

# Az orvosi megismerés módszertana

Ferenci Tamás

tamas.ferenci@medstat.hu

<http://www.medstat.hu/>

<https://www.youtube.com/c/FerenciTamas>

Utoljára frissítve: 2022. december 2.

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés
- Az empirikus orvosi kutatások alapgondolata és a kauzalitás
- A confounding problémája
- A confounding megoldásai: megfigyelés és kísérlet
- Okozati következtetések levonása megfigyeléses adatokból
- A véletlen szerepéről
- Bizonyitekok hierarchiája és összessége
- A végpont és meghatározása

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés

- Az empirikus orvosi kutatások alapgondolata és a kauzalitás
- A confounding problémája
- A confounding megoldásai: megfigyelés és kísérlet
- Okozati következtetések levonása megfigyeléses adatokból
- A véletlen szerepéről
- Bizonyitekok hierarchiája és összessége
- A végpont és meghatározása

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

# Az orvosi kutatások egy általános sémaja felé

- Az orvostudomány egy jelentős része – egész története alatt – igen egyszerű alakban megfogalmazható kérdések megválaszolásából, kutatásából állt:
  - A koponyalékelés hatásos-e a migrén ellen?
  - Hat-e ez a gyógyfű a cukorbetegségre?
  - Csökkenti-e ez a vérnyomáscsökkentő gyógyszer-jelölt a vérnyomást, azaz hatásos-e?
  - Okoz-e hasmenést mint mellékhatást ugyanez a gyógyszer-jelölt?
  - Magasfeszültségű távvezeték közelében tartózkodás növeli-e a rákkockázatot?
  - A vöröshús-fogyasztás növeli-e a rákkockázatot?
  - A dohányzás hogyan hat az infarktus utáni túlélésre?
  - A császármetszéssel születés növeli-e a T1DM rizikót?
  - Az anyai acetaminophen-szedés növeli annak valószínűségét, hogy a gyerek autista lesz?
- Hogyan adjunk ezekre választ...?

# Vizsgálati módszerek

Ilyen és ehhez hasonló kérdések számos módon vizsgálhatóak:

- Elméleti (biológiai) megfontolások
- Matematikai modellek
- Más területekről vett analógiák
- In vitro kísérletek
- Állatkísérletek
- stb. stb.

De ami most számunkra a legfontosabb lesz: a (humán) *empirikus* vizsgálatok

# Az empirikus orvoslás talán első példája

10 És mondá az udvarmesterek fejedelme Dánielnek: Félek én az én uramtól, a királytól, aki megrendelte a ti életeket és italotokat; minek lássa, hogy a ti orcátok hitványabb amaz ifjakénál, akik egykorúak veletek? és így búnbe kevernétek az én fejemet a királynál. 11 És mondá Dániel a felügyelőnek, akire az udvarmesterek fejedelme bízta vala Dánielt, Ananiást, Misáelt és Azariást: 12 Tégy próbát, kérlek, a te szolgáiddal tíz napig, és adjanak nékünk zöldségféléket, hogy azt együnk, és vizet, hogy azt igyunk. 13 Azután mutassák meg néked a mi ábrázatunkat és amaz ifjak ábrázatát, akik a király ételével élnek, és aszerint cselekedjél majd a te szolgáiddal. 14 És engede nékik ebben a dologban, és próbát tőn velük tíz napig. 15 És tíz nap mulva szébbnek látszék az ő ábrázatuk, és testben kövérebbek valának mindenazonál az ifjaknál, akik a király ételével élnek vala.

