

**REICZIGEL JENŐ:
VÁLOGATOTT FEJEZETEK A BIOSTATISZTIKÁBÓL**

**Szent István Egyetem
Állatorvostudományi Kar**

Budapest, 2005.

Tartalomjegyzék

1. Bevezetés	3
2. Az empirikus vizsgálatokkal kapcsolatos alapfogalmak	4
3. Az epidemiológiai vizsgálatok általános kérdései	7
4. Morbiditási és mortalitási ráták	8
5. Epidemiológiai vizsgálati elrendezések	12
6. Diagnosztikai tesztek	17
7. Klinikai kísérletek	27
8. Kísérleti elrendezések	32
9. Túlélési görbék elemzése	37
10. Bioassay	41
Irodalom	44

1. BEVEZETÉS

E jegyzet első változatának alapjául az ELTE survey statisztikus szakirányra szakosodott szociológus hallgatói számára 1998 őszén tartott biostatistika kurzus anyaga szolgált, amelyet a SZIE ÁOTK-n 2000 tavaszán tartott PhD kurzushoz némileg kibővítettem és átdolgoztam. A tanfolyam célja, hogy áttekintést adjon a biostatistikának a leendő kutatók számára legfontosabb – az elemi statisztikai ismereteken túlmenő – fejezeteiről, és ezzel segítséget nyújtson a szakterületükön folyó kutatómunkában való eredményes részvételhez. Az anyag feltételezi, hogy a hallgatók már túl vannak az első bevezető statisztika kurzuson, tehát ismerik legalább a legfontosabb az alapfogalmakat, úgymint valószínűség, várható érték, konfidencia-intervallum, statisztikai próba, stb.

Biostatistikán az orvos- és állatorvos-tudomány, az egészségügy, a biológia, az ökológia és a mezőgazdaság területén alkalmazott statisztikai módszereket értjük. A különböző szakterületek vizsgálati módszerei sajátosan eltérőek. Az egészségügyi és az ökológiai vizsgálatok sokváltozós, nagymintás, legtöbbször nem kísérletes, átfogó megközelítésével szemben a mezőgazdasági kísérletekben pénzügyi, az orvosi-állatorvosi kísérletekben emellett még etikai megfontolások miatt is jellemző, hogy specifikus, célzott kérdésekre keresnek választ, kis mintákat és kevés változót használnak.

Az orvosi-állatorvosi alkalmazások esetén fontos, hogy mindig egyértelműen különbséget tegyünk az orvosi, illetve a statisztikai értelemben vett szignifikancia között. Gyakori eset, hogy nem tudjuk, hogy értjük, amikor azt halljuk vagy olvassuk, hogy "a különbség nem szignifikáns". Az angol nyelvben az „orvosilag szignifikáns” helyett ma már általánosan a „klinikailag releváns” kifejezést használják, ami követendő példa, hiszen így egyértelmű lesz a fogalmazás, azaz hogy „a különbség nem szignifikáns” vagy „a különbség irreleváns”.

A jegyzet által felölelt anyag természetesen csak válogatásnak tekinthető a biostatistika területéről. Nevezetesen kitérünk az epidemiológiai vizsgálatok különféle módszereire és vizsgálati elrendezéseire, a diagnosztikai tesztek értékelésének néhány kérdésére, a klinikai kísérletek tervezésére és kiértékelésére, a leggyakrabban használt kísérleti elrendezésekre, a túlélési görbék elemzésére, valamint a bioassay típusú vizsgálatokra.

2. EMPIRIKUS VIZSGÁLATOKKAL KAPCSOLATOS ALAPFOGALMAK

A statisztikát az orvosi alkalmazásokban a betegségek kóroktanával (etiológia), diagnosztizálásával és terápiájával kapcsolatos elemzésekre használják.

A vizsgálatok leíró vagy analitikus jellegűek lehetnek, mi a továbbiakban csak az analitikus vizsgálatokkal foglalkozunk. Ezek célja általában oksági, vagy legalábbis – ha a vizsgálati elrendezés oksági kapcsolat kimutatását nem teszi lehetővé – valószínűsíthető oksági összefüggések kimutatása.

A kísérlet (*experiment, trial*) olyan vizsgálat, amelyben a kísérletező aktív beavatkozást („kezelés”) végez a kísérleti alanyokon, és ennek hatását vizsgálja.

Az “observational study” (nem kísérletes vizsgálat, felmérés, „passzív” megfigyelés) kifejezést a tervezett klinikai kísérlet (*planned clinical trial*) ellentétéként használják, amikor is a “kezeléseket a természet allokalja”, más szóval az egyedek kezelés szerinti csoportokba sorolását nem a kutatók, hanem az élet végzi. (Így a kezelések random allokalásáról, amelyre olyan sok gondot fordítanak a kísérletekben, itt szó sem lehet.) A “kezelések” különféle jótékony vagy káros (többnyire az utóbbi) tényezők, életmód, stb. Ezek általában diszkrét változók, a célváltozó lehet folytonos vagy diszkrét. Tipikus, hogy két csoporttal dolgozunk, egy a vizsgált hatásnak kitett (*exposed*), és egy kontroll (*non-exposed, control*) csoporttal.

Elfogadott nézet, hogy oksági kapcsolat kimutatására csak a kontrollált randomizált kísérletek (*controlled randomized trial, CRT*) alkalmasak, amelyekben a kezelt és a kontroll csoport között az alkalmazott kezelést kivéve semmilyen különbség nincs, tehát a célváltozóban megmutatkozó különbség csakis ennek tulajdonítható (vagy esetleg a véletlen játékának).

A vizsgálatokat gyakran kategorizálják aszerint is, hogy retrospektívek vagy prospektívek. Prospektívnek nevezünk egy vizsgálatot, ha a mérések / megfigyelések a vizsgálat során történnek, az adatok a vizsgálat során keletkeznek. Retrospektív vizsgálat esetén az elemzés a vizsgálat időpontjához képest korábbi eseményekkel, illetve korábban keletkezett adatokkal történik. (Jegyezzük meg, hogy míg a megfigyeléses vizsgálatok között mindkét típus előfordul, a kísérletek mindig prospektív típusúak.)

A következőkben a vizsgálatok tervezésével kapcsolatos alapfogalmakat foglaljuk össze röviden:

Tényező, faktor (*factor*) – „független változó”, „ható tényező” (kezelés, gyógyszer, gyógyszeradag, stb.) aminek hatását vizsgálni kívánjuk.

Megfigyelési egység (*experimental unit*) az az egység a kísérletben, amelyre egy bizonyos kezelést vagy kezeléskombinációt alkalmazunk (egy állat, egy alom, egy sejt kultúra, egy petri-csésze táptalaj, stb.).

Mastitis kísérletben az állatot vagy a beteg tőgyegyedet tekintjük megfigyelési egységnek? Milyen szempontokat vegyünk figyelembe a választáshoz?

Két szemcsepp összehasonlításakor a beteget, vagy a szemet tekintünk megfigyelési egységnek? Mi az előnye az egyik vagy másik választásnak?

Takarmányozási kísérletben lehet-e az egyes állatokat tekinteni megfigyelési egységnek, ha a takarmányfogyasztást / súlygyarapodást technológiai okok miatt (pl. baromfinál) csak ketrecenként tudjuk mérni? Mi lehet a hátránya annak a megoldásnak, ha a kísérleti állatokat elkülönítve, külön ketrecben tartunk?

Családterápia vizsgálata esetén a családot vagy az egyes családtagokat tekintjük megfigyelési egységnek? Lehet olyan vizsgálatot tervezni, amelyikben mindkettő egyszerre vizsgálható?

Célváltozó vagy kimeneti változó (*outcome variable, response variable*) olyan megfigyelhető vagy mérhető változó, amely alkalmas a tényezők hatásának mérésére (a megfigyelési egység állapotának, kezelésre adott válaszána jellemezésére). Lehet bináris (0/1, pl. túlélés, megbetegedés, gyógyulás, stb.) vagy számszerű (baktériumtelepek száma, átmérője, vérnyomás, közérzet-pontszám, stb.). Egy kísérletben több célváltozó is lehet, ekkor ki kell jelölni egy elsődlegest.

Milyen szempontok szerint válasszunk kimeneti változót?

- legyen jól definiált (a „vérkép értékelése alapján” nem elég pontos, a reprodukálhatóság miatt „anyag és módszer” szerű leírás kell),
- legyen a rendelkezésre álló eszközökkel kellő pontossággal mérhető,
- minél kevesebb szubjektív torzítást tartalmazzon (vakág, szimultán kiértékelés több kiértékelő személy bevonásával),
- fontos szempont, hogy a vizsgálat célja hatás (*effect*) vagy terápiás hatás (*efficacy*) kimutatása (az elsőből nem következik a második - pl. egy rizikófactor kiiktatása nem biztos, hogy valóban csökkenti a betegség kockázatát!).

Ismétlés (*replication*) – kísérleteknél ugyanazt a kezelést (kezeléskombinációt) általában több megfigyelési egységen szokták alkalmazni. Ennek az az értelme, hogy

- csökkentsék annak az esélyét, hogy a kísérlet egyes megfigyelési egységek kiesése (elhullás, sikertelen mérés, stb.) miatt értékelhetetlenné válik,
- kiderüljön, mekkora a kezelésre adott válasz szóródása populációban,
- az eredmények a populációra általánosíthatók legyenek,
- növekedjék a vizsgálat pontossága.

Mi a véleményünk arról, ha valaki – az előbb említett baromfi takarmányozási kísérletben, amikor csak ketrecenként tud mérni – a pontosságot úgy kívánja növelni, hogy megkétszerezi az egy ketrecben tartott állatok számát?

Párhuzamos vizsgálati elrendezésről (*parallel design*) beszélünk, ha a vizsgálatban két csoport (kezelt-kontroll), esetleg több csoport összehasonlításából kívánunk következtetéseket levonni a kezelés(ek) hatására nézve. A csoportok a populációból vett véletlen minták, amelyek lehetnek függetlenek (*independent samples*) vagy párosítottak, amikor a két minta összetartozó párokból áll, például férj és feleség testsúlya, ugyanazon anya első és második gyermekének születési súlya, stb. (*matched/paired/dependent samples*), de minden megfigyelési egység csak egy csoportba van besorolva (vö. ismételt mérések és crossover elrendezés).

Kontroll (*control*) a nem kezelt csoport, amihez a kezeltet hasonlítjuk, vagy a standard kezelést kapott csoport, amihez az új, kiértékelendő kezelést kapottakat hasonlítjuk.

A „történelmi kontroll” (*historical control*) – korábbi betegek adatainak kontrollként való felhasználása – sok hibalehetőséget rejt magában (egészségi állapot, higiénia, életmód, kiegészítő kezelések, kezelési eljárások időbeli változása).

Az önkontrollos (kezelés előtt-után) elrendezés nem helyettesítheti a kontrollt, mert a placebohatást nem képes kiszűrni.

Célpopuláció és vizsgált populáció (*target population, sampled population*). A célpopuláció az a populáció, amelyre a vizsgálat eredményeit alkalmazni kívánjuk, a vizsgált populáció pedig az a populáció, amelyből az elemzés alapjául szolgáló véletlen mintát vettük. Ha e két populáció nem azonos, akkor a kapott eredmény érvényessége megkérdőjelezhető.

Belső és külső validitás (*internal, external validity*). A vizsgálat belső validitásán azt értjük, hogy a kapott eredmény a valóban érvényes a vizsgált mintára (az elemzés nem tartalmaz torzítást), a külső validitás pedig azt jelenti, hogy érvényes a kívánt célpopulációra (azaz általánosítható). Ehhez a belső validitáson túlmenően még arra is szükség van, hogy a mintavétel se okozzon torzítást.

A vizsgálat belső validitását veszélyezteti például, ha a kiértékeléskor egy olyan statisztikai módszert, amelynek feltétele, hogy a vizsgált változó normális eloszlású legyen, nem normális eloszlású változó elemzésére alkalmazunk.

A vizsgálat külső validitása válik kérdésessé, ha az eredményeket a háziorvosi praxisra vonatkozóan kívánjuk felhasználni, a vizsgálatot viszont egy egyetemi klinika beteganyagán végezzük, ahol jellemzően súlyosabb vagy komplikált esetek fordulnak elő.

Zavarás, elfedés (*confounding*). Előfordulhat, hogy két változó között fennálló összefüggést egy harmadik, zavaró változó (*confounder variable*) hatása kiolt (elfed). Hasonlóképpen előfordulhat az is, hogy két független változó között egy harmadik változó hatása látszólagos összefüggést teremt.

Példák:

1. Egy angol kutatásban például összefüggést találtak a felekezeti hovatartozás és a trombózisos esetek gyakorisága között. Alaposabb vizsgálódás fényt derített rá, hogy a látszólagos összefüggés a fogamzásgátlással kapcsolatos attitűdök közötti különbséggel, valamint az orális fogamzásgátlók trombozist előidéző hatásával magyarázható.

2. Egy irodában január folyamán mérik a belső és a külső hőmérsékletet és közöttük negatív korrelációt kapnak (minél melegebb volt kinn, annál hidegebb volt benn). Természetesen ennek oka az, hogy a fűtő a kazán teljesítményét mindig a külső hőmérséklethez igazítja (kaphattak volna akár pozitív, akár 0 korrelációt is attól függően, hogy a fűtő milyen ügyes, vagy – automata vezérlés esetén – a termosztát hol van elhelyezve és mennyire jól szabályoz).

3. Egy egyetem felvételi statisztikája alapján

	Férfiak	Nők	Összesen
Felvettek	3000	2000	5000
Elutasítottak	7000	8000	15000
Összesen	10000	10000	20000

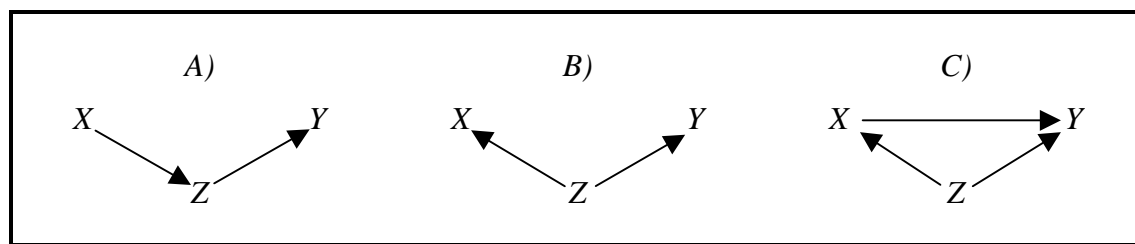
felmerült a nemi diszkrimináció gyanúja, mivel a jelentkező férfiaknak 30, a nőknek pedig csak 20%-át vették fel, és statisztikailag ez a különbség szignifikáns. (Vajon melyik statisztikai próbát alkalmazták?) Amikor megvizsgálták, hogy melyik kar milyen mértékben felelős ezért, meglepő eredményt kaptak:

	Természettud. Kar			Gazdaságtud. Kar			Orvostud. Kar		
	Férfiak	Nők	Össz.	Férfiak	Nők	Össz.	Férfiak	Nők	Össz.
Felvettek	1500	500	2000	1300	700	2000	200	800	1000
Elutasítottak	1500	500	2000	2700	1300	4000	2800	6200	9000
Összesen	3000	1000	4000	4000	2000	6000	3000	7000	10000

Hogyan értelmezhetjük ezt a furcsa eredményt?

