mások viszont (pl. 1) „c“-vel. Többek között ezért is he-

lyesebb a halmazelmélet ideograficus jeleinek használata, mert a  $(\bar{B} \cap T)$  jelzés félreérthetetlen: minden olyan eset, amelyben a teszt pozitív (T), noha a betegség nincs jelen ( $\bar{B}$ ). A B körnek a baloldali hold alakú része  $(B \cap \bar{T})$  jelképezi mindazokat az eseteket, amelyekben a teszt ál-negatív. Ezt a halmazt a contingencia-táblázatokkal elemző szerzők többsége „c”-vel jelöli, kivételt Barbara McNeyl képez, aki itt használja a „b” betűt. Végül a négyszögnek az a területe, amely mind a két körön kívül esik  $(\bar{B} \cap \bar{T})$  a reálian-negatív esetek halmazát ábrázolja, vagyis azoknak az eseteknek felel meg, amelyekben a teszt negativitása egybeesik a betegség hiányával.

Az ál-pozitív és ál-negatív hányadosok kérdésében teljesen egyetértünk a Lancet szerkesztőségi cikkével: két különböző ál-pozitív és ugyan-csak két ál-negatív hányadost kell megkülönböztetnünk. Hogy minél könnyebben érthetővé váljék ez a látszólag elvont kérdés, számítsuk ki ezeket a hányadosokat egy konkrét példán, amit egy előző dolgozatunkból vettünk, és amelyben bizonyos EKG kritériumok diagnosztikai értékét elemeztük a kamra hypertrophiák körismézésében. A baloldali kör foglalja magába azt a 43 esetet, amelyben a klinikai, röntgen, haemodynamikai (szívkatóéterezés) stb. vizsgálatok alapján megállapítottuk, hogy a betegnek jobb-kamra hypertrophiája van. A jobboldali kör jelképezi azokat az eseteket, amelyekben egy bizonyos EKG kritérium constellatio volt megfigyelhető, jelen esetben „intrinsicoid deflexio nagyobb mint 0,03 a  $V_1$ -ben”, amihez legalább még egy jobb-kamra hypertrophiára utaló kritérium társult (amplitudó kritérium, vagy QRS-tengely kritérium, vagy

**B=JOBBKAMRA HYPERTROPHIA; T=INTRINSEC.DEFL.V<sub>1</sub> > 0.03**  
**PLUSZ MEG EGY JKH. +RA UTALÓ EKG KRITERIUM**



2. ábra

incomplete jobb-szár-block, vagy ST-T kritériumok, bármelyik a felsorolt négy közül). Ezt a kritérium kombinációt 23 esetben találtuk: 22 esetben jobb-kamra hypertrophiával együtt és 1 esetben ennek híján (vagyis 1 ál-pozitív teszt). Végül 56 olyan esetünk volt, ahol az EKG tünetcsoport is hiányzott és jobb-kamra hypertrophiája sem volt a betegeknek (reálisan negatív volt a teszt.) Betegeinknek ezt az elosztását („repartitioját“) a 2. ábra szemlélteti.

Ami az *ál-pozitív hányadost* illeti, ennek kiszámítása úgy történik, hogy az egy talált esetet valamihez viszonyítjuk, mégpedig egy olyan halmazhoz, amelynek ez a „halmaztöredék“, amit  $(\overline{B} \cap T)$ -vel jelöltünk, része és ez vagy  $\overline{B}$  (vagyis mindazok, akiknek nincs jobb-kamra hypertrophiájuk), vagy pedig T (vagyis mindazok, akiknek a nevezett EKG tünetcsoportja meg volt). Ezek szerint a két ál-pozitív hányados:

$$a) \quad \frac{(\overline{B} \cap T)}{\overline{B}} = \frac{1}{56+1} = 0,017$$

$$b) \quad \frac{(\overline{B} \cap T)}{T} = \frac{1}{22+1} = 0,043$$

Érthető, hogy egy teszt annál fajlagosabb (specifícusabb), minél kisebb az ál-pozitív hányadosa! Rendben van, de melyik a két ál-pozitív hányados közül. Az is érthető, hogy az ál-pozitív coefficiens a complementer értéke a specificitásnak, vagyis:

$$- \text{ál-pozitív hányados} + \text{specificitás} = 1,00$$

Az epidemiológusok számára voltaképp az a fontos, hogy a „screening“ céljából használt teszt, hány százalékban osztályozza helyesen az egészségeseket és a betegeket. Érthető ezért, hogy az epidemiológusok az a)-pont alatt jelölt ál-pozitív hányadost használják, de ebből az is adódik, hogy amit ők specificitásnak hívnak, azt az alábbi képlet alapján számítják:

$$\text{Epidemiológusok „specificitása“: } \frac{(\overline{B} \cap T)}{\overline{B}} = \frac{56}{56+1} = 0,982, \text{ ami valójában}$$

complementer értéke az a)-típusú ál-pozitív hányadosnak, mert  $0,017 + 0,982$  az gyakorlatilag  $= 1,00$ . Az epidemiológusok által így kiszámított „specificitás“ a következő feltételes valószínűségnek felel meg: P(T| $\overline{B}$ ) vagy más szóval, ha nincs jelen a betegség, akkor milyen valószínűséggel lesz negatív a teszt, vagy úgy is mondhatjuk, hogy a teszt negativitása hány százalékban osztályozza helyesen az egészségeseket (a kapott értéket 100-zal szorozva megkapjuk a %-os értéket) — és íme, az epidemiológus elérte a célját. De a klinikus nem! A klinikusnak egy másik feltételes valószínűsége volna szüksége, mégpedig arra, hogy mi az értéke a P( $\overline{B}$ |T)-nek, vagyis ha pozitív egy teszt, akkor milyen valószínűséggel szenved a páciens B betegségben. Ennek, amit mi a pozitív teszt fajlagosságának nevezünk, a b)-típusú ál-pozitív hányados a complementer értéke. Újabbban egyes epidemiológusok is kiszámítják ezt a valószínűséget, de mivel a negatív teszt érzékenységének megjelölésére a „specificitás“ kifejezést már felhasználták, új kifejezést kellett keresniök és a „predic-

tive value“ (talán „jósló értéknek“ nevezhetnénk magyarul) mellett kö-  
tötték ki. A pozitív teszt specificitását („jósló értékét“) az alábbi képlet  
alapján számíthatjuk ki:

$$\frac{(B|T)}{T} = \frac{22}{22+1} = 0,956, \text{ ami valójában a complemter értéke a b)-}$$

típusú ál-pozitív hányadosnak, mert  $0,043 + 0,956$  gyakorlatilag  $= 1$ .  
Miért pont az ilyen jellegű feltételes valószínűség kell a klinikusnak? Egy-  
szerűen azért, mert mi a feltételt (vagyis, hogy pozitív a teszt) ismerjük,  
és ebből próbálunk arra következtetni, hogy fennáll-e, és milyen valószínűs-  
séggel a B betegség.

Az ál-negatív hányadosnak is két típusát különíthetjük el:

$$\text{a) } \frac{(B|\bar{T})}{\bar{T}} = \frac{21}{21+56} = 0,272$$

$$\text{b) } \frac{(B|\bar{T})}{B} = \frac{21}{21+22} = 0,488$$

Az a)-típusú ál-negatív hányados a pozitív teszt érzékenységének  
(amit az epidemiológusok egyszerűen sensitivitásnak neveznek) comple-  
mter értéke. A pozitív teszt érzékenységének kiszámítása:

$$\frac{(B|T)}{B} = \frac{22}{22+21} = 0,511, \text{ ennek complemter értéke az a)-típusú ál-}$$

negatív hányados, mert  $0,488 + 0,511$  gyakorlatilag  $= 1,00$ . A pozitív teszt  
érzékenységének kiszámításával a következő feltételes valószínűséget kap-  
juk:  $P(T|B)$  vagyis, ha fennáll a B betegség, milyen valószínűséggel lesz  
pozitív a teszt, vagy más módon fogalmazva, a teszt pozitivitása hány szá-  
zalékban osztályozza helyesen a betegeket, a betegeknek hány százalékát  
„szűri ki“.

A b)-típusú ál-negatív hányados a negatív teszt fajlagosságának (amit  
az epidemiológusok a negatív teszt jósló-értékének — predictive value —  
neveznek) complemter értéke. Ennek kiszámítási képlete:

$$\frac{(B|\bar{T})}{\bar{T}} = \frac{56}{56+21} = 0,727. \text{ Itt is megfigyelhető, hogy a b)-típusú ál-ne-}$$

gativ hányados és a negatív teszt összege:  $0,272 + 0,727$  gyakorlatilag  
 $= 1,00$ . A negatív teszt specificitásának kiszámítása a következő feltéte-  
les valószínűséget eredményezi:  $P(\bar{B}|\bar{T})$  vagyis, ha negatív teszt, milyen  
valószínűséggel nincs jelen a betegség. Ez megint csak egy olyan feltéte-  
les valószínűség, ami a klinikust érdekli elsősorban, hiszen itt az ismert-  
nek elfogadott feltétel a teszt negativitása, ebből próbálunk következtetni  
a betegség hiányára.