Dokumentálás helye: Biblia, Dániel könyve, 1. fejezet (Károli Gáspár fordítása)

# Mik ezzel a bajok?

Nagyon jó, de felmerül azért pár kérdés is:

- Dániel beszerezte a Regionális Kutatásetikai Bizottság engedélyét a kutatáshoz?
- A résztvevők teljes írásos tájékozott beleegyezéssel vettek részt a kísérletben?
- Regisztrálta Dániel, természetesen előzetesen, a kutatást nemzetközi adatbázisban (pl. ClinicalTrials.gov-on)?
- Nem világos a végpont meghatározása: a „szebbnek látszik az ö ábrázatuk” pontosan milyen módon került operacionalizálásra? Hiányzik a használt kvantitatív mérési eljárás kellő pontosságú megadása.
- Nem derül ki, hogy a kísérleti alanyok randomizálásra kerültek-e, illetve milyen módszerrel.
- Nem világos, hogy a vizsgálók, illetve az alanyok vakosítva voltak-e az ételek tekintetében.
- Az eredményközlés elégtelen: hiányzik a végpontokon mért numerikus kimenet, és szignifikanciára vonatkozó statisztikai próba dokumentálása.

DE!

DE! A fenti mégis fantasztikus: felmerült a gondolat (kb. i.e. 200-ban vagyunk!), hogy a kérdést **empirikus** alapon kell megválaszolni! **Tények alapján** (nem szent iratok, sámánok, vakszerencse vagy tapasztalati sejtés alapján)!

# A világ legszélesebb körben alkalmazott terápiás eljárása



# A világ legszélesebb körben alkalmazott terápiás eljárása

A középkorból:



# A világ legszélesebb körben alkalmazott terápiás eljárása

Az ókori Görögországból:



# A világ legszélesebb körben alkalmazott terápiás eljárása

XIX. századi fényképen:



# A világ legszélesebb körben alkalmazott terápiás eljárása

Tibetből:



# A világ legszélesebb körben alkalmazott terápiás eljárása

Közel-Keletről:



# Az (igen késői) folytatás



Pierre-Charles-Alexandre Louis (1787-1872)

# Louis vizsgálata (1828, 1834, 1835)

Túlélők adatai:

1	2	3	4	5	6	7	8	9									
10	3	7	3	19	3	19	3	28	2	13	1	24	2	19	2	35	1
12	2	10	2	29	3	12	2	17	3	16	2	12	4	12	1	11	2
14	2	12	2	20	2	15	2	40	2	23	3	19	2	18	1	17	2
				20		22	4	13	2	35	5	18	2	20	3	30	3
				16	3	12	4	21	2	17	2	15	2	13	2		
				17	4	21	2	13	2			27	2	21	2		
				25	3												
				28	4												
				40	2												
				16	2												
				12	4												
12	2½	10	2½	20	3	20	8	22	2	21	2½	19	2½	17	2	23	2

The figures upon the horizontal line above the columns indicate the day when the first bleeding was performed ; the figures on the left in each column mark the duration of the disease ; those on the right, the number of bleedings ; and those on the horizontal line below, show the mean duration of the disease and the average number of bleedings.

Morabia A. P. C. A. Louis and the birth of clinical epidemiology. J Clin Epidemiol. 1996 Dec;49(12):1327-33.

# Louis vizsgálata (1828, 1834, 1835)

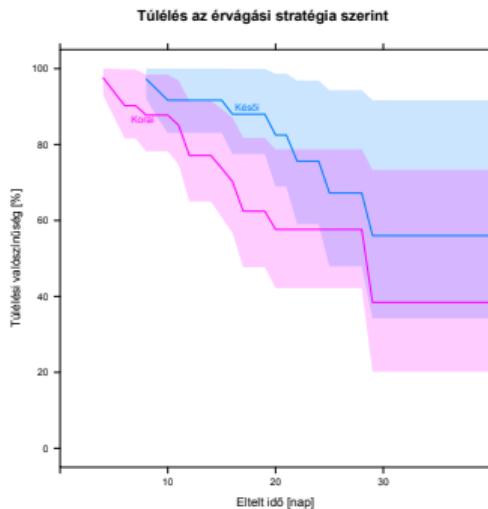
Elhunytak adatai:

1	2	3	4	5	6	7	8	9
6 5 18	53 5 65	4 1 57	29 2 19	16 4 58	62 4 20	20 2 68	25 1 40	22 1 50
12 3 69	16 2 54	29 4 46	8 2 63	10 2 40				
8 2 65	6 3 30	12 1 85	9 4 24	29 3 24				
12 1 55	6 4 47	15 3 37						
17 7 75	47 2 75	17 1 67						
	11 4 45	20 3 22						
6 5 18	20 3 3-5 66	15 3 51	20 2 $\frac{1}{2}$ 49	11 3 48	33 3 28	20 2 68	25 1 40	22 1 50

Morabia A. P. C. A. Louis and the birth of clinical epidemiology. J Clin Epidemiol. 1996 Dec;49(12):1327-33.

# Louis vizsgálata (1828, 1834, 1835)

Kiértékelés:



$$\text{RR}=1,76 \ (95\% \text{ CI: } [0,905–3,41], \ p = 0,0984)$$

Morabia A. P. C. A. Louis and the birth of clinical epidemiology. J Clin Epidemiol. 1996 Dec;49(12):1327-33.

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés
- **Az empirikus orvosi kutatások alapgondolata és a kauzalitás**
- A confounding problémája
- A confounding megoldásai: megfigyelés és kísérlet
- Okozati következtetések levonása megfigyeléses adatokból
- A véletlen szerepéről
- Bizonyitekok hierarchiája és összessége
- A végpont és meghatározása

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

# Az orvosi kutatások egy általános sémaja felé

- Az orvostudomány egy jelentős része – egész története alatt – igen egyszerű alakban megfogalmazható kérdések megválaszolásából, kutatásából állt:
  - A koponyalékelés hatásos-e a migrén ellen?
  - Hat-e ez a gyógyfű a cukorbetegségre?
  - Csökkenti-e ez a vérnyomáscsökkentő gyógyszer-jelölt a vérnyomást, azaz hatásos-e?
  - Okoz-e hasmenést mint mellékhatást ugyanez a gyógyszer-jelölt?
  - Magasfeszültségű távvezeték közelében tartózkodás növeli-e a rákkockázatot?
  - A vöröshús-fogyasztás növeli-e a rákkockázatot?
  - A dohányzás hogyan hat az infarktus utáni túlélésre?
  - A császármetszéssel születés növeli-e a T1DM rizikót?
  - Az anyai acetaminophen-szedés növeli annak valószínűségét, hogy a gyerek autista lesz?
- Hogyan adjunk ezekre választ empirikus adatok alapján?

# Az általános séma

- Van egy *expozíció*nk
  - ...koponyalékelés, gyógyfű, gyógyszerbevitel, távvezeték közelében tartózkodás stb.
- Van egy *végpontunk*
  - ...migrén-súlyosság, cukorbetegség-súlyosság, vérnyomás, rák stb.
- És ami érdekel minket: a kettő közti *okozati* kapcsolat, a kauzalitás!

# Mit nevezünk kauzalitásnak, avagy az ideális vizsgálati módszer

- Vesszük a vizsgálat szempontjából érdekes csoportot és megnézzük, hogy exponálva őket mi történik velük végpont szempontjából...
- ...utána bekapcsoljuk az időgépet, visszamegyünk a kutatás kezdetére, majd – miközben természetesen minden más változatlan – megnézzük, hogy mi történik akkor, ha nincsenek exponálva...
- ...és a kettő *különbsége* az expozíció *hatása!* (Különbsége: a dolog csak összehasonlításban értelmezhető!)
- Sajnos ennek elvégzése sok esetben gyakorlati problémákba ütközik
- Így más módszert kell választanunk: exponált (vagy különbözőképpen exponált) és nem exponált csoportokat hasonlítunk egymáshoz
- De vigyázzunk, ezzel behozunk egy *feltételezést*: hogy a nem exponált csoport mutatja, hogy mi történt *volna*, ha az exponált nem kapott *volna* expozíciót
- (Ez a kauzalitás ún. 'counterfactual' modellje)

# Az empirikus orvosi kutatások alaplogikája

Az expozíció akkor van okozati összefüggésben a végponttal, ha a csak az expozícióban eltérő csoportok eltérnek a végpontban, mégpedig olyan mértékben, ami már nem tudható be a véletlen ingadozásnak.