A confounding legjellemzőbb típusait láthatjuk az alábbi ábrán. Mindig *Z* jelöli a confounder változót. Az *A*) és *B*) esetben *Z* közvetett, illetve látszólagos összefüggést kelt *X* és *Y* között, a *C*) esetben pedig megváltoztatja az *X* és *Y* közötti valódi összefüggést (erősítheti, gyengítheti, akár teljesen elfedheti vagy ellentétes előjelűre is változtathatja). Az előbbi trombózisos példa az *A*), a másik kettő pedig a *C*) típusba sorolható. A harmadik példában a „nem” és a „felvétel” közötti összefüggés az összesített táblában a

karonkéntival ellentétes irányúvá válik. Ez a jelenség a Simpson-paradoxon: egy rétegenkénti pozitív összefüggés az összesítésben negatívnak mutatkozik (vagy fordítva).



Ha két változó között nehezen értelmezhető, vagy az elméletileg várttól eltérő összefüggést kapunk, mindig jusson eszünkbe a confounding lehetősége. Keressünk potenciális confoundereket (a szakirodalomból meríthetünk ötleteket), és próbáljuk meg kiszűrni a hatásukat. Erre egy egyszerű módszer az, hogy a potenciális confounder szerint homogén csoportokat, rétegeket képezünk és az X és Y közötti összefüggést rétegenként vizsgáljuk (lásd majd az 5. fejezetben). Mivel egy rétegen belül Z állandó (vagy közel állandó), hatása itt nem mutatkozhat. Tehát ha a rétegenkénti elemzés eltérő eredményt ad az egybenitől, akkor Z valóban confounder.

Jegyezzük meg, hogy egy változó biztosan nem lehet confounder, azaz nem zavarhatja az X és Y változók közötti kapcsolat kimutatását, ha X és Y közül legalább az egyikől független (pontosítás alább). Ezt célozza mind a véletlen mintaválasztás, mind pedig a kezelések random allokálása.

Ha egy potenciális zavaró Z változóra az alábbi két állítás közül:

1. Y és Z feltételesen függetlenek X -re nézve,
2. X és Z függetlenek,

akár csak az egyik is igaz, akkor Z nem zavarhatja az X és Y közötti kapcsolat kimutatását, legalábbis olyan értelemben, hogy ekkor vagy $E(Y|X,Z) = E(Y|X)$, vagy $E(Y|X,Z)$ az $E(Y|X)$ torzítatlan becslése.

Bizonyítás:

a) 1-ből egyenesen következik $E(Y|X,Z) = E(Y|X)$,

b) ha 2 fennáll, akkor $E(Y|X) = \int yf(y|x)dy = \int y \left(\int f(y|x,z)f(z|x)dz \right) dy = \int \left(\int yf(y|x,z)dy \right) f(z)dz = E(E(Y|X,Z))$, azaz $E(Y|X,Z)$ az $E(Y|X)$ torzítatlan becslése.

Ha a kezelt és a kontroll csoport is egyszerű véletlen minta, akkor a hatás becslése torzítatlan (legalábbis nincs mintavételi torzítás). Akkor is ez a helyzet, ha egyesítésük egyszerű véletlen minta és a kezelés allokálása randomizált (ekkor ugyanis a kezelés típusa független bármely potenciális zavaró változótól).

3. AZ EPIDEMIOLOGIAI VIZSGÁLATOK ÁLTALÁNOS KÉRDÉSEI

Az epidemiológia jelentése járványtan, és kialakulása idején (a múlt század végén) valóban a járványok keletkezését és terjedését kutatta. Mára azonban már tágabb területet ölel fel, különféle betegségek előfordulásának jellemző mintázatait vizsgálja (morbiditás, mortalitás), sőt nemcsak betegségeket, hanem egyéb biológiai tulajdonságokét (pl. balkezesség) és életvitelbeli szokásokét (pl. dohányzás vagy vegetarizmus) is. Az epidemiológiai vizsgálatok jellemzően nem kísérletező, hanem „observational” típusúak (bár vannak kivételek).

Ha a célváltozó bináris, akkor általában a hatásnak kitett és ki nem tett csoportok rizikóját vetik össze, vagy additíve, vagy multiplikatíve (azaz vagy azt nézik, mennyivel, vagy azt hogy hányszorosára nő a rizikó a nem kitett csoporthoz képest). A relatív rizikó (RR , *relative risk*, lásd később) az egyén szempontjából méri a kockázat növekedését, illetve

csökkenését egy bizonyos rizikófaktorra vonatkozóan. A tényezőnek tulajdonítható rizikó (*AR, attributable risk*, lásd később) azt méri, hogy populáció szinten hány emberéletet követel egy bizonyos kockázati tényező (pl. a dohányzás), vagy hány élet menthető meg egy bizonyos intézkedés bevezetésével (pl. a biztonsági öv használatának kötelezővé tételével).

Folytonos célváltozó esetén is használhatunk multiplikatív vagy additív mértéket a tényező hatásának mérésére. Ha \bar{Y}_e a célváltozó átlaga a kitett csoportban, \bar{Y}_n pedig a kontrollban, akkor \bar{Y}_e / \bar{Y}_n multiplikatív, $\bar{Y}_e - \bar{Y}_n$ pedig additív mérték. Hogy adott esetben melyiket érdemes használni, arra nézve általános szabályt nehéz mondani, elsősorban a mértékek értelmezhetőségét kell figyelembe venni az adott körülmények között.

Voltaképpen mindig az egyedi hatásokra vagyunk kíváncsiak, ezért ha a hatás nem, kor, vagy valamely más szempont szerint változik, akkor az elemzést is csoportonként kell elvégezni (adott esetben ez függhet a mérték megválasztásától, lehet például, hogy $\bar{Y}_e - \bar{Y}_n$ csoportonként változik, de \bar{Y}_e / \bar{Y}_n nem).

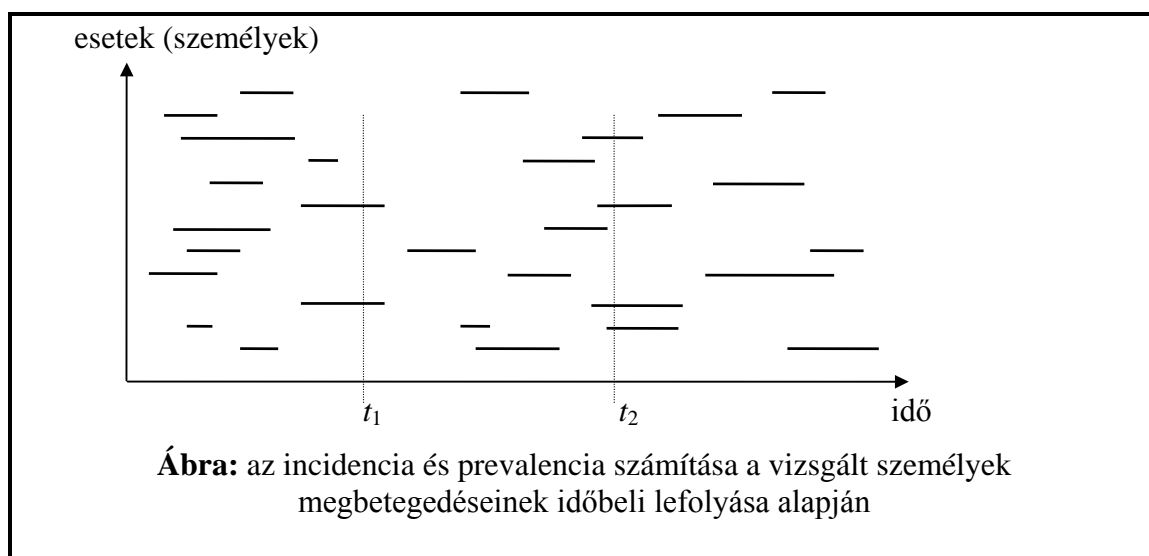
Keresztmetszeti (*cross-sectional*) vizsgálatnak nevezzük az olyan vizsgálatokat, amelyekben a potenciális ható tényezők és a célváltozó mérése egyazon időpontban történik. A tényezők és a célváltozó közötti asszociáció szerencsés esetben ilyen vizsgálatból is kimutatható, de mivel a hatás a célváltozóban általában bizonyos késleltetéssel jelenik meg, jobb eredményt várhatunk olyan elrendezésű vizsgálatról, amelyben a célváltozó mérése későbbi időpontban történik, mint a ható tényezőké („*cohort*” és „*case-control*” vizsgálatok, lásd az 5. fejezetben).

4. MORBIDITÁSI ÉS MORTALITÁSI RÁTÁK

A mortalitási és a morbiditási ráták a betegségek, illetve halálokok gyakoriságának mérésére, a gyakoriságok változásának számszerűsítésére, összehasonlítására szolgálnak. Ne felejtsük el, hogy a ráták szigorúan véve mindig adott időszakra és populációméretre vonatkoznak, a mértékegységük megadásánál is gondoljunk erre, például a “5 eset /év/1000 fő” egyértelmű, de az “5 eset /év” nem.

Egy betegség incidenciája csak az adott időszakban észlelt új eseteket, prevalenciája pedig az összes (új és nem új) eseteket magában foglalja. Az incidencia mindig időszakra vonatkozik, a prevalencia vonatkozhat időpontra (*point prevalence*), vagy időszakra is (*period prevalence*), bár az utóbbi értelmezést ritkábban is használják (és kicsit problematikusabb is az ismétlődések miatt).

Az ábra alapján a $t_1 - t_2$ időszakra vonatkozó incidencia számításához 11 esetet kell figyelembe venni, a t_1 időpontbeli prevalencia számításakor 2 esetet, a $t_1 - t_2$ időszakra vonatkozó prevalencia számításához pedig 13-at.



Ha mortalitási rátáról van szó, hogyan értelmezhető az incidencia és a prevalencia?

Mondjon véleményt a következő – egy betegség incidenciájával és prevalenciájával kapcsolatos – állításokról (az incidenciát és a prevalenciát ugyanarra a populációra és időszakra vonatkoztatva):

- az incidencia mindig nagyobb, mint a prevalencia
- az incidencia mindig kisebb, mint a prevalencia
- bármelyik lehet nagyobb is, kisebb is, egyenlők is lehetnek.

Véleményét példákkal is támassza alá!

Egy morbiditási vagy mortalitási ráta vonatkozhat a teljes populációra (*overall rate, crude rate*) vagy csak egy csoportra (*specific rate*), így beszélhetünk például nem- vagy korszpecifikus morbiditási rátákról. Ha a ráta korcsoportonként más és más, akkor a teljes populációra vonatkozó ráta a populáció korstruktúrájától is függ (gondoljunk a teljes valószínűség tételére!). Két populáció (például két ország) összehasonlításakor ez a hatás (a korstruktúrák különbözősége esetén) zavaró lehet, ezért a rátákat az összehasonlításhoz standardizálni kell. Az alábbi példában a korszpecifikus prevalenciák az I. populációban rendre alacsonyabbak, a korstruktúra különbségei miatt mégis a teljes populációra számított prevalencia a II. populációban alacsonyabb.

Példa: reumás megbetegedések előfordulása két populációban

	I. populáció		II. populáció	
	prevalencia	korstruktúra	prevalencia	korstruktúra
fiatalok (<30)	5%	20%	5%	40%
középkorúak (30-60)	15%	40%	20%	40%
idősek (>60)	35%	40%	40%	20%
teljes populáció	21%*		18%	

*számítási mód: $0.05 \cdot 0.20 + 0.15 \cdot 0.40 + 0.35 \cdot 0.40 = 0.21$

A standardizálásra kétféle eljárást szoktak használni. Az úgynevezett direkt standardizáláshoz egy standard populációra van szükség, amely jobb híján lehet a két összehasonlítandó populáció egyesítése. Az összehasonlításhoz mindkét populáció specifikus rátáival kiszámítjuk a standard populációra vonatkozó rátát, majd e két standard ráta hányadosát használjuk a populációk közötti különbség kifejezésére. Ezt a mutatót CMR-nek (*comparative mortality/morbidity ratio*) vagy CMF-nek (*comparative mortality/morbidity factor*) nevezik. (Figyelem, ez nem ráta, hanem egy dimenzió nélküli arányszám!).

Példa: összehasonlítás direkt standardizálás segítségével az előző példa (reumás megbetegedések) adataival

	standard populáció korstruktúrája	I. populáció korszpecifikus rátái	II. populáció korszpecifikus rátái
fiatalok (<30)	30%	5%	5%
középkorúak (30-60)	40%	15%	20%
idősek (>60)	30%	35%	40%
standard populáció rátája		18%	21.5%

$$\text{CMF} = 18 / 21.5 = 83.7\%$$

Az úgynevezett indirekt standardizáció standard rátákkal történik. Ezt alkalmazhatjuk például, ha arra vagyunk kíváncsiak, hogy hogyan változna a morbiditási ráta *A* országban, ha az egészségügyi ellátás azonos szinten volna a *B* országgal (és így a korszpecifikus ráták azonosak volnának, persze a korstruktúra nem). A tényleges és az ilyen módon standardizált ráták hányadosát SMR-nek (*standardized mortality/morbidity ratio*) vagy néha SIR-nek (*standardized incidence ratio*) nevezik. (Figyelem, ez sem ráta, hanem egy dimenzió nélküli arányszám!)

Példa: indirekt standardizálás az előző példa (reumás megbetegedések) adataival

	II. populáció korstruktúrája	II. populáció korszpecifikus rátái	standard korszpecifikus ráták
fiatalok (<30)	40%	5%	5%
középkorúak (30-60)	40%	20%	15%
idősek (>60)	20%	40%	35%
teljes populáció rátája		18% (tényleges)	15% (standard)

$$\text{SMR} = 18 / 15 = 120\%$$

További fontos mérőszámok. Egy betegség okainak felderítését célzó (=etiológiai) vizsgálatban a potenciális okok, úgynevezett rizikó-faktorok elemzése történik különböző módszerekkel. A fő kérdés mindig az, hogy a vizsgált rizikó-faktor megléte valóban befolyásolja-e a betegség előfordulásának esélyeit. Ennek vizsgálata statisztikailag a tapasztalt előfordulási valószínűségek összehasonlítása a faktor hatásának kitett (*exposed*) és nem kitett (*non-exposed*) populációban.

Az összehasonlításra sokféle mutatót használnak, a leggyakrabban alkalmazott mérőszámokat az alábbi táblázatban foglaljuk össze. A táblázatban alkalmazott jelölések a következők:

B+: a betegség bekövetkezése (morbiditás / mortalitás / incidencia), *B-*: a betegség hiánya,

F+: a potenciális rizikófaktor megléte (expozíció), *F-*: a potenciális rizikófaktor hiánya.

A táblázatban felsorolt mérőszámok mindegyike a kockázatot a betegség előfordulási valószínűségeként értelmezi, amely lehet feltételes (egy részpopulációban) vagy feltétel nélküli (az egész populációban). Megjegyezzük, hogy léteznek más elven alapuló, bonyolultabb mérőszámok is, amelyek több szempontot, nemcsak az előfordulási valószínűséget veszik figyelembe.

Táblázat: az epidemiológiai vizsgálatokban leggyakrabban használt mérőszámok

Mérőszám	Angol elnevezése	Értelmezése
kockázat a rizikófaktor hiányában	risk of those nonexposed	$R_{F-} = P(B+ F-)$
kockázat a rizikófaktor fennállása esetén	risk of those exposed	$R_{F+} = P(B+ F+)$
átlagos kockázat a populációban	population risk	$PR = P(B+)$
relatív rizikó	relative risk vagy risk ratio	$RR = R_{F+} / R_{F-}$
rizikó-különbség	risk difference*	$RD = R_{F+} - R_{F-}$
a rizikófaktornak tulajdonítható kockázat	attributable risk vagy ethiologic fraction	$AR = (R_{F+} - R_{F-}) / R_{F+}$
a populációbeli átlagos kockázat rizikófaktorának tulajdonítható része	population attributable risk	$PAR = (PR - R_{F-}) / PR$

*sokhelyütt ezt a mutatót az attributable risk egy változataként említik

5. EPIDEMIOLOGIAI VIZSGÁLATI ELRENDEZÉSEK

Az alábbiakban egy betegség okainak felderítését célzó vizsgálati elrendezésekkel foglalkozunk. A két leggyakrabban használt vizsgálati elrendezés a kohorsz és a “case-control” vizsgálat, amelyek elsődlegesen a mintaválasztás módjában különböznek, ebből adódóan azonban tulajdonságaik jelentősen eltérnek.