Látszólag unalmas terminológiai vitába bocsátkozik cikkünk, hiszen  
az epidemiológusok is kiszámítják mindazt amit mi, csak a negatív teszt  
érzékenységét „specificitásnak“ nevezik, és azt, amit mi fajlagosságnak  
(specificitásnak) nevezünk, azt ők jósló-értéknek (predictive value) hívják.  
Egyetértünk a Lancet szerkesztőjei cikkével, hogy bármily unalmasnak  
tűnnék is ez a vita, mégis szükséges, mert a mindennapi gyakorlatban a  
klinikus fajlagosság vagy specificitás alatt azt érti amit mi, nem pedig a  
negatív teszt érzékenységét. Erre a legkirívóbb példát két igen tekintélyes

szerző, D. F. Ranshoff és A. R. Feinstein (4) szolgáltatják: miután elfogadják és használják az epidemiológusok terminológiáját, a közleményük utolsó bekezdésében a következőket írják: „A high sensitivity suggest that the test is good to »rule out« disease; a high specificity suggest a »rule in« usage“. Az utóbbi állítás igaz volna, ha a specificitás alatt azt értenék mint mi, az első állítás semmiképpen nem helyes. A valóságban a betegség kizárására azok a tesztek alkalmasak, amelyeknek a negatív teszt fajlagossága magas (negatív test predictive value), a betegség pozitív kóris-mézésére pedig azok a tesztek a legalkalmasabbak, amelyeknek a pozitív teszt fajlagosságuk (epidemiológusok predictive value-je) közelíti meg az 1,00-et. Feinstein, mint klinikus, a mindennapi beszédben ugyanazt érti specificitásnak, mint mi, és mint általában minden klinikus. Így magyarázható a Freud könyvének megfelelő fejezetébe is beillő „elvtetés“.

A szerkesztőségbe érkezett: 1979. október 15-én.

### Irodalom

1. McNeyl Barbara, Keeler E., Adelstein S. J.: New-Engl. J. Med. (1975), 293, 211; 2. Olosz E., Voloc N., Monoki I.: Rev. Med. (1977), 23, 21; 3. \*\*\* The Lancet (editorial), (1979), vol. 1, nr. 8120, 809; 4. Ranshoff D. F., Feinstein A. R.: New-Engl. J. Med. (1978), 299, 926; 5. Rose G., Barker D. J. P.: Brit. Med. J. (1978), 2, 1070; 6. Nilius R., Neef L.: Das deutsche Gesundheitswesen (1979), 34, 577.

E. Olosz, P. Koválszki, I. Monoki, Hajna Kiss-Olosz

#### THE ANALYSIS OF THE DIAGNOSTICAL VALUE OF SOME TESTS FROM THE CLINICIAN'S STANDPOINT

The diagnostical value of some tests has been calculated by means of the theory of sets, as well as by means of the probability theory illustrated by Venn diagrams. The formulas used and the probabilities which resulted are the following:

$$\text{Sensitivity } T: \frac{(D \cap T)}{D} = \dots\dots\dots P(T|D)$$

$$\text{Sensitivity } \bar{T}: \frac{\overline{D \cap T}}{\bar{D}} = \dots\dots\dots P(\bar{T}|\bar{D})$$

$$\text{Specificity } T: \frac{D \cap \bar{T}}{T} = \dots\dots\dots P(D|\bar{T})$$

$$\text{Specificity } \bar{T}: \frac{\overline{D \cap \bar{T}}}{\bar{T}} = \dots\dots\dots P(\bar{D}|\bar{T})$$

This terminology used by the authors is different from the epidemiological one, which became official being accepted by WHO, but it coincides both with the everyday clinical language and with the semantic-logical meaning of these notions. The epidemiologists name the sensitivity of the negative test (T) as "specificity", and the specificity of the positive and negative test as the "predictive value" of the positive and negative test. The authors consider that if the proposed terminology were accepted, the ambiguities of the WHO nomenclature mentioned in the leading article of Lancet from the 14th of April, 1979 could be eliminated.