# Tehát...

Ha...

- a gyógyszert szedők vérnyomása csökken a velük mindenben azonos, de gyógyszert nem szedő csoporthoz képest
- a gyógyszert szedők között többször fordul elő hasmenés a velük mindenben azonos, de gyógyszert nem szedő csoporthoz képest
- a távvezeték közelében felnövők között több lesz a rákos mint a velük mindenben azonos, de nem távvezeték közelében felnövő csoport körében
- a több vörös húst fogyasztók körében több lesz a rákos mint a velük mindenben azonos, de kevesebb vörös húst fogyasztók körében
- a dohányzók nagyobb arányban halnak meg a szívinfarktus után a velük mindenben azonos, de nem dohányzó csoport tagjai
- a császármetszéssel születők között több lesz az 1-es típusú cukorbeteg mint a velük mindenben azonos, de nem császármetszéssel világra jött csoport tagjai között

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés
- Az empirikus orvosi kutatások alapgondolata és a kauzalitás
- **A confounding problémája**
  - A confounding megoldásai: megfigyelés és kísérlet
  - Okozati következtetések levonása megfigyeléses adatokból
  - A véletlen szerepéről
  - Bizonyitekok hierarchiája és összessége
  - A végpont és meghatározása

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

# A confounding problémája

- Naiv gondolat arra, hogy hogyan vizsgáljuk a kérdéseket: több rákos van a távvezeték közelében élők körében? a több vörös húst evők körében? több T1DM-es van a császármetszéssel születők körében?
- Roppant csábító, de... *teljesen fals!*
- A távvezeték körében élő emberek nem *csak* a távvezeték közelében élésben térnek el a nem távvezeték közelében élőktől, a több vörös húst evők nem *csak* a több vörös hús evésében térnek el a kevesebb vörös húst evőktől, a császármetszéssel születők nem *csak* a császármetszéssel születésben térnek el a pvn születőktől stb. stb.

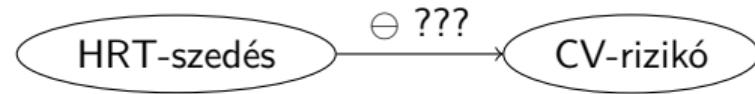
Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health. 2001;22:189-212.

# A confounding problémája

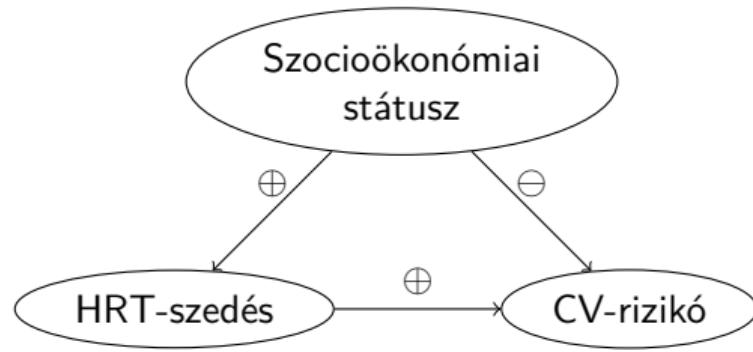
- Innén kezdve, ha találunk is különbséget a végpontban, nem tudhatjuk, hogy az mi miatt van: a vizsgált eltérés (azaz az expozíció) miatt, a vizsgált eltéréssel együtt járó valamely egyéb eltérés miatt, vagy a kettő valamilyen keveréke miatt
- Ha a két csoport nem csak az expozícióban tér el, onnantól biztosan *nem* tudhatjuk, hogy ha találunk is különbséget, annak mi az oka!
- (Az ideális kutatási módszerre visszautalva: az a gond, hogy az összehasonlító csoportra nem igaz, hogy mutatja mi történne az exponált csoportban, ha nem lett volna expozíció...)

Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health. 2001;22:189-212.

# A confounding problémája



# A confounding problémája



## Nézzünk egy számszerű példát is erre

CV események előfordulása:

	Nem szed HRT-t	Szed HRT-t
Alacsony szocioökonómiai státusz	4% (240/6000)	6% (12/200)
Magas szocioökonómiai státusz	1% (50/5000)	2% (60/3000)
Összességében	2,6% (290/11000)	2,3% (72/3200)

# A confounder definíciója tehát

Változó, amire egyszerre igaz, hogy:

- ① összefügg az expozícióval (tehát eltérnek e szerint az exponált és nem exponált csoportok), és
- ② önmagában is hat a végpontra

A HRT-s példában:

- ① A szocioökonómiai státusz összefügg a gyógyszerszedéssel (mert alacsonyban csak  $200/6200 = 3,23\%$  szed HRT-t, a magasban viszont  $3000/8000 = 37,5\%$ )
- ② A szocioökonómiai státusz önmagában is hat a végpontra (1%-ról 4%-ra, illetve 2%-ról 6%-ra emeli a kockázatot)

# Nincs confounding, ha a változó nem függ össze az expozícióval

	Nem szed HRT-t	Szed HRT-t
Alacsony szocioökonómiai státusz	4% (186/4650)	6% (93/1550)
Magas szocioökonómiai státusz	1% (60/6000)	2% (40/2000)
Összességében	2,3% (246/10650)	3,7% (133/3550)

(Mindkét csoportban  $1550/6200 = 2000/8000 = 25\%$  a szedők aránya)

# Nincs confounding, ha a változó nem hat a végpontra

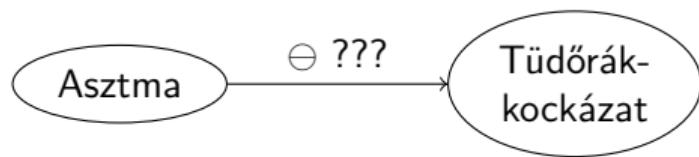
	Nem szed HRT-t	Szed HRT-t
Alacsony szocioökonómiai státusz	3% (180/6000)	5% (10/200)
Magas szocioökonómiai státusz	3% (150/5000)	5% (150/3000)
Összességében	3% (330/11000)	5% (160/3200)

## Két megjegyzés

- Megoldja a confounding problémáját, ha rájövünk, hogy mi a confounder, és a szerint „lebontva” (rétegezve) nézzük az eredményeket – erre a gondolatra később még visszatérünk
- Nem kötelező, hogy csak két (exponált és nem exponált) csoport legyen, sőt, általában a kauzalitási következtetés robusztusabb is, ha nem egyszerűen azt tudjuk kimutatni, hogy az exponáltaknál nagyobb a kockázat, hanem azt is, hogy az egyre jobban (hosszabban, tovább stb.) exponáltaknál egyre nagyobb a kockázat

# A confounding problémája

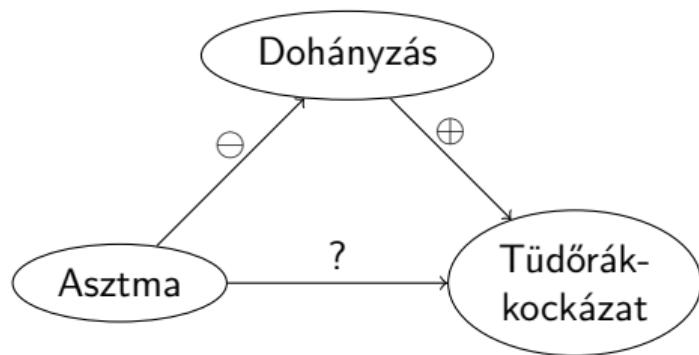
Amikor azt mondunk a confounding első feltételeként, hogy a változó összefügg az expozícióval, akkor mindenki, hogy mi hat mire és hogyan (egyik a másikra, másik az egyikre, harmadik mindkettőre stb.), csak az a tény számít, hogy korreláltak