Kohorsz vizsgálat. A kohorsz, követéses vagy longitudinális (*cohort, follow-up, longitudinal*) típusú vizsgálat egészséges személyekkel indul, akiknek egy része ki van téve egy tényező (a potenciális betegségkókozó, például valamely munkahelyi vagy környezeti ártalom) hatásának, a többiek pedig nem. A vizsgált személyeket követjük és feljegyezzük a bekövetkező megbetegedéseket, majd ennek alapján összehasonlítjuk a tényezőnek kitett és a kontroll csoportot.

A kohorsz vizsgálatok nyújtják a legtöbb információt a tényezőnek a betegséggel kapcsolatos hatásáról és lehetővé teszik a kockázat direkt mérését, azonban igen költségesek, idő- és munkaigényesek. Több szerző (különösen régebben) azonosította a kohorsz vizsgálatokat a prospektív vizsgálatokkal. Ez félrevezető, ugyanis egy kohorsz vizsgálat is lehet retrospektív. Ekkor a kohorsz levéltári vagy irattári adatok visszakeresésével állítható elő (például egy üzem dolgozói, vagy egy település lakói egy korábbi időpontban). Így a vizsgálat időigényessége jelentősen csökkenthető, de a munkaigényesség már nem annyira (a követés miatt).

“Case-control” vizsgálat. A “case-control” vizsgálat a betegségben szenvedők (“esetek”, “cases”) és a betegségben nem szenvedők (“kontrollok”, “controls”) egy-egy csoportját választja kiindulásul, és e személyek előéletében igyekszik azonosítani a betegség kialakulását befolyásoló tényezőket. Ez történhet interjúzással vagy kérdőívvel, de akárhogy is, a “case-control” vizsgálat mindenképpen retrospektív. Egyszerűsége és olcsósága miatt sokkal gyakrabban alkalmazzák, mint a kohorsz típusút.

A kétféle típusú vizsgálat közötti néhány fontos különbséget foglaltunk össze az alábbi táblázatban.

Kohorsz	Case-control
A tényezőre koncentrálnak, egy tényezőnek többféle hatása is vizsgálható	A betegségre koncentrálnak, egy betegségnek többféle oka is vizsgálható
Alkalmazása akkor előnyös, ha a vizsgált tényező ritka	Alkalmazása akkor előnyös, ha a vizsgált betegség ritka
Lehetővé teszi az incidenciát és a relatív rizikó direkt mérését	

Becslések kohorsz vizsgálat esetén. Most rátérünk arra, hogy hogyan becsülhetők az előbbieken felsorolt mutatószámok a mintából kohorsz vizsgálat esetén. Foglaljuk össze a vizsgálat eredményeit egy 2x2-es táblázatban.

Betegség	+	-
----------	---	---

Faktor		
+	a	b
-	c	d

Kohorsz vizsgálat esetén az $RR = R_{F+} / R_{F-}$ kifejezés számlálója is és nevezője is külön-külön becsülhető, mégpedig $R_{F+} = a / (a+b)$ és $R_{F-} = c / (c+d)$. A relatív rizikó becslése innen

$$RR_B = a / (a+b) : c / (c+d).$$

A becslés logaritmus, $\ln(RR_B)$ közelítőleg normális eloszlású, $E(\ln(RR_B)) = \ln(RR)$ és $\text{var}(\ln(RR_B)) = 1/a + 1/c$. Ennek alapján konfidencia intervallumot is adhatunk $\ln(RR)$ -re, nevezetesen

$$\ln(RR_B) \pm u_\alpha \sqrt{1/a + 1/c}$$

alakban, ahol u_α a normális eloszlás kritikus értéke (95%-os megbízhatósági szinthez $u_\alpha = 1.96$, 90%-oshoz 1.65).

Példa: azbeszttel való munkavégzés rákot okozó hatását elemezték egy üzem dolgozóin kohorsz vizsgálatban (136 exposed, 150 kontroll) 10 éves követéssel. Az eredményeket az alábbi táblázat foglalja össze (fiktív adatok):

Rákos megb.	+	-
Azbeszt		
+	47	89
-	18	132

Az egyes rizikók és a relatív rizikó becslése:

$$R_{F+} = 47 / 136 = 0.346, R_{F-} = 18/150 = 0.120, RR_B = 0.346 / 0.12 = \mathbf{2.87}.$$

A relatív rizikó logaritmus $\ln(RR_B) = \ln(2.87) = 1.054$. A 95%-os konfidencia-intervallum a relatív rizikó logaritmusára a fenti képlet szerint:

$$\left(1.054 - 1.96 \cdot \sqrt{1/47 + 1/18} \dots 1.054 + 1.96 \cdot \sqrt{1/47 + 1/18} \right).$$

A számításokat elvégezve ez (0.514 ... 1.601), amiből a relatív rizikóra a **(1.673 ... 4.958)** konfidencia-intervallumot kapjuk.

Számolás a “person-years” elv alapján: az éves incidenciaráta számolásakor figyelembe vesszük a követésben eltöltött évek számát, nem csak a vizsgált személyek számát (ezek szorzata adja meg a vizsgálat méretét, azaz a “number of persons at risk”-et).

Case-control vizsgálatok. Mintaválasztás, bevonási kritériumok. Beteg: általában minden beteg a praxisból. Kontroll: elvileg hasonlóknak kell lennie, kivéve a betegséget

(más betegek u. abból a kórházból, esetleg kórházi látogatók, szomszédok vagy véletlen minta a populációból). Párosítás lehetséges eloszlás szerint vagy egyedileg, 1-2 változó szerint. A párosításnak előnyei (teszt erejének növekedése, zavaró változók kiküszöbölése) és hátrányai is vannak (a párosító változók hatása nem vizsgálható).

A kérdőívvel, vagy interjúval kapcsolatos fő nehézségeket az emlékezési torzítás és a válaszmegtágadás jelenti, amelyek megfelelő kezelése befolyásolja a vizsgálat validitását.

Case-control vizsgálatban a relatív rizikó becslése nem olyan egyszerű, mint kohorsz vizsgálatban. Sem a R_{F+} , sem pedig a R_{F-} nem becsülhető, ezért a relatív rizikót az esélyhányadossal szokás becsülni. Ugyanazt a 2x2-es táblázatot használva, mint az előbb,

Betegség	+	-
Faktor		
+	a	b
-	c	d

és az esélyhányadost (*odds ratio*) OR -rel, becslését OR_B -rel jelölve, $OR_B = a / b : c / d$, ami annál kevésbé tér el az RR -től, minél kisebb a R_{F+} és a R_{F-} .

Bár az RR és az OR számszerűleg különbözhetnek, viszonyukkal kapcsolatban érdemes megjegyezni az alábbi összefüggéseket:

- mindkét mutatóra az 1-nél nagyobb értékek jelentik a kockázat növekedését, az 1-nél kisebbek a kockázat csökkenését, az 1 érték azt jelzi, hogy a kockázat nem változik
- ha valamelyikük eggyel egyenlő, akkor a másik is, ha valamelyikük egynél kisebb/nagyobb, akkor a másik is,
- az RR értéke mindig 1 és az OR értéke között van,
- ha statisztikai próbával tarthatónak bizonyul az $H_0:RR=1$, vagy a $H_0:OR=1$ (azaz, hogy „a kockázat nem változik”), akkor a másik is tartható

Az esélyhányadosra is igaz, hogy becslésének logaritmus, $\ln(OR_B)$ közelítőleg normális eloszlású, és hogy $E(\ln(OR_B)) = OR$, $\text{var}(\ln(OR_B))$ pedig $1/a + 1/b + 1/c + 1/d$. Ennek alapján a konfidencia intervallum $\ln(OR)$ -re:

$$\ln(OR_B) \pm u_\alpha \sqrt{1/a + 1/b + 1/c + 1/d},$$

ahol u_α ismét a normális eloszlás kritikus értéke.

A minta rétegzésére például akkor van szükség, ha zavaró változó hat (tudjuk, hogy ilyenkor előfordulhat, hogy minden rétegben 1-nél nagyobb OR esetén is adódhat az összevont táblában 1-nél kisebb OR , gondoljunk a confoundingról tanultakra, vagy a Simpson-paradoxonra). Ebben az esetben az OR becslésére (többek között) használhatjuk a Mantel-Haenszel féle módszert. Ha a rétegek száma r , és az i -edik réteghez a

Betegség	+	-
Faktor		
+	a_i	b_i
-	c_i	d_i

táblázat tartozik ($i = 1, 2, \dots, r$), és N_i -vel jelöljük a táblázat összegét ($N_i = a_i + b_i + c_i + d_i$) akkor az OR Mantel-Haenszel-féle becslése a következő

$$OR_{MH} = \frac{\sum_i \frac{a_i d_i}{N_i}}{\sum_i \frac{b_i c_i}{N_i}}$$

A Mantel-Haenszel becslés logaritmus is közelítőleg normális eloszlást követ, varianciájára pedig az alábbi közelítő képlet érvényes, ha az N_i gyakoriságok elég nagyok:

$$\text{var}(\ln(OR_{MH})) = \frac{\sum_i v_i w_i^2}{(\sum_i w_i)^2},$$

ahol $v_i = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}$ és $w_i = \frac{b_i c_i}{N_i}$.

Ennek alapján konfidencia-intervallumot is adhatunk a populációban a zavaró változó hatásának kiküszöbölése után érvényes $\ln(OR)$ értékre (és ebből az OR értékére is), mégpedig

$$\ln(OR_{MH}) \pm u_\alpha \sqrt{\text{var}(\ln(OR_{MH}))}$$

alakban, ahol u_α ismét a normális eloszlás kritikus értéke. (Ez a konfidencia-intervallum azonban ugyancsak közelítőleg érvényes, részben a Mantel-Haenszel eljárás alkalmazása, részben a többszöri közelítgetés miatt).

Példa: case-control vizsgálatban a táplálkozás magas vérnyomásra gyakorolt hatását vizsgálták (150 „eset”, 150 „kontroll”). Az alábbi táblázat szerint a betegek között több a zsíros ételeket rendszeresen fogyasztó, mint a kontroll csoportban:

Magas vérnyomás	+	-
Zsíros ételek		
+	89	57
-	61	93

Az esélyhányados becslése a táblázatból: $OR_B = 89 \cdot 93 / (61 \cdot 57) = 2.38$,

az esélyhányados logaritmus: $\ln(OR_B) = \ln(2.38) = 0.867$.

95%-os konfidencia-intervallum az esélyhányados logaritmusára:

$$\left(0.867 - 1.96 \cdot \sqrt{\frac{1}{89} + \frac{1}{57} + \frac{1}{61} + \frac{1}{93}} \dots 0.867 + 1.96 \cdot \sqrt{\frac{1}{89} + \frac{1}{57} + \frac{1}{61} + \frac{1}{93}} \right),$$

azaz – a számításokat elvégezve – (0.403 ... 1.331), amiből magára az esélyhányadosra az (1.50 ... 3.78) konfidencia-intervallum adódik.

Amikor a táblázatot életkor szerint bontották („rétegezték”), az alábbi két táblát kapták:

≤ 45 év			> 45 év		
Magas vérny. Zsíros ét.	+	-	Magas vérny. Zsíros ét.	+	-
+	8	30	+	81	27
-	10	60	-	51	33

A rétegenkénti esélyhányadosok:

$$OR_B^{(1)} = 8 \cdot 60 / (10 \cdot 30) = 1.60$$

$$OR_B^{(2)} = 81 \cdot 33 / (51 \cdot 27) = 1.94$$

Tehát az előbb kapott magas esélyhányados részben az életkorbeli különbségek zavaró hatásának tulajdonítható. Az esélyhányados Mantel-Haenszel-féle becslése az egyes rétegekbeli esélyhányadosok segítségével:

$$OR_{MH} = \frac{\frac{8 \cdot 60}{108} + \frac{81 \cdot 33}{192}}{\frac{10 \cdot 30}{108} + \frac{51 \cdot 27}{192}} = 1.85.$$

A konfidencia-intervallumhoz most is szükség van az esélyhányados logaritmusára:

$\ln(OR_{MH}) = \ln(1.85) = 0.615$. Ennek varianciája a fenti képlet szerint

$$\text{var}(\ln(OR_{MH})) = \frac{\left(\frac{1}{8} + \frac{1}{30} + \frac{1}{10} + \frac{1}{60}\right)\left(\frac{10 \cdot 30}{108}\right)^2 + \left(\frac{1}{81} + \frac{1}{27} + \frac{1}{51} + \frac{1}{33}\right)\left(\frac{51 \cdot 27}{192}\right)^2}{\left(\frac{10 \cdot 30}{108} + \frac{51 \cdot 27}{192}\right)^2} = 0.073,$$

ahonnan a konfidencia-intervallum az esélyhányados logaritmusára

$$(0.615 - 1.96 \cdot \sqrt{0.073} \dots 0.615 + 1.96 \cdot \sqrt{0.073}) \text{ azaz } (0.085 \dots 1.145),$$

és magára az esélyhányadosra (1.09 ... 3.14).

A nyers (korrigálatlan) és a Mantel-Haenszel-féle OR hányadosa egy lehetséges mértéke a zavaró változó által okozott torzításnak, amelyet “*confounding risk ratio*”-nak neveznek.

$$CRR = OR_{nyers} / OR_{MH}.$$

A fenti példában ez $2.38 / 1.85 = 1.29$.

A rizikófaktorok tulajdonítható kockázat becslése case-control elrendezésből. Az „attributable risk” a definíció alapján egyszerűen átírható és a szokásos közelítő módszerrel, a relatív rizikót az esélyhányadossal helyettesítve becsülhető:

$$AR = (R_{F+} - R_{F-}) / R_{F+} = 1 - 1/RR \approx 1 - 1/OR .$$

Hasonlóan módon adható közelítő becslés a „population attributable risk”-re is

$$PAR = (PR - R_{F-}) / PR = 1 - \frac{P(B+|F-)}{P(B+)} = 1 - \frac{P(F-|B+)}{P(F-)} \approx 1 - \frac{P(F-|B+)}{P(F-|B-)} .$$

Mindkét közelítés arra támaszkodik, hogy ha a betegség ritka, a potenciális rizikófaktor populációbeli előfordulása jól közelíthető a nem betegek közötti előfordulási arányával, azaz $P(F+) \approx P(F+|B-)$, illetve $P(F-) \approx P(F-|B-)$.

6. DIAGNOSZTIKAI TESZTEK

A diagnosztikai tesztek valamely mért mennyiség (vérnyomás, fehérvérsejtek száma, stb.) alapján vonnak le következtetést egy betegség fennállásával kapcsolatban. Értékelésükre általában a következő mutatókat használják (a képletekben T+/- a pozitív/negatív teszteredményt, B+/- pedig a betegség meglétét/hiányát jelenti):

a teszt szenzitivitása (*sensitivity*): $SE = P(T+|B+)$,

a teszt specifitása (*specificity*): $SP = P(T-|B-)$,

a teszt pozitív prediktív értéke (*positive predictive value*): $PV+ = P(B+|T+)$,

a teszt negatív prediktív értéke (*negative predictive value*): $PV- = P(B-|T-)$.

A mutatók értelmezése az alábbi táblázaton alapul (vegyük észre a rokonságot a statisztikai hipotézis-vizsgálattal).

Táblázat: diagnosztikai teszt értékelése

Betegség	van B+	nincs B-
Teszt-eredmény		
T+	$TP = B+ \cap T+$ (true positive)	$FP = B- \cap T+$ (false positive)
T-	$FN = B+ \cap T-$ (false negative)	$TN = B- \cap T-$ (true negative)

A szenzitivitás a helyes pozitív diagnózis valószínűsége, azaz a pozitív diagnózis valószínűsége, feltéve, hogy az illető valóban beteg. A specifitás a helyes negatív diagnózis valószínűsége, azaz a negatív diagnózis valószínűsége, ha az illető a vizsgált betegségben nem szenved (nem feltétlenül egészséges, csak a vizsgált betegségre negatív). Nyilvánvaló, hogy egy hasonló tünetekkel, elváltozásokkal járó másik betegség csökkentheti a diagnosztikai teszt specifitását.

A prediktív értékek a fordított feltételes valószínűségek, tehát hogy az illető beteg, feltéve, hogy teszteredménye pozitív, illetve, hogy egészséges, feltéve, hogy teszteredménye

negatív. Figyelem, a prediktív értékek nem csak a teszt tulajdonságaitól függenek, hanem a betegség prevalenciájától is! (A betegség prevalenciája a $P(B+)$ valószínűséggel azonos.)

A prediktív értékeket a Bayes-tétel segítségével számíthatjuk ki a szenzitivitásból és specificitásból:

$$PV+ = \frac{SE \cdot prev}{SE \cdot prev + (1 - SP) \cdot (1 - prev)}$$

$$PV- = \frac{SP \cdot (1 - prev)}{(1 - SE) \cdot prev + SP \cdot (1 - prev)}$$

A felsorolt mutatók meghatározása legtöbbször nem az egész populáció, hanem egy minta adataiból történik. Vegyük észre, hogy ehhez nem elegendők csupán a teszteredmények, hanem ismernünk kell a valódi állapotot is, amelynek meghatározása általában egy, a vizsgált tesztnél bizonyítottan sokkal megbízhatóbb teszt segítségével kell, hogy történjék.

Példa: prosztaták diagnosztizálása a prosztata manuális vizsgálatával, 300 vizeleti nehézségekkel kórházba utalt beteg adatainak alapján (forrás: Giunan et al., N.Engl.J.Med, 1980, 499-593). Elfogadjuk, hogy a biopsziával nyert minta szövettani eredményét a betegség pontos indikátorának tekinthetjük.

	Szövettan eredménye		
Manuális vizsgálat eredménye	+	-	
+	48	25	69
-	21	206	231
	73	227	300

E minta alapján a manuális vizsgálat szenzitivitása $SE = 48/73 = 0.66$, specificitása pedig $SP = 206/227 = 0.91$. A pozitív prediktív érték $PV+ = 48/69 = 0.70$, a negatív prediktív érték pedig $PV- = 21/231 = 0.09$. Ne felejtjük el azonban, hogy ezek a prediktív értékek egy olyan populációra érvényesek, amelyben a prevalencia $prev = 73/300 = 0.24$ (a vizeleti problémával kórházba utaltak populációja). Ha egy másik populációban a prevalencia kisebb, akkor ott a pozitív prediktív érték kisebb, a negatív pedig nagyobb lesz. Például egy olyan populációban, ahol a prosztaták prevalenciája csupán 1%, ott a vizsgálat pozitív prediktív értéke

$$\frac{0.66 \cdot 0.01}{0.66 \cdot 0.01 + (1 - 0.91) \cdot (1 - 0.01)} = \frac{0.0066}{0.0957} = 0.0069,$$

negatív prediktív értéke pedig

$$\frac{0.91 \cdot (1 - 0.01)}{0.91 \cdot (1 - 0.01) + (1 - 0.66) \cdot 0.01} = \frac{0.9009}{0.9043} = 0.9962.$$

Ha egy teszt szenzitivitását és/vagy specificitását egy mintából kiszámoljuk, akkor ezzel egy becslést adunk a teszt szenzitivitásának és/vagy specificitásának a populációra vonatkozó ismeretlen értékére. Ilyen esetben az ismeretlen értékre konfidencia-intervallumot is megadhatunk. Mivel mindkét mutató valószínűség, a valószínűségre adható szokásos konfidencia-intervallumokat használhatjuk. Alább megadjuk a jelenleg legjobbnak tekinthető aszimptotikus (nagy mintákra érvényes) konfidencia-intervallum, az adjusztált Wald-féle intervallum képletét. Erre csak abban az esetben van szükség, ha vagy nincs módunk számológépet használni, vagy a mintanagyság 1000-nél nagyobb, egyébként a Stern-féle egzakt módszert érdemes használni (leírását lásd Reiczigel, Stat.Med., 2003, 611-621, a megfelelő számítógépes program letölthető a <http://www.univet.hu/users/jreiczig/statmed2003/> címről).

Az adjusztált Wald-féle konfidencia-intervallum képlete:

$$\left(p_{adj} - u_{tábl} \sqrt{\frac{p_{adj}(1-p_{adj})}{n_{adj}}}, p_{adj} + u_{tábl} \sqrt{\frac{p_{adj}(1-p_{adj})}{n_{adj}}} \right),$$

ahol a betűk jelentése

	szenzitivitás becslése esetén	specificitás becslése esetén
n_{adj}	a betegek száma + $u_{tábl}^2$	az egészségesek száma + $u_{tábl}^2$
p_{adj}	$\frac{\text{a helyes pozitív diagnózisok száma} + u_{tábl}^2/2}{n_{adj}}$	$\frac{\text{a helyes negatív diagnózisok száma} + u_{tábl}^2/2}{n_{adj}}$
$u_{tábl}$	kritikus érték a normális eloszlás táblázatából (95%-os intervallumhoz 1.96, 90%-oshoz 1.65)	

Ha arra vagyunk kíváncsiak, hogy két diagnosztikai teszt szenzitivitása vagy specificitása megegyezik-e ($H_0: SE_1 = SE_2$, illetve $H_0: SP_1 = SP_2$), akkor két eset lehetséges:

1. két független mintánk van, az egyiket az egyik tesztet értékeltük ki, a másikon a másikat,
2. ugyanazon a mintán értékeltük ki mindkét tesztet.

Mindkét esetben végül is a szokásos khi-négyzet tesztet – vagy ha lehet, inkább a Fisher-féle egzakt tesztet – használjuk az összehasonlításra, de más-más táblázatra alkalmazva.

1. Két független minta

1. minta	$B+$	$B-$	2. minta	$B+$	$B-$
$T+$	TP_1	FP_1	$T+$	TP_2	FP_2
$T-$	FN_1	TN_1	$T-$	FN_2	TN_2

A 2x2-es táblázatok, amelyekre a khi-négyzet vagy a Fisher-féle egzakt tesztet alkalmaznunk kell:

Szenzitivitások összehasonlítására:		Specifitások összehasonlítására:	
TP_1	TP_2	FP_1	FP_2
FN_1	FN_2	TN_1	TN_2

Példa: Egy kutatásban (Karen, A. et al., *Reprod. Nutr. Dev.*, 43, 577-586) ultrahangos vizsgálatot alkalmaztak juhok vemhességének megállapítására (a vemhesség pontos állapotjelzőjének az állatok hormonális állapotát tekintették). Az alábbi táblázatok a mesterséges megtermékenyítés utáni 36. és 50. napi ultrahangos vizsgálat eredményeit foglalják össze. Feltételezve, hogy a két vizsgálatot független mintákon végezték, eltér-e a két időpontbeli vizsgálat szenzitivitása, illetve specifitása?

36. nap	vemhes	nem vemh.
$T+$	13	14
$T-$	18	110
	$SE=0.419$	$SP=0.887$

50. nap	vemhes	nem vemh.
$T+$	15	3
$T-$	14	116
	$SE=0.517$	$SP=0.974$

Az összehasonlításokhoz a Fisher-féle egzakt próbát (vagy ha erre nincs lehetőségünk, közelítő megoldásként a khi-négyzet próbát) az alábbi táblázatokon kell elvégezni:

Szenzitivitások összehasonlítása	
13	15
18	14
$p = 0.6051$	

Specifitások összehasonlítása	
14	3
110	116
$p = 0.0102$	

A kapott p -értékek (Fisher-egzakt teszt) alapján állíthatjuk, hogy a szenzitivitások nem különböznek szignifikánsan, de a specifitások igen.

2. Ugyanaz a minta mindkét teszthez

Ekkor az összehasonlításához csak azokat a mintaelemeket érdemes felhasználni, amelyekre a két teszt eltérő eredményt adott. Ha az alábbi táblázatból indulunk ki, akkor ez azt jelenti, hogy az összehasonlításban csak az a , b , c , d számokat fogjuk felhasználni, e -t és f -et nem.

	B+	B-
(i) T1+ és T2-	a	b
(ii) T1- és T2+	c	d
(iii) T1 és T2 azonos eredményt ad	e	f

Ha a két teszt szenzitivitása azonos, akkor a táblázatbeli (i) és (ii) eset minden mintaelemre 50-50% valószínűséggel fordul elő, tehát a szenzitivitások egyenlősége vizsgálható úgy, hogy azt teszteljük, hogy a táblázatbeli $a/(a+c)$ arány szignifikánsan eltér-e 50%-tól. Hasonlóképpen a specificitások egyenlőségének vizsgálatához azt teszteljük, hogy a $b/(b+d)$ arány szignifikánsan eltér-e 50%-tól. Erre bármely szokásos eljárás használható, de ha lehet, végezzünk egzakt binomiális próbát.

Példa: Tegyük fel most, hogy az előbbi példában szereplő, a vemhesség megállapítására hivatott ultrahangos vizsgálatokat ugyanazon a 200 állaton végzik el mindkét időpontban, és az alábbi eredményeket kapják:

	vemhes	nem vemh.
(i) a 36. napon +, az 50. napon – állatok száma	1	18
(ii) a 36. napon –, az 50. napon + állatok száma	6	3
(iii) mindkét napon azonos eredmény	37	135

A szenzitivitások ez esetben sem térnek el szignifikánsan ($p=0.1250$, binomiális próba $n=7$, $p=0.5$, $x=1$), a specificitások igen ($p=0.0015$, binomiális próba $n=21$, $p=0.5$, $x=3$).

Ha a két teszt összehasonlításában mindkét mutatót figyelembe szeretnénk venni, akkor az a kérdés, vajon a szenzitivitás vagy a specificitás a fontosabb, illetve hogy hogyan kombináljuk össze őket olyan mutatóvá, amelyik mindkettőt megfelelően tekintetbe veszi. Ilyenek léteznek, például a Youden-index ($SE + SP - 1$) vagy a helyes besorolások aránya ($accuracy$, $SE \cdot prev + SP \cdot (1 - prev)$).

A fenti táblázatban az a és d adja meg annak a gyakoriságát, hogy a T1 teszt helyesen dönt, a T2 hibásan, ugyanakkor a c és b annak, hogy T1 hibásan, T2 pedig helyesen. Ha azt a kérdést akarjuk tesztelni, hogy a helyes döntések aránya a két tesztnél egyenlő-e, akkor azt tesztelhetjük, hogy az $(a+d)/(a+b+c+d)$ arány szignifikánsan eltér-e 50%-tól.

Az ilyen kombinált mutatók hasznossága minden esetben a kétféle elkövethető hiba („fals pozitív” és „fals negatív”) következményeinek megítélésétől függ. Ezek a következmények egyébként gyakran kifejezhetők, mint plusz költségek. A fals pozitív diagnózis költségvonzata az alapos kivizsgálás költsége (amely végül kimutatja, hogy a betegség nem áll fenn), a fals negatív pedig az elkésett kezelés többletköltsége (gyakran ez a nagyobb összeg). A fejezet végén példát mutatunk egy egyszerű költségelemzésre.

Érdekes mutató a likelihood hányados (*likelihood ratio*), amely akkor kap fontos szerepet, ha a diagnózishoz több teszt eredményét akarjuk összekombinálni. A pozitív

teszteredményhez tartozó likelihood hányadost LR_+ , a negatív teszteredményhez tartozót LR_- jelöli. Értelmezésük:

$$LR_+ = \frac{\text{szenzitivitás}}{1 - \text{specifititás}} = \frac{P(T+ | B+)}{1 - P(T- | B-)} = \frac{P(T+ | B+)}{P(T+ | B-)},$$

$$LR_- = \frac{1 - \text{szenzitivitás}}{\text{specifititás}} = \frac{1 - P(T+ | B+)}{P(T- | B-)} = \frac{P(T- | B+)}{P(T- | B-)}.$$

A képlet szerint LR_+ (illetve LR_-) azt mondja meg, hogy hányszor akkora a pozitív (illetve negatív) teszteredmény valószínűsége, ha a páciens valóban beteg, mint ha nem beteg. Annál jobb a teszt, minél nagyobb az LR_+ és minél kisebb az LR_- .

Annak megértéséhez, hogy hogyan használhatók a likelihood hányadosok több teszteredmény kombinálásához, szükségünk van az „odds” fogalmának ismeretére.

Egy E esemény „odds”-án az $odds(E) = \frac{P(E)}{P(\bar{E})} = \frac{P(E)}{1 - P(E)}$ hányadost értjük, ahol \bar{E} az E

esemény ellentett eseménye ($\bar{E} = \text{„nem } E\text{”}$). Egy esemény valószínűsége és „odds”-a közötti összefüggést az alábbi táblázat illusztrálja:

valószínűség	1/100	1/10	1/4	1/2	3/4	9/10	99/100
odds	1/99	1/9	1/3	1	3	9	99

Látható, hogy az 1/2-nél nagyobb valószínűségek „odds”-a 1-nél nagyobb szám. Az esemény valószínűségét az odds-ból visszszámolni a $P(E) = \frac{odds(E)}{1 + odds(E)}$ képlettel lehet.

A pozitív teszteredmény likelihood hányadosának algebrai átalakításával azt kapjuk, hogy

$$LR_+ = \frac{P(T+ | B+)}{P(T+ | B-)} = \frac{\frac{P(B+ | T+)P(T+)}{P(B- | T+)P(T+)}}{\frac{P(B- | T+)P(T+)}{P(B-)}} = \frac{P(B+ | T+)}{P(B- | T+)} \cdot \frac{P(B-)}{P(B+)},$$

ahonnan

$$\frac{P(B+ | T+)}{P(B- | T+)} = LR_+ \cdot \frac{P(B+)}{P(B-)}.$$

Itt a baloldalon lévő törtet „post-test odds”-nak, a jobboldalit „pre-test odds”-nak nevezik. A „pre-test odds” az az esély (jobban mondva, „odds”), amelyet a teszt elvégzése előtt adunk a szóban forgó betegségnek, a „post-test odds” pedig az, amelyet a teszt elvégzése után pozitív teszteredmény esetén tulajdonítunk a betegségnek. Ezek szerint LR_+ a betegség esélyeinek („odds”-ának) változását fejezi ki, feltéve, hogy a teszteredmény pozitív.

Teljesen hasonló számítás vezet oda, hogy $\frac{P(B+ | T-)}{P(B- | T-)} = LR_- \cdot \frac{P(B+)}{P(B-)}$, azaz LR_- a

betegség esélyeinek megváltozása a teszt előttihez képest akkor, ha a teszt-eredmény negatív.