Pontosan emiatt a továbbiakban erre az első feltételre duplanyilas jelzést fogunk használni

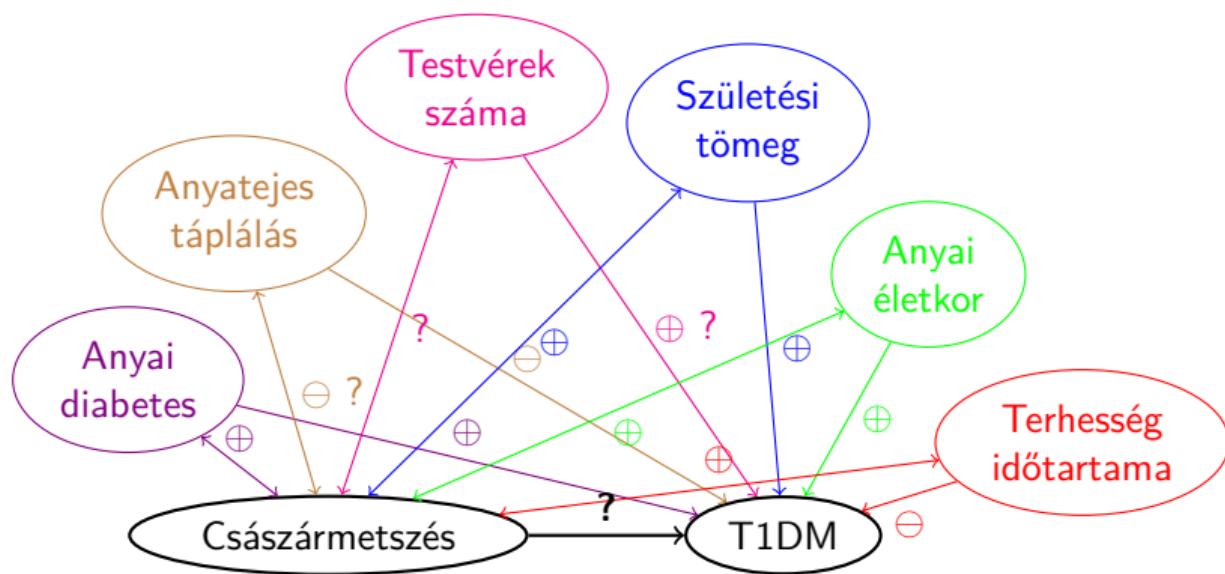
# A confounding problémája

Amikor azt mondunk a confounding első feltételeként, hogy a változó összefügg az expozícióval, akkor mindenki, hogy mi hat mire és hogyan (egyik a másikra, másik az egyikre, harmadik mindketten stb.), csak az a tény számít, hogy korreláltak



Pontosan emiatt a továbbiakban erre az első feltételre duplanyilas jelzést fogunk használni

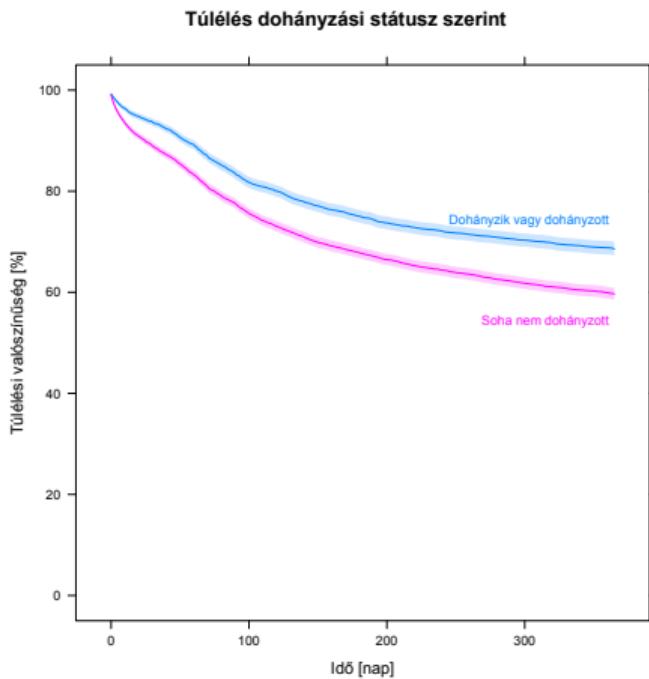
# A confounding problémája



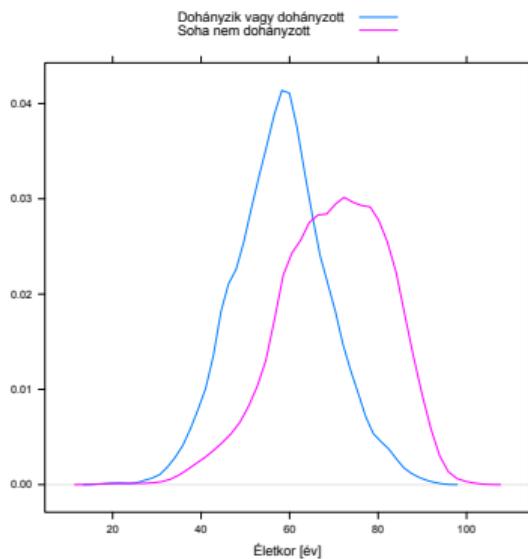
Cardwell CR, Stene LC, Joner G, et al. Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. Diabetologia.

2008 May;51(5):726-35.

# Egy klinikaibb példa a confoundingra: infarktus utáni halálozás és a dohányzás



# Egy klinikaibb példa a confoundingra: infarktus utáni halálozás és a dohányzás



Aune E, Røislien J, Mathisen M, Thelle DS, Otterstad JE. The "smoker's paradox" in patients with acute coronary syndrome: a systematic review. BMC Med. 2011 Aug 23;9:97.

# A Simpson-paradoxon mint confounding

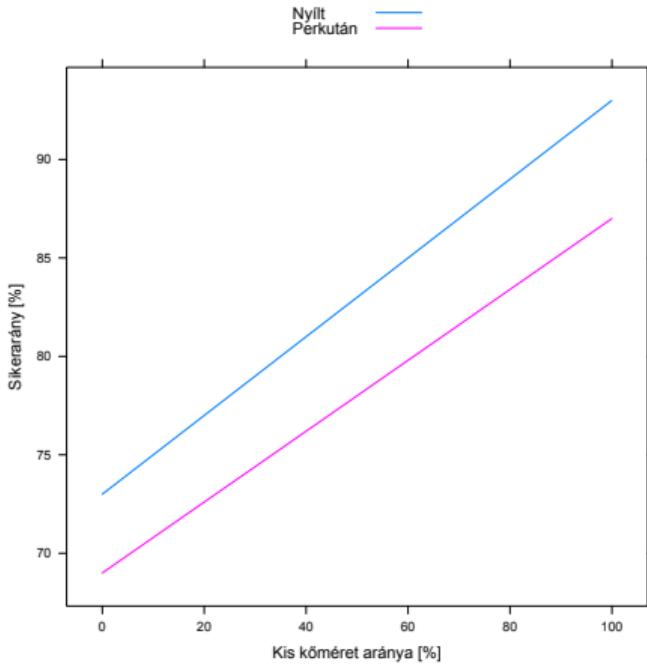
	Nyílt feltárás	Perkután eljárás
Összességében	78% (273/350)	83% (289/350)

# A Simpson-paradoxon mint confounding

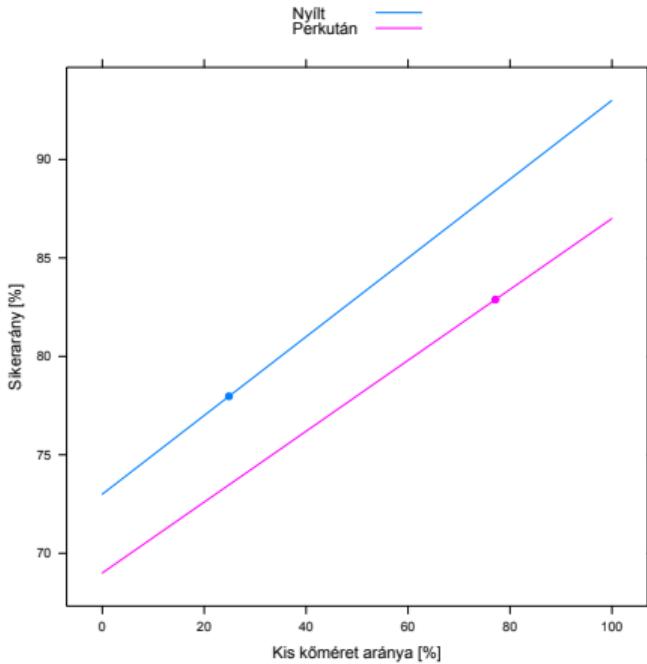
	Nyílt feltárás	Perkután eljárás
Kőátmérő < 2cm	93% (81/87)	87% (234/270)
Kőátmérő $\geq$ 2cm	73% (192/263)	69% (55/80)
Összességében	78% (273/350)	83% (289/350)

Julious SA, Mullee MA. Confounding and Simpson's paradox. BMJ. 1994 Dec 3;309(6967):1480-1. Hernán MA, Clayton D, Keiding N. Int J Epidemiol. 2011 Jun;40(3):780-5. Baker SG, Kramer BS. Good for women, good for men, bad for people: Simpson's paradox and the importance of sex-specific analysis in observational studies. J Womens Health Gend Based Med. 2001 Nov;10(9):867-72. Wainer H. The BK-Plot: Making Simpson's Paradox Clear to the Masses. Chance. 2002;15(3):60-62. Baker SG, Kramer BS. The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? BMC Med Res Methodol. 2002 Nov 13;2:13.

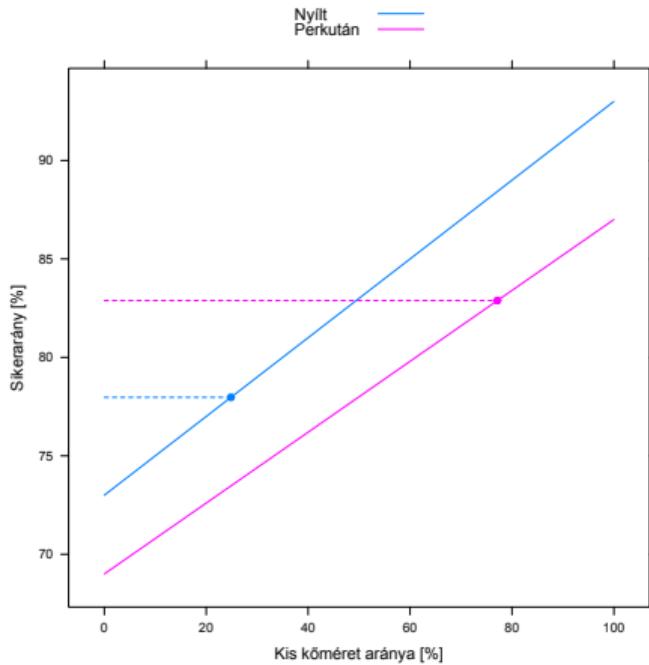
# A Simpson-paradoxon grafikus magyarázata: BK-ábra



# A Simpson-paradoxon grafikus magyarázata: BK-ábra