Több teszt alkalmazása esetén a teszteredmények kombinálásához a megfelelő LR-eket (a teszteredménytől függően LR+ vagy LR-) lánc-szerűen össze kell szorozni. (A kapott „post-test odds” a végén – a fent említett módon – visszaszámolható valószínűséggé.)

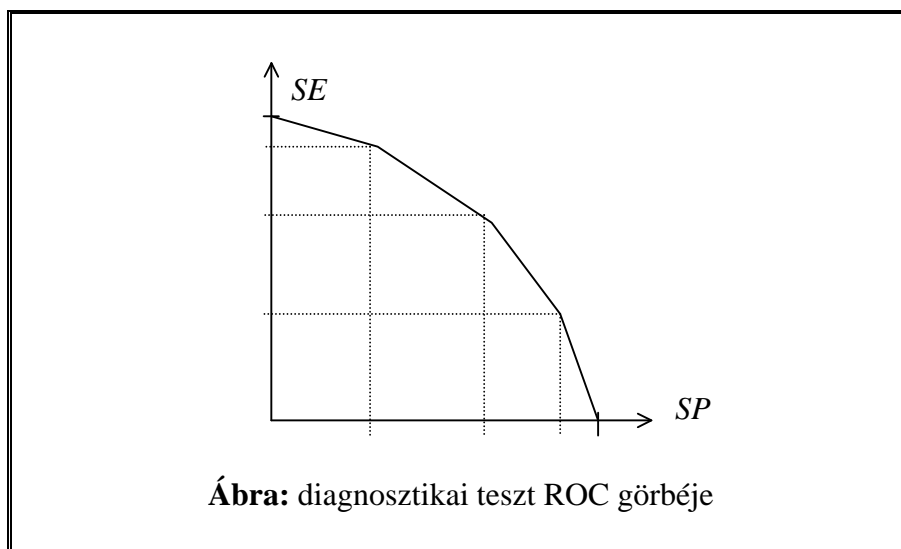
Példa: torokgyulladás esetén arra, hogy a kórokozó streptococcus-e, többek között a következő táblázatban szereplő diagnosztikai tesztek szokták alkalmazni, amelyek szenzitivitására és specificitására vonatkozóan az irodalomban az alábbi adatokat találtuk. Az utolsó sorban az ezekből számított LR+ értékeket tüntettük fel.

	magas láz (>38.3)	strept. fertőzés a beteg környezetében	a kórokozó kitenyésztése
szenzitivitás	0.33	0.18	0.73
specificitás	0.89	0.92	0.96
LR+	3.00	2.25	18.25

Számoljuk ki, hogy ha a streptococcus okozta torokgyulladás prevalenciája az adott populációban 10%, mi az esélye, hogy a vizsgált betegnél is ez a helyzet

- a) ha a betegnek 39 fokos a láza:
 - a „pre-test odds” itt a prevalenciából adódik: $0.1 / 0.9 = 0.11$,
 - a „post-test odds” $0.11 \cdot 3.00 = 0.33$,
 - az ebből visszszámolt valószínűség $0.33 / 1.33 = 0.25$.
- b) ha 39 fokos a láza, és a testvére éppen most esett át streptococcus okozta fertőzésen:
 - mivel itt két pozitív teszteredmény van, a „post-test odds” $0.11 \cdot 3.00 \cdot 2.25 = 0.74$,
 - az ebből visszszámolt valószínűség $0.74 / 1.74 = 0.43$.
- c) ha mindezek mellett még a tenyésztés is pozitív eredményt adott:
 - most három pozitív teszteredménnyel kell számolnunk, így a „post-test odds” $0.11 \cdot 3.00 \cdot 2.25 \cdot 18.25 = 13.51$,
 - az ebből visszszámolt valószínűség pedig $13.51 / 14.51 = 0.93$.

Árnyaltabb összehasonlításra ad lehetőséget, ha a diagnosztikai tesztek a ROC görbékkel jellemezzük (ROC = *receiver operating characteristic*, a rádiótechnikából kölcsönzött kifejezés, néha OC görbének is nevezik). A ROC görbe a teszt szenzitivitása és specificitása közötti összefüggést leíró függvény görbéje (lásd az ábrát). Ha a szóban forgó teszt egy mért változó (pl. vérnyomás) értékén alapul, mondjuk a teszt pozitív, ha a mért érték egy bizonyos határértéket meghalad, akkor a határérték változtatásával együtt változik a teszt szenzitivitása és specificitása, és ilyen módon a ROC görbe egyes pontjait kapjuk.



Lássunk egy példát! Tegyük fel, hogy egyfajta vesebetegséget jelezhet, ha a vizeletben egy bizonyos anyag szintje megemelkedik. Mivel az anyag kimutatása egyszerű, diagnosztikus tesztként be kívánják vezetni, ezért 100 beteg és 100 a szóban forgó betegségben nem szenvedő emberen kipróbálják.

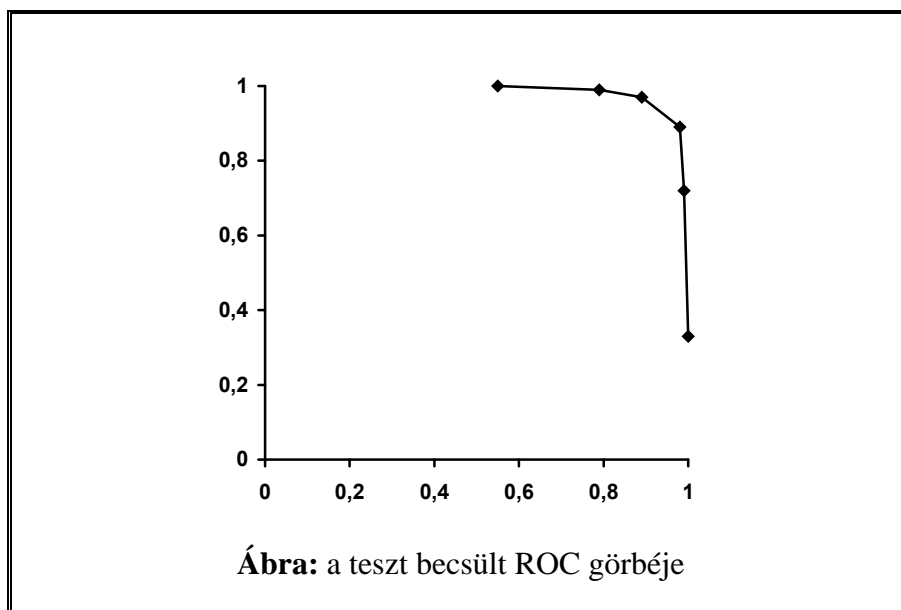
Példa: a teszt ROC görbéjének szerkesztéséhez szükséges adatok

Szint (mmol/l)	<5	5-10	11-20	21-30	31-50	51-100	>100	Össz.
Egészséges	55	24	10	9	1	1	0	100
Beteg	0	1	2	8	17	39	33	100

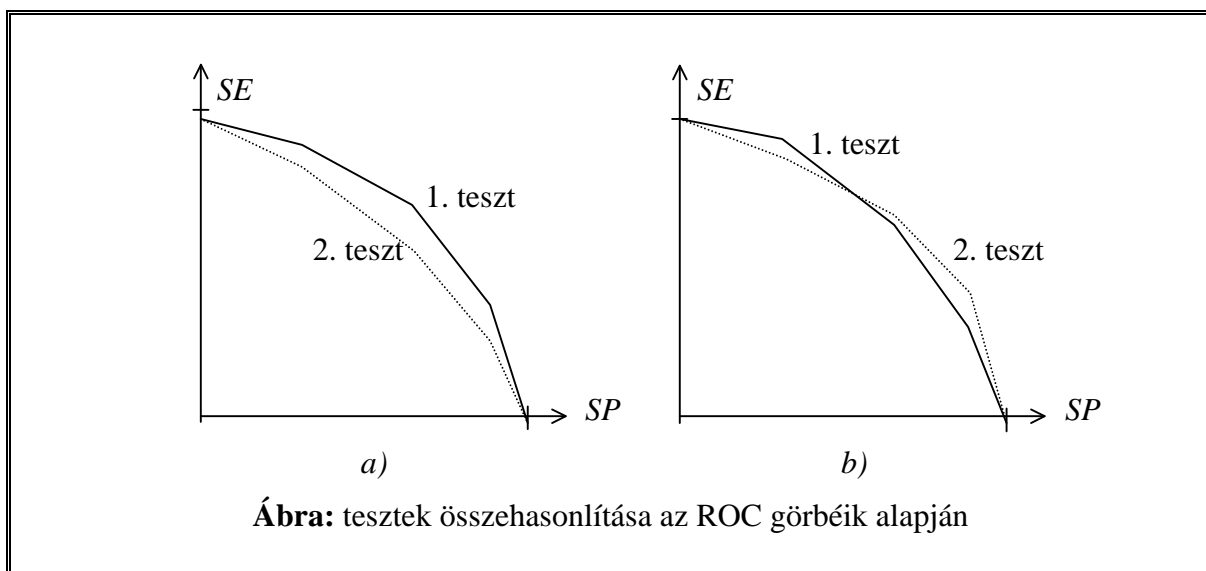
A fenti táblázat alapján az ROC görbe mintából becsült pontjait a szenzitivitás és specifitás definíciója szerint kapjuk.

Táblázat: a teszt ROC görbéjének pontjai

Választott határérték	≥ 5	≥ 11	≥ 21	≥ 31	≥ 51	≥ 101
Specifitás $SP = P(T-/B-)$	0.55	0.79	0.89	0.98	0.99	1.00
Szenzitivitás $SE = P(T+/B+)$	1.00	0.99	0.97	0.89	0.72	0.33



Tegyük fel, hogy két tesztet az ROC görbéik alapján akarunk összehasonlítani. Ha a görbék az *a*) ábrán látható viszonyban vannak, azt mondhatjuk, hogy az 1. teszt globálisan jobb, mint a 2. teszt. Kevésbé szerencsés esetben a görbék viszonya olyan, amelyet a *b*) ábra mutat, ami azt jelenti, hogy bizonyos specificitás értékek esetén az 1. teszt szenzitivitása a nagyobb. Ilyen esetben globális mutatóként a ROC görbe alatti területet szokás alkalmazni (persze ez ellen is ugyanolyan kifogások emelhetők, mint bármelyik másik kombinált mutató alkalmazása ellen).



A fentiek olyan diagnosztikai tesztekre vonatkoznak, amelyek egy mért változón alapulnak. Ha több változó figyelembe vételével szeretnénk diagnosztikai tesztet konstruálni, akkor alkalmazhatunk hierarchikus döntési struktúrát („döntési fát”), amelyben az egyes változók értékeit egymás után vesszük figyelembe, vagy olyan modellt, amelyben az egyes változók szimultán vesznek részt a döntésben (pl. logisztikus regressziós modellt).

Most lássunk egy példát egy egyszerű költségelemzésre, amelyet például egy szűrővizsgálat bevezetése előtt célszerű elvégezni, hogy fel lehessen mérni a szűrővizsgálat bevezetésével elérhető megtakarítás mértékét! Tegyük fel, hogy

1. a teszt költsége 1.500 Ft, szenzitivitása 97%, specificitása 99%,
2. az alapos kivizsgálás költsége 24.000 Ft,
3. a kezelés költsége korai felismerés esetén 200.000 Ft,
4. a kezelés költsége késői felismerés esetén 2.500.000 Ft,
5. a betegség prevalenciája a célpopulációban 0.2%.

Az egy főre jutó átlagos költség a teszt alkalmazása esetén

$$1500 + 0.998 \cdot 0.01 \cdot 24.000 + 0.002 \cdot \{0.97 \cdot (24.000 + 200.000) + 0.03 \cdot 2.500.000\} = 2324 \text{ Ft}$$

Az egy főre jutó átlagos költség a teszt alkalmazása nélkül

$$0.002 \cdot 2.500.000 = 5000 \text{ Ft.}$$

Vegyük észre, hogy az elérhető megtakarítás nem csak a teszt szenzitivitásától és specificitásától függ, hanem a betegség prevalenciájától is. Ez azt jelenti, hogy a megtakarítás nemcsak a teszt javításával növelhető, hanem a célpopuláció jó megválasztásával is (azaz annak az okos megválasztásával, hogy kikre terjedjen ki a szűrés).

(Fa-diagram?)

7. KLINIKAI KÍSÉRLETEK

A klinikai kísérlet (*clinical trial*) olyan tudományosan megtervezett kísérlet (*planned experiment*), amelyben egészséges személyek vagy betegek vesznek részt, és amelynek célja bizonyos terápiákra (gyógyszer vagy egyéb kezelés) vonatkozó ismeretszerzés, illetve ezen ismeretek pontosítása (Pocock, 1983).

A klinikai kísérletek speciális tulajdonságai:

- a minta nem áll rendelkezésre egyidőben,
- a megfigyelési egységek élőlények, ezért nem zárható ki az interakciójuk a kísérletet végzővel, etikai kérdésekre különösen tekintettel kell lenni.

A gyógyszerkísérletek osztályozása

- I. fázis: gyógyszerkinetika, toxicitás vizsgálata (egészséges önkéntesek, kérdés inkább a biztonságosság, mint a hatékonyság, $n \leq 100$)
- II. fázis: előkészítő hatásvizsgálatok, dózis-hatás elemzés (betegek, állandó felügyelet, $n \leq 200$)
- III. fázis: terápiás hatás klinikai vizsgálata (kezelés és placebo vagy kezelés és standard kezelés összehasonlítása)
- *Amikor klinikai kísérletről beszélünk, általában erre gondolunk*
 - *Ez nem csak gyógyszeres kezeléssel, hanem más terápiával is értelmes*
- IV. fázis: bevezetés utáni további vizsgálatok, többnyire felmérések (mellékhatások felderítése, marketing)

Talán a legnagyobb kísérlet a gyermekbénulás elleni Salk-féle vakcina vizsgálata volt (USA, 1954, $n \approx 1.8$ millió).

Kísérlettervezés

A protokollnak tartalmaznia kell a következőket:

- a kísérlet céljának megfogalmazása (specifikus hipotézisek is!)
- a célpopuláció megválasztása (beválasztási-kizárási kritériumok)
- a kezelés pontos specifikációja
- a hatás mérése (elsődleges-másodlagos hatásmutató / kimeneti változó / végpont)
- a kontroll típusa, a randomizáció módja (rétegzés?), a betegek tájékoztatása, a hozzájárulás-kérés módja, vakság
- az alkalmazandó statisztikai elemzés, a minimálisan szükséges mintaelemszám
- a vizsgálat végrehajtásának terve: egy- vagy többcentrumos-e, a betegek követésének módja, teendők a követés félbeszakadása esetén, stb.
- a kísérlet monitorozása, adatkezelés, adminisztratív és egyéb felelősségek tisztázása, stb.

Klinikai kísérletek egy fontos csoportosítása:

Kérdésfeltevés:	tudományos/elméleti (<i>theoretical design</i>)	pragmatikus (<i>pragmatic design</i>)
Cél:	okszági összefüggés kimutatása	klinikai döntéselőkészítés
Minta:	lehetőleg homogén egy faktor hatását vizsgáljuk	reprezentatív a célpopulációra nézve komplex terápia hatását vizsgáljuk

Ha azt a feltevést, hogy a daganat ellen alkalmazott besugárzás hatása előzetes gyógyszeres kezeléssel fokozható, kísérletileg is igazolni szeretnénk, akkor az alábbi két kísérleti elrendezés közül melyiket válasszuk?

Idő	1. kísérleti elrendezés		2. kísérleti elrendezés	
	I. (kezelt)	II. (kontroll)	I. (kezelt)	II. (kontroll)
1. hét	gyógyszer	placebo	gyógyszer	besugárzás
2. hét	gyógyszer	placebo	gyógyszer	besugárzás
3. hét	gyógyszer	placebo	gyógyszer	besugárzás
4. hét	besugárzás	besugárzás	besugárzás	
5. hét	besugárzás	besugárzás	besugárzás	
6. hét	besugárzás	besugárzás	besugárzás	

Beválasztási-kizárási kritériumok (*inclusion/exclusion criteria, eligibility criteria*) azok a feltételek, amelyek alapján eldöntjük, hogy mely páciensek (kísérleti egységek) kerüljenek a mintába. Vonatkozhatnak pl. a beteg életkorára, a betegség súlyosságára, más betegség fennállására, stb. Szerepük az, hogy biztosítsák,

- a kísérlet reprodukálhatóságát,
- a kiesés drop-out esélyének csökkentését,
- a csoportok homogenitását,
- a kísérlet külső validitását,
- a confounding elkerülését.

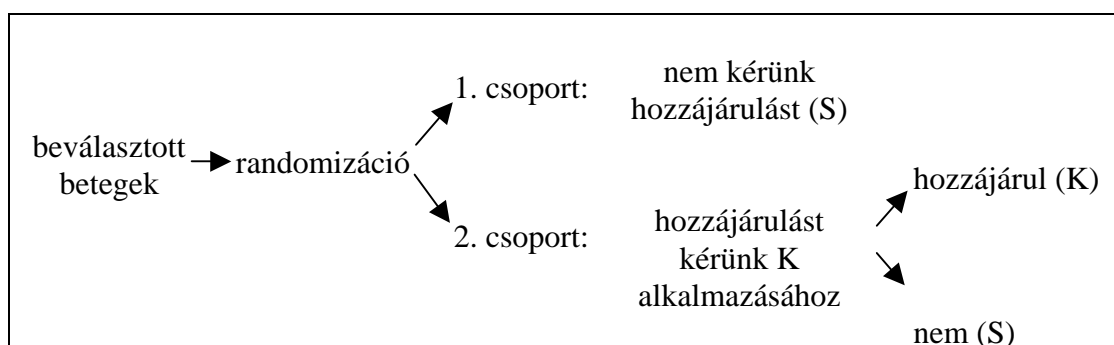
Ha a célváltozó alacsony vagy magas értéke beválasztási kritérium, akkor a „regression to the mean” („visszatérés az átlaghoz”) látszólagos hatást okozhat! Példa: tegyük fel, hogy egy kísérletben vérnyomáscsökkentő szert vizsgálunk, és az egyik beválasztási kritérium az, hogy a vizsgálatkor a páciens szisztolés vérnyomása ≥ 160 Hgmm legyen. Ha két hét elteltével a beválasztott betegeken a mérést megismételjük, kezelés nélkül is nagy valószínűséggel azt tapasztaljuk, hogy a mért értékek átlaga kisebb, mint az első mérés volt. Ezt a különbséget a két mérés között kapott kezelésnek tulajdoníthatnánk, ha nem alkalmaznánk placebo-kontroll csoportot, amelyben ez a hatás szintén megjelenik.

A „visszatérés az átlaghoz” jelensége olyan esetekben tapasztalható, amikor a megfigyelt érték nem elhanyagolható mértékű véletlen ingadozást mutat a „valódi” vagy „jellemző” érték, pl. átlagérték körül. (Például a vérnyomás mért értéke a páciensre jellemző átlagérték körül ingadozik.) Ha ilyenkor két mérést végzünk, természetes, hogy az első

méréskor legnagyobb értéket mutató páciensek nem azonosak a másodikban legnagyobb értéket mutatókkal. Ezért az első méréskor a legnagyobb értéket mutatókat kiválasztva, a második méréskor átlagosan az elsónél valamivel kisebb értékre számíthatunk. Hasonló a helyzet akkor is, ha azokat választjuk be, akik egy küszöbérték alatti értéket produkálnak: ekkor a második mérés átlaga várhatóan az elsónél nagyobb lesz.

Általános elv: mindig először kell dönteni a kísérletbe való beválaszthatóságról, majd tájékoztatni a beteget és megkérni a hozzájárulását, és csak azután lehet dönteni a kezelés típusáról, randomizálva.

Egy alternatíva: a Zelen-féle "randomized consent design" (Pocock, 1983, 108-109.)
(S - standard kezelés, K - kísérleti kezelés)



Módszerek a szisztematikus hibák elkerülésére. kontroll, randomizáció, vaktság, "analysis by intention to treat".

Előfordulhat, hogy a terápia hatását mérő változó (célváltozó) megfigyelt értékét befolyásolják a kezeléssel kapcsolatos elvárások (különösen, ha a célváltozót nem műszeresen mérik, pl. tünetek súlyossága), ami az eredményeket eltorzíthatja.

Vak (*blind*), illetve kettős vak (*double-blind*) kísérletről akkor beszélünk, ha a páciens, illetve sem a páciens, sem a kísérletben részt vevő orvosok – legfőképp aki a célváltozó mérését végzi – nem ismerik, hogy a páciens melyik kezelést kapja. A vaktság szerepe a placebo-hatás kivédése (minthogy így a kezelt és a kontroll csoport egyformán részesedik belőle).

Olykor sok fantázia kell ahhoz, hogy hogyan titkosíthatunk egy kísérletet, ha a kezelés nem csupán egy tableta bevétele, hanem pl. fizioterápia vagy műtét.

A párhuzamos csoportos kísérletekben a potenciális zavaró változók hatásának kiszűrése érdekében fontos a randomizáció, vagyis az egyedek véletlenített besorolása az egyes kísérleti csoportokba.

Randomizációs módszerek

- teljes randomizáció: előre kisorsolt, borítékolt kezelések a teljes tervezett mintára.
- blokkos randomizáció: ha biztosítani akarjuk, hogy ne különbözzék lényegesen az egyes kezelések előfordulási gyakorisága akkor sem, ha a kísérletet a tervezettnél hamarabb be kell fejezni, a kisorsolást végezzük úgy, hogy minden valahányadik (10-ik, 20-ik, stb.) beteg után az egyes kezelések ugyanannyiszor forduljanak elő.
- kiegyensúlyozott randomizáció: ha az a cél, hogy bizonyos változók, pl. nem, kor, stb. az egyes kezelési csoportokban azonos megoszlást mutassanak, akkor az egyes kezeléseket

sorsoljuk mindig változó valószínűségekkel, mégpedig aszerint változtatva, hogy a soron következő beválasztott beteg férfi vagy nő, idős vagy fiatal, stb. (NB. Itt nem lehet előre borítékolni!)

Analysis by intention to treat („a szándékolt kezelés szerinti elemzés”): minden beteget abban a kezelési csoportban elemzünk, ahová eredetileg be volt sorolva – akkor is, ha abbahagyta a kezelést, és akkor is, ha egy kontrollcsoportbeli páciens menet közben kezelésre szorult! (Ellentéte a "per protocol analysis", amelyben mindenkit a ténylegesen kapott kezelésnek megfelelő csoportban elemzünk.) Az "intention to treat" mellett szól – de egyébként is tanulságos – az alábbi klasszikus példa, amelyben (leírását lásd a New England J. of Med., 1980, ??????) a koleszterin-szint csökkentésére szánt clofibrate nevű szert és placebot hasonlítottak össze kettős vak, randomizált kísérletben. A hatásmutató az ötéves mortalitás volt, nevezetesen 18% a kísérleti, 19% a kontroll csoportban. Kiderült azonban, hogy a kísérleti csoportban csak a résztvevők kétharmada szedte valóban a gyógyszert, és ezek körében a mortalitás 15%-os volt, szignifikánsan kisebb, mint a kontrollé. Ez azt látszott bizonyítani, hogy a szer hatásos. Azonban a további elemzés során az is kiderült, hogy a kontroll csoportban szintén csak a résztvevők kb. 2/3-a szedte a tablettákat, és ezek körében is csupán 15%-os volt a mortalitás.

Fontos követelmény a kísérletekkel szemben a reprodukálhatóság. A reprodukálhatóság és az általánosíthatóság azonban gyakran egymásnak ellentmondó követelmények. Például egy kísérletet végezhetünk egy bizonyos beltenyésztett egértörzsszel vagy genotipikusan heterogén mintával. Homogén mintát választva a reprodukálhatóság javul, de az általánosíthatóság romlik, heterogén mintával pedig fordítva.

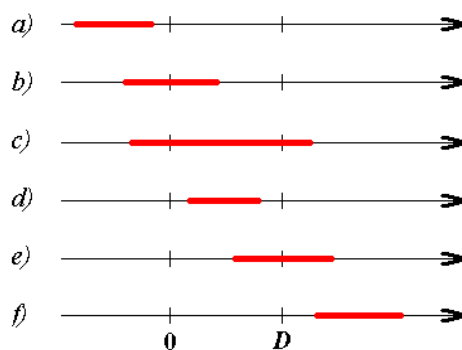
Etikai kérdések

- A beteg tájékoztatása és önkéntes hozzájárulása szükséges a kísérletben való részvételhez.
- A randomizáció csak akkor etikus, ha a terápiák között nincs tudományosan bizonyított különbség.
- stb.

A negatív eredmény interpretációja

Ha a kísérleti kezelést nem találjuk szignifikánsan jobbnak, mint a standard kezelést (vagy a placebót, amennyiben placebó-kontrollos volt a kísérlet), felmerül a kérdés, hogy ezért vajon a hatás csekély volta, vagy a mintaelemszám elégtelensége felelős. Ilyenkor segíthet, ha ismerjük (pontosabban, ha ki tudjuk számítani) az alkalmazott statisztikai teszt erejét.

Szintén segíthet az interpretációban, ha a hatásnagyságra (pontosabban a kísérleti és a standard kezelés hatása közötti különbségre) konfidencia-intervallumot tudunk készíteni. Az alábbi ábra bemutatja, hogy milyen következtetéseket lehet levonni e konfidencia-intervallum elhelyezkedéséből. Az ábrán a 0 pont a két kezelés hatásának egyenlőségét reprezentálja, a negatív tartomány azt, hogy a kísérleti kezelés hatása elmarad a standard kezelésétől, D pedig a legkisebb, klinikailag még releváns különbséget jelöli a kísérleti és a standard kezelés hatása között, tehát azt a minimális különbséget, amelynél még érdemes a standard kezelést az újjal felváltani. (Ezt megfogalmazhatjuk abszolút vagy relatív különbségként, az adott szakterületen szokásos módon.)



- a) Az új kezelés rosszabb
- b) Az új kezelés nem jobb, mint a hagyományos (vagy rosszabb, vagy pedig a különbség klinikailag irreleváns)
- c) Az eredmény nem egyértelmű, nagyobb kísérletre van szükség.
- d) Az új kezelés jobb, de a különbség klinikailag irreleváns
- e) Az új kezelés jobb, de a különbség nem biztos, hogy klinikailag releváns
- f) Az új kezelés jobb.

Ekvivalencia-kísérletek (equivalence trials)

Ekvivalencia-kísérletről akkor beszélünk, ha két kezelés – például egy új és a standard – hatásának egyenértékűségét szeretnénk bizonyítani. Jelöljük a két kezelést A-val és B-vel (pl. két gyógyszer), a hatás mérésére szolgáló számszerű változót pedig X-szel (pl. vérnyomás, fájdalom-szám, stb).

Ha ilyenkor a hatások egyenlőséget választanánk nullhipotézisnek, akkor – még ha az eredmény a H_0 megtartása lenne is – csak azok hinnének benne (továbbra is), akik már a kísérlet előtt sem kérdőjelezték meg.

Az ekvivalenciát csak az bizonyíthatja a kétkedők számára is meggyőzően, ha az eredmények alapján mind a

$$H_0': B \text{ hatásosabb, mint } A,$$

mind pedig a

$$H_0'': B \text{ kevésbé hatásos, mint } A$$

hipotézisek elvethetők.

Itt persze a „hatásosabb”-at a klinikailag még releváns különbség figyelembe vételével szokták érteni. Például ha a hatások közötti 15%-nál kisebb különbség orvosi szempontból már nem releváns, azaz a kezelések orvosilag egyenértékűnek tekinthetők, ha

$$0.85 \mu_A < \mu_B < 1.15 \mu_A$$

ahol μ_A és μ_B az X célváltozó átlagértéke a két kezelés mellett, akkor a fenti két hipotézisvizsgálat

$$H_0': 1.15\mu_A \leq \mu_B,$$

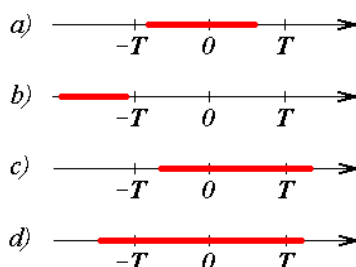
illetve

$$H_0'': \mu_B \leq 0.85\mu_A$$

-ra módosul. (Megjegyezzük, hogy bár az ekvivalencia-tartományt leggyakrabban valóban relatív különbség – százalékos eltérés – formájában szokás definiálni, éppenséggel használhatnánk abszolút különbséget is, minden ugyanúgy menne akkor is.)

Ha az eredmények mindkét nullhipotézis elvetéséhez vezetnek, akkor a kísérlet statisztikailag (empirikusan) bizonyította a kezelések ekvivalenciáját.

A kezelések ekvivalenciájáról dönthetünk konfidencia-intervallumok segítségével is: ezt a módszert is eléggé elterjedt. Lényege az, hogy konfidencia-intervallumot számolunk a két kezelés hatása közötti eltérésre ($\mu_A - \mu_B$, vagy μ_A / μ_B , attól függően, hogy abszolút vagy relatív eltéréssel kívánunk dolgozni), és a kapott intervallum helyzete alapján döntünk az ekvivalenciáról. Ha az ekvivalencia-tartományt a $(-T, T)$ intervallum jelöli (amely az előbbi $\pm 15\%$ -os relatív eltérések megfelelője), akkor a konfidencia-intervallum helyzete lényegében az alábbiak valamelyike lehet (illetve a *b*) és *c*) eseteknek a szimmetrikus párja is előfordulhat)



- A két kezelés ekvivalens
- Az *A* kezelés kevésbé hatásos mint a *B*
- Az *A* kezelés legalább olyan jó, mint a *B* (ekvivalens vagy jobb)
- A kísérlet nem hozott egyértelmű eredményt a két kezelés viszonyát tekintve (nagyobb kísérletre van szükség)

8. KÍSÉRLETI ELRENDEZÉSEK

Már említettük, hogy a párhuzamos csoportos kísérletekben a potenciális zavaró változók hatásának kiszűrése érdekében fontos a randomizáció, vagyis az egyedek véletlenül besorolása az egyes kísérleti csoportokba (az epidemiológiai vizsgálatokkal kapcsolatban már láttuk, hogy olyan változók, amelyek függetlenek a kezeléstől, nem lehetnek zavaró változók). Lehetőség van azonban a potenciális zavaró tényezők hatásának szisztematikus kontrollálására is. Ilyen kérdésekkel a kísérlettervezés, és azon belül a kísérleti elrendezések elmélete (*experimental design*) foglalkozik.

Ha például az a gyanúnk, hogy egy tényező nem kívánatos hatást gyakorol a célváltozó értékére, akkor ezt a tényezőt lehet úgy kontrollálni, hogy teljes randomizálás helyett a tényező szerint rétegzünk és minden rétegben mindegyik kezeléssel ugyanannyit allokálunk, csupán az egyes rétegeken belül randomizálunk. Így az egyes kezelésekhöz tartozó átlagértékeket a tényező azonos mértékben befolyásolja. Ezt a *rétegzett randomizációt* alkalmazó elrendezést *véletlen blokkos elrendezésnek* nevezik (a rétegek a blokkok), a teljes randomizációt pedig *teljes véletlen elrendezésnek*.

A kísérleti elrendezések elméletét először a mezőgazdasági kísérletekkel kapcsolatban alkalmazták, sok szakkifejezés eredete is erre utal, például a megfigyelési egységeket hagyományosan *parcellának* (*plot*) nevezik. Lássunk most mi is egy példát egy növénytermesztési kísérlettel kapcsolatban. A kísérlethez kijelölt (mondjuk téglalap alakú) táblát a megfelelő számú, azonos méretű parcellára osztják (a parcellák száma az alkalmazandó kezelések száma szorozva a megkívánt pontossághoz szükséges ismétlésszámmal). Ha a parcellákat minden szempontból azonosnak lehet tekinteni, akkor a kezeléseket random módon célszerű a parcellákhoz allokálni (teljes véletlen elrendezés). Ha azonban ismeretes, hogy mondjuk kelet-nyugati irányban változik a talaj vízellátottsága, és ez hatással van a termés várható mennyiségére, akkor olyan blokkokat kell alkotni, amelyeken belül a vízellátottság-beli különbségek már elhanyagolhatók, és a kezeléseket úgy allokálni, hogy minden blokkba minden kezelésből azonos számú – általában 1 – ismétlés essék (és blokkon belül persze randomizálni).

Példa: teljes véletlen és véletlen blokkos elrendezés növénytermesztési kísérletben (három kezelés öt ismétléssel)

homogén tábla

K1	K2	K1	K3	K1
K2	K3	K2	K1	K3
K1	K3	K2	K3	K2

a) teljes véletlen elrendezés
(teljesen randomizált)

inhomogenitás

B1	B2	B3	B4	B5
K1	K2	K1	K3	K2
K2	K3	K3	K2	K1
K3	K1	K2	K1	K3

b) véletlen blokkos elrendezés
(blokkon belül randomizált)

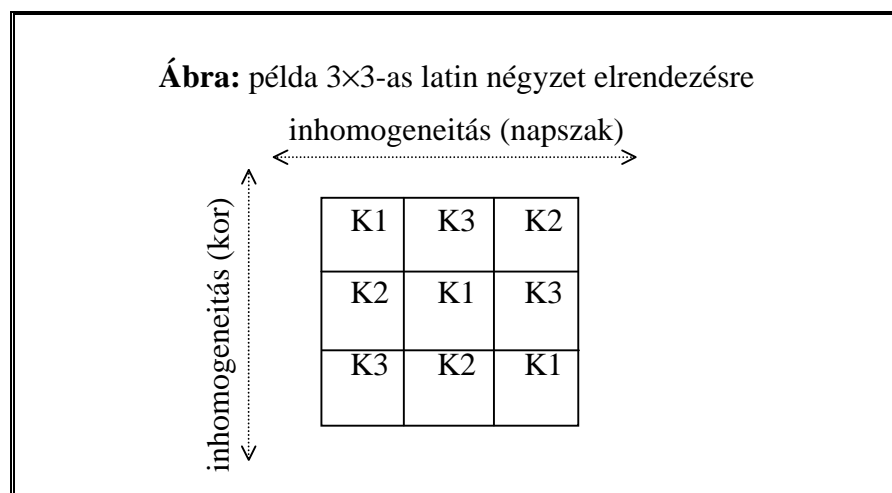
Természetesen orvosi kísérletben is lehet szerepe a blokkolásnak. Például ha a vizsgálat időigényes, és az a gyanúnk, hogy a mért értékek napszak szerint tendenciózan változnak, akkor érdemes blokkos elrendezést választani (pl. reggeli, délelőtti, délutáni, esti blokk, minden blokkban minden kezelésből egy mérés).

Ha a kísérletet varianciaelemzéssel értékeljük ki, akkor blokkos elrendezés esetén a teljes négyzetösszeget az alábbi módon bontjuk fel (jelölje a kezelések számát a , a blokkokét b , és az egyszerűség kedvéért minden blokkban minden kezelésből egy ismétlés legyen):

Négyzetösszeg	Szabadsági fok
Teljes	$ab-1$
Kezelés hatása	$a-1$
Blokk hatás	$b-1$
Reziduális	$(a-1)(b-1)$

Az elrendezés garantálja, hogy a kezelés hatása és a blokk-hatás szétválasztható, az pedig, hogy a szórásfelbontásban a blokk-hatásnak tulajdonítható szórás el van különítve a véletlen hibától, erősebb tesztet eredményez. (Igaz viszont, hogy a kezelés \times blokk interakció, amennyiben van ilyen, nem választható szét a véletlen hibától, de hát nem szabad telhetetlennek lenni.)

Latin négyzet elrendezés. Egy fokkal bonyolultabb elrendezés arra az esetre, ha egytényezős kísérletben két potenciális zavaró változót szeretnénk szimultán kontrollálni, mondjuk, ha az előző, orvosi példában a napszak szerinti változás mellett a mért értékek függenek a páciens életkorától is. A megoldás: mindkét zavaró tényező szerint blokkokat képzünk, mégpedig mindkettőből ugyanannyi blokkot, mint az összehasonlítandó kezelések száma (mondjuk, ha 3 kezelést kell összehasonlítani, akkor a blokkok lehetnek napszak szerint reggel, délelőtt, délután, kor szerint fiatal, középkorú, idős), és a kezeléseket úgy allokáljuk, hogy minden kezelésből mindkét tényező minden blokkjába ugyanannyi essék. Az alábbi ábrán egy ilyen elrendezés látható:



Érdekes kérdés, hányféle különböző módon lehet k kezelést elrendezni egy $k \times k$ méretű négyzetben úgy, hogy minden kezelés minden sorban és minden oszlopban pontosan egyszer forduljon elő. Mindenesetre a kezeléseket e feltétel teljesítése mellett véletlenszerűen kell allokálnunk (praktikusan például egy szabályos elrendezésből kiindulva a sorok és oszlopok véletlen cserélgetésével). Látható, hogy a latin négyzet elrendezés akkor kivitelezhető könnyen, ha a zavaró változók folytonosak, mert ekkor könnyen hozhatunk létre belőlük épp a kezelések számával megegyező számú kategóriát.

Latin négyzet elrendezés esetén a varianciaelemzésben a teljes négyzetösszeget az alábbi módon bontjuk fel (a a kezelések száma, de ugyanennyi a sorok és az oszlopok száma is):

Négyzetösszeg	Szabadsági fok
Teljes	a^2-1
Kezelés hatása	$a-1$
Sor hatás	$a-1$
Oszlop hatás	$a-1$
Reziduális	$(a-2)(a-1)$

Az eddigi kísérletekben minden megfigyelési egységen egyetlen mérést végeztünk. Többféle ok miatt (de leggyakrabban az egyedek közötti különbségek miatt) szükség lehet arra, hogy minden egyes egyedben több mérést végezzünk. Az ilyen kísérleteket ismételt méréses kísérletnek (*repeated measures experiment*) nevezzük. Ide tartozik többek között a következő, klinikai kísérletek esetén gyakran alkalmazott elrendezés.

Crossover elrendezés. Két kezelés (A és B) összehasonlítása esetén elképzelhető, hogy bár a kezelések hatása különbözik, de olyan nagy szóródást mutat a kísérleti alanyok között, hogy két párhuzamos csoporttal dolgozva a kezelések közötti különbség a nagy csoporton belüli szórások miatt nem lesz kimutatható. Ekkor a két kezelés közötti különbséget csak úgy van remény kimutatni, ha mindkét kezelést ugyanazon az alanyon alkalmazzuk, egyiket a másik után.

Ezt az elrendezést legtöbbször krónikus betegségekkel kapcsolatban alkalmazzák (asztma, magas vérnyomás, cukorbetegség, reuma, stb.), amikor a kezeléstől gyógyulás nem várható, csak a tünetek enyhítése, az állapot javítása. Ilyenkor, ha az elsőnek alkalmazott kezelést abbahagyjuk, az állapot visszaáll a kezdeti állapotra, és a második kezelést is ki lehet próbálni. Az eljárás nem alkalmazható, ha a kimenetel gyógyulás, sem akkor, ha a kimenetel halál. Alkalmazása nem előnyös, ha a kezelések túl hosszantartóak, mert ilyenkor a drop-out valószínűsége jelentősen megnő.

A legegyszerűbb elrendezésben két csoporttal dolgozunk, az egyik csoportban a kezeléseket AB , a másikban BA sorrendben alkalmazzuk („ AB/BA design”).

IV) A beteg - ismeretlen okból - eltűnik a követésből (*drop-out*). Ez nehézségeket okozhat az elemzésben, ha az eltűnés nem független a kezelés eredményétől (ha pl. valószínűsíthető, hogy a gyógyultak nem jönnek vissza a kontroll-vizsgálatokra, vagy pl. ha azok, akik nem javulnak, más megoldást keresnek).

A túlélési görbék elemzéséhez az alábbi függvényeket szokták használni.

1) A beavatkozástól az eseményig eltelt idő (T) eloszlás-, illetve sűrűségfüggvénye

$$F(t) = P(T < t), \text{ illetve ha } F(t) \text{ abszolút folytonos, } f(t) = F'(t).$$

2) A "túlélési függvény" (*survival function*)

$$S(t) = P(T > t) = 1 - F(t) - P(T = t).$$

Tegyük fel a továbbiakban, hogy $F(t)$ abszolút folytonos, ekkor a $P(T = t)$ tagokkal nem kell vesződnünk.

3) A "hazard függvény" (*hazard rate function*)

$$\lambda(t) = f(t) / S(t) = F'(t) / S(t).$$

$\lambda(t)$ annak a kockázatát méri, hogy az esemény éppen a t időpontban (illetve annak végtelen kis környezetében) következik be, feltéve, hogy addig nem következett be. Felhasználva, hogy $F(t) = 1 - S(t)$,

$$\lambda(t) = -S'(t) / S(t) = -d/dt \ln(S(t)).$$

4) A "kumulatív hazard függvény" (*cumulative hazard function*)

$$\Lambda(t) = \int_0^t \lambda(u) du = -\ln(S(t)), \text{ amellyel } S(t) = e^{-\Lambda(t)}.$$

A túlélési függvény mintából való becslésére a leggyakrabban használt két nemparaméteres eljárás a „life table” módszer és a Kaplan-Meier módszer. Mielőtt ezeket ismertetnénk, lássuk, hogy néz ki egy minta ebben az esetben. Ha nem lenne cenzorálás, akkor a minta egy adatsort tartalmazna, az egyedek túlélési idejét. A cenzorálás azt eredményezi, hogy egyes egyedekre nem a túlélési idő, hanem egy annál kisebb szám szerepel, a követés hossza (ezeket egy, a szám után írt + jellel fogjuk jelölni). Az értékeket nagyság szerint rendezhetjük, így például egy 10 elemű minta esetén az adatok a következő formát öltik (pl. műtét utáni túlélési idők hónapban):

4, 6+, 7, 7+, 9, 9, 10+, 13, 15, 22+

Ebben az adatsorban 4 cenzorált megfigyelés van, a második, 4-ik, 7-ik és az utolsó.

A „life table” módszer

Ennél a módszernél az időt önkényesen intervallumokra osztjuk, és ezekre az intervallumokra vonatkozóan végezzük el a túlélési függvény becslését. Minden intervallumban az alábbi módon számolunk:

$$\text{esemény empirikus valószínűsége} = \text{események száma} / \text{összlétszám},$$

ahol a cenzorálás miatt az összlétszámot többféleképpen is érthetjük:

1. összlétszám az intervallum elején

(nem törődünk a cenzorálással, azaz úgy vesszük, hogy a cenzoráltak mind túlélnek az intervallum végéig)

2. összlétszám az intervallum elején, mínusz a cenzorált megfigyelések (redukált minta)
3. összlétszám az intervallum elején, mínusz $\frac{1}{2}$ cenzorált megfigyelések („arany középút” – a következő példában ezt használjuk)

A túlélési függvény becslését az

$$(1 - \text{az esemény empirikus valószínűsége az intervallumban})$$

kifejezések összeszorozásával kaphatjuk, hiszen a túlélés valószínűsége egy intervallum végéig nem más, mint a

$$\text{túlélés valószínűsége az interv. elejéig} \cdot (1 - \text{az esem.emp.valószínűsége az interv.ban}).$$

Illusztrációként végezzük el a számításokat az előző adatsorral! Válasszunk 3 hónapos intervallumokat!

Intervallum	Létszám az interv. elején	Események száma	Cenz. száma	Korrigált létszám	Esemény empirikus valósz.	Túlélési függvény becslése
$0 \leq t < 3$	10	0	0	10	0	1
$3 \leq t < 6$	10	1	0	10	.100	.900
$6 \leq t < 9$	9	1	2	8	.125	.788
$9 \leq t < 12$	6	2	1	5.5	.364	.501
$12 \leq t < 15$	3	1	0	3	.333	.334
$15 \leq t < 18$	2	1	0	2	.500	.167
$18 \leq t < 21$	1	0	0	1	0	.167
$21 \leq t < 24$	1	0	1	0.5	0	.167

A Kaplan-Meier módszer

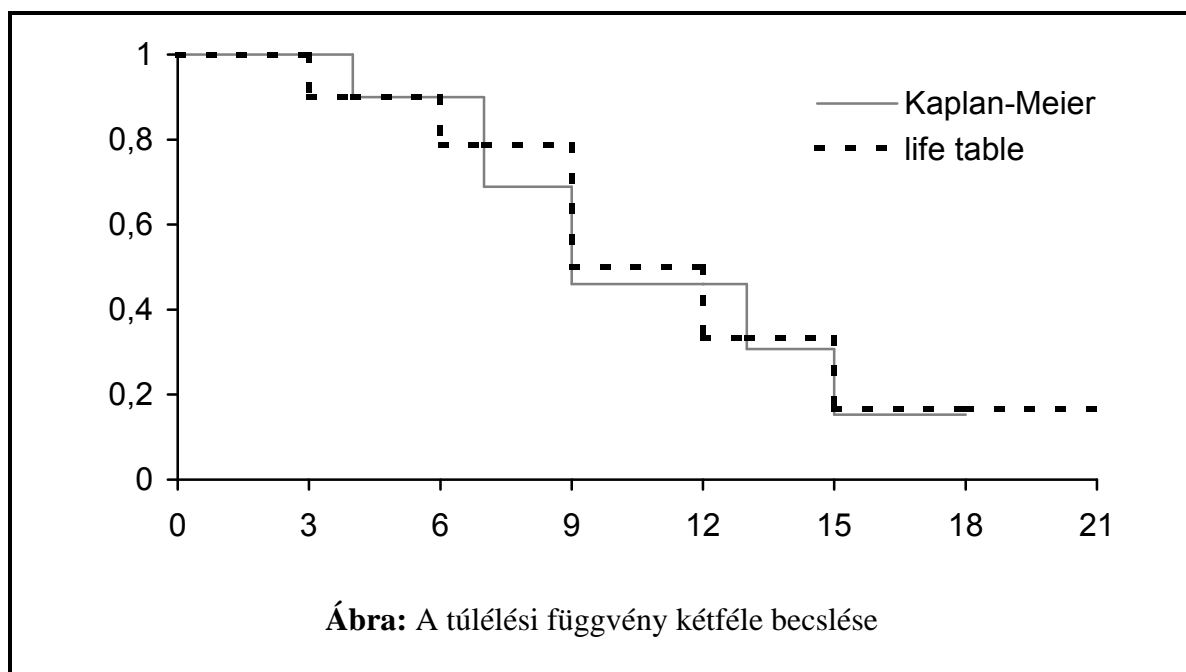
Elsődlegesen abban különbözik a „life table” módszertől, hogy itt nem önkényesen választott intervallumokkal dolgozunk, hanem az intervallumokat az események időpontja határozza meg. Egyéb tekintetben a számítások hasonlóan történnek. (Eltérések: itt az intervallumok jobbról zártak, azaz jobboldalukon van egyenlőségjel, a cenzoráltakat teljes egészében levonjuk, mint az előbb a redukált minta esetén, és azonos időpontra eső esemény és cenzorálás esetén úgy vesszük, mintha a cenzorálás végtelenül kicsi idővel később történt volna).

Illusztrációként lássuk a számításokat az előző adatokkal:

$$4, 6+, 7, 7+, 9, 9, 10+, 13, 15, 22+$$

Intervallum	Létszám az interv. elején	Események száma	Cenz. Száma	Korrigált létszám	Esemény empirikus valósz.	Túlélési függvény becslése
$0 < t \leq 4$	10	1	0	10	.100	.900
$4 < t \leq 7$	9	1	1	8	.125	.689
$7 < t \leq 9$	7	2	1	6	.333	.460
$9 < t \leq 13$	4	1	1	3	.333	.307
$13 < t \leq 15$	2	1	0	2	.500	.153
$15 < t \leq 18$	1	0	0	1	.000	.153

Az alábbi ábrán a kapott túlélési görbék szokásos grafikus ábrázolása látható.



A Mantel-Haenszel-féle teszt két túlélési görbe összehasonlítására

Túlélési görbék összehasonlítására többféle eljárás is ismeretes. Jellemző eset, amikor két görbét kell összehasonlítani, erre legtöbbször a Mantel-Haenszel tesztet használják.

A nullhipotézis az, hogy a két csoportban a hazard mindig azonos, vagyis az, hogy az események különböző időpontokban következnek be a két csoportban, csupán a véletlennek tulajdonítható.

A teszt statisztika az események tényleges és várható száma közötti (megfelelően normált) különbségek összege az egyik csoportra vonatkozóan, amelynek null-eloszlása aszimptotikusan standard normális eloszlás. A tényleges és várható számok közötti

különbségeket csak azokra az időpontokra kell összegeznünk, amelyekben volt esemény. A következő példa segíthet, ha mindezt kézzel kell végigszámolnunk. Legyenek az adatok:

1. csoport: 1, 2+, 5, 5+, 10, 18+,

2. csoport: 4, 4, 7, 8+, 10, 10+.

Idő	1. csoport		2. csoport		Létszám			Es.össz.	Eltérés
	Es.	Cen.	Es.	Cen.	Össz.	1. cs.	2. cs.		
1	1	0	0	0	12	6	6	1	$\frac{1-1\cdot6/12}{\sqrt{1\cdot11\cdot6\cdot6/12^2\cdot11}} = 1.00$
2	0	1	0	0					
4	0	0	2	0	10	4	6	2	$\frac{0-2\cdot4/10}{\sqrt{2\cdot8\cdot4\cdot6/10^2\cdot9}} = -1.22$
5	1	1	0	0	8	4	4	1	$\frac{1-1\cdot4/8}{\sqrt{1\cdot7\cdot4\cdot4/8^2\cdot7}} = 1.00$
7	0	0	1	0	6	2	4	1	$\frac{0-1\cdot2/6}{\sqrt{1\cdot5\cdot2\cdot4/6^2\cdot5}} = -0.71$
8	0	0	0	1					
10	1	0	1	1	4	2	2	2	$\frac{1-2\cdot2/4}{\sqrt{2\cdot2\cdot2\cdot2/4^2\cdot3}} = 0.00$
18	0	1	0	0					

Összesen: 0.07
(5% krit. érték = 1.96)

azaz a különbség a két csoport túlélési függvényei között nem szignifikáns.

Az "Eltérés" oszlop számítási módja a fenti táblázatban:

$$Eltérés = \frac{[1.csop.es.] - \frac{[Es.össz.] \cdot [1.csop.létsz.]}{[Össz.létsz.]}}{\sqrt{\frac{[Es.össz.] \cdot [Össz.létsz. - Es.össz.] [1.csop.létsz.] [2.csop.létsz.]}{[Össz.létsz.]^2 [Össz.létsz. - 1]}}$$

10. BIOASSAY

Egyes, az orvosi- és biológiai kutatásban használatos anyagok (vakcinák, vérekészítmények, antibiotikumok, stb.), hatóanyagtartalma, erőssége nem mérhető ki fizikai vagy kémiai mérési módszerekkel, ezért a hatás erősségét egy biológiai rendszernek (sejttenyészetnek, kísérleti állatnak, stb.) az anyagra adott válaszát megmérve lehet csak számszerűsíteni. Az ilyen mérési módszerek neve „*biological assay*” vagy röviden „*bioassay*”. Gyakran az a helyzet, hogy egy új készítményt („*teszt*”) egy már ismert készítménnyel („*standard*”) kell összehasonlítani a biológiai hatásukat tekintve („*comparative assay*”). A „*standard*”

készítmény lehet egy ismert anyag nemzetközileg szabványosított mennyisége („nemzetközi egység”).

Direkt assay esetén egy adott biológiai válasz eléréséhez szükséges dózist mérik ki, indirekt assay esetén csak néhány rögzített dózissal dolgoznak és az ezekre adott válaszokból vonnak le következtetéseket (ez utóbbi módszer a gyakoribb). Ha a biológiai válasz folytonos mennyiség, akkor „kvantitatív assay”-ről, ha pedig dichotom változó, akkor „kvantális assay”-ről beszélünk.

Kvantitatív assay esetén úgy képzeljük, hogy a válasz a dózis valamely monoton függvénye, amelyet „dózis-válasz görbének” (*dose-response curve*) nevezünk. Természetesen azt is feltesszük, hogy a válasz random komponenst is tartalmaz, biológiai változatosság, illetve mérési hibák miatt. A modell tehát

$$y = f(x) + \varepsilon$$

ahol x a dózis, y a válasz, ε pedig 0 várható értékű véletlen hiba (gyakran azt is felteszik, hogy normális eloszlású). Jegyezzük meg, hogy ez a modell implicite feltételezi, hogy az egyedek közötti biológiai variabilitás a függvény alakját nem érinti.

Ha két anyagot hasonlítanak össze, az egyszerűség kedvéért azt szokták feltételezni, hogy a két anyag hatása között csak mennyiségi különbség van, úgy, mint ha az egyik a másiknak a hígítása lenne. Formálisan ez azt jelenti, hogy a két anyag „dózis-válasz görbéi” közötti kapcsolat a következő

$$y_1 = f(x) + \varepsilon \quad \text{és} \quad y_2 = f(\rho x) + \varepsilon,$$

ahol ρ az úgynevezett relatív erősség (*relative potency*). E feltevés híján az összehasonlítás nehezzé válik, ha például a két görbe metszi egymást, akkor valamely dózis-tartományban az egyik, egy más dózis-tartományban pedig a másik anyag hatásosabb.

A továbbiakban először a „kvantitatív assay” két legelterjedtebb típusát ismertetjük.

Parallel line assay

Ennél a assay-típusnál az a kiindulási feltevés, hogy a válasz a dózis logaritmusának lineáris függvénye (vö. a Weber-Fechner-törvénnyel), azaz

$$y_1 = f(x) + \varepsilon = \alpha + \beta \ln(x) + \varepsilon \quad (\text{standard}),$$

$$y_2 = f(\rho x) + \varepsilon = \alpha + \beta \ln(\rho x) + \varepsilon = \alpha + \beta \ln(\rho) + \beta \ln(x) + \varepsilon \quad (\text{teszt}),$$

azaz ha a dózis-válasz görbéket egy olyan koordináta-rendszerben ábrázoljuk, amelyben az x tengelyen logaritmikus a skála, két párhuzamos egyenest kapunk (neve is innen ered).

A kiértékelést egy speciális eseten, a „6 pontos parallel line assay”-n (*6-point parallel line assay*) mutatjuk be. Ekkor mindkét anyaggal 3 dózist (kis, közepes, nagy) alkalmaznak.

Kiértékelés varianciaelemzéssel. A varianciaelemzés szokásos, legegyszerűbb alapeseténél (főhatások és interakció tesztelése) finomabb elemzésre nyújt lehetőséget a *kontrasztok* alkalmazása. A kontrasztok a kezelésekből képzett lineáris kombinációk (azzal a feltétellel, hogy az együtthatók összege 0), amelyek nulla volta egyenként tesztelhető. Akkor hasznos ez a megközelítés, ha sikerül a feladat szempontjából értelmes jelentéssel bíró kontrasztokat találni. Két kontrasztot ortogonálisnak nevezünk, ha az együtthatóik szorzatösszege 0. Az ortogonalitás azért fontos tulajdonság, mert bizonyítható, hogy ortogonális kontrasztok ugyanabból a mintából egymástól függetlenül tesztelhetők. A 6 pontos bioassay esetében például a következő teszteket (de nem csak ezeket!) végezhetjük megfelelően megválasztott kontrasztok segítségével.

- C1 – a standard és a teszt készítmény hatása közötti átlagos különbség,
 C2 – a dózis hatása (az egyenesek közös meredeksége),
 C3 – az egyenesek párhuzamosságtól való eltérése,
 C4 – a dózis-hatás görbe lineáristól való eltérése a standardra,
 C5 – a dózis-hatás görbe lineáristól való eltérése a teszt készítményre.

Az alábbi táblázat mutatja a C1, ... C5 kontrasztokhoz tartozó együtthatókat (S1, S2, S3 a standard, T1, T2, T3 a teszt készítmény 3-3 dózisékat jelöli). Könnyű meggyőződni róla, hogy a fenti kontrasztok ortogonálisak, vagyis hogy a fenti tesztek függetlenek egymástól.

Táblázat: a 6 pontos parallel line assay kiértékelése ortogonális kontrasztok segítségével

	S1	S2	S3	T1	T2	T3
C1	-1	-1	-1	+1	+1	+1
C2	-1	0	+1	-1	0	+1
C3	-1	0	+1	+1	0	-1
C4	+1	-2	+1	0	0	0
C5	0	0	0	+1	-2	+1

Megjegyezzük, hogy a 6 pontos parallel line assay kiértékelhető regressziószámítás alkalmazásával is, de talán a varianciaelemzés a fenti kontrasztokkal fogalmilag tisztább és könnyebben megérthető.

Egy egyszerűbb típus a 4 pontos parallel line assay, amikor mindkét készítménynek csak 2 dózisékat alkalmaznak. Természetesen itt kevesebb dolgot lehet tesztelni, csak 3 ortogonális kontraszt alkotható (4 csoport esetén a szabadsági fok 3). Szokásosan ezek a következők:

- C1 – a standard és a teszt készítmény hatása közötti átlagos különbség,
 C2 – a dózis hatása (az egyenesek közös meredeksége),
 C3 – az egyenesek párhuzamosságtól való eltérése,

Táblázat: a 4 pontos parallel line assay kiértékelése ortogonális kontrasztok segítségével

	S1	S2	T1	T2
C1	-1	-1	+1	+1
C2	-1	+1	-1	+1
C3	-1	+1	+1	-1

Slope ratio assay

Ennél az assay-típusnál az a kiinduló feltevés, hogy a válasz lineárisan függ a dózistól, azaz a két dózis-válasz görbe a következő:

$$y_1 = f(x) + \varepsilon = \alpha + \beta x + \varepsilon \quad (\text{standard}),$$

$$y_2 = f(\rho x) + \varepsilon = \alpha + \beta \rho x + \varepsilon \quad (\text{teszt}),$$

Látható, hogy a két regressziós egyenes egyenletében közös a konstans tag (α). Ezt, valamint a görbék linearitását célszerű tesztelni az assay kiértékelésekor. A kiértékelés szokásosan regressziószámítás segítségével történik. (A linearitás és a közös α tesztelésére a regressziós egyenletbe további tagokat kell bevonni és tesztelni azok 0 voltát.) A relatív erősség (ρ) becslésére egyszerűen az alábbi formula szolgál:

$$\hat{\rho} = \frac{\hat{\beta}_2}{\hat{\beta}_1}.$$

A két legegyszerűbb elrendezés ebből a típusból az 5 pontos és a 3 pontos slope ratio assay. Az elsőben mindkét készítményből 2 dózist, valamint egy 0 dózist alkalmaznak, a másodikban a 0 dózison kívül mindkét anyagból csupán 1 dózist alkalmaznak. (Ez utóbbi már nem teszi lehetővé sem a linearitás, sem pedig a közös α tesztelését.)

Végül röviden említést teszünk a dichotom válasz esetén használatos kvantális assay-k kiértékelésére használatos logit és probit modellekről.

Kvantális assay esetén, mivel a válasz dichotom, azt szokták feltételezni, hogy a dózis és a pozitív válasz valószínűsége között áll fenn regressziós kapcsolat. Mivel a valószínűség korlátos tartományban mozoghat (a dózis viszont végtelen tartományban), speciális regressziós modellekre van szükség. Itt segít a logit transzformáció, amely kölcsönösen egyértelmű leképezést létesít a $(0,1)$ és a $(-\infty,\infty)$ között:

$$\text{logit}(p) = \ln \frac{p}{1-p}.$$

A logit modell egy, a válasz logit transzformációját felhasználó lineáris regressziós modell:

$$\text{logit}(p) = \alpha + \beta x.$$

Hasonló elven alapul a másik modell is. A probit transzformáció a $(0,1)$ és a $(-\infty,\infty)$ közötti kölcsönösen egyértelmű leképezés céljára a normális eloszlás sűrűségfüggvényét, $\Phi(x)$ -t használja fel

$$\text{probit}(p) = \Phi^{-1}(p),$$

amivel a regresszió az alábbi alakot ölti:

$$\text{probit}(p) = \alpha + \beta x.$$

IRODALOM

Ahlbom, A. - Norell, S. E. *Introduction to modern epidemiology*. Epidemiology Resources Inc., Chestnut Hill, 1990.

Altman, D. G. *Practical statistics for medical research*. Chapman & Hall, 1991.

-
- Armitage, P. - Berry, G. *Statistical methods in medical research* (3. Ed.). Blackwell Science, Oxford, 1994.
- Breslow, N. E. - Day, N. E. *Statistical methods in cancer research. Vol 1: The design and analysis of case-control studies*. International Agency for Research on Cancer, Lyon, 1980.
- Breslow, N. E. - Day, N. E. *Statistical methods in cancer research. Vol 2: The design and analysis of cohort studies*. International Agency for Research on Cancer, Lyon, 1987.
- Cohran, W. S. *Planning and analysis of observational studies*. Wiley, 1983.
- Finney, D. J. *Statistical method in biological assay*. Griffin, 1978.
- Fleiss, J. L. *The Design And Analysis of Clinical Experiments*. Wiley, 1986.
- Freeman, P. R. The Performance of the 2-stage Analysis of 2-treatment, 2-period Crossover Trials. *Statistics in Medicine*, **8**, 1421-1432. (1989)
- Holland, P. W. Statistical and causal inference. *JASA*, 1986, pp.945-970.
- Miller, R. G. *Survival analysis*. Wiley, 1981.
- Pocock, S. J. *Clinical trials: a practical approach*. Wiley, 1983.
- Rothman, K. J. *Modern Epidemiology*. Little Brown, 1986.
- Yamaguchi, K. *Event history analysis*. Sage, London, 1991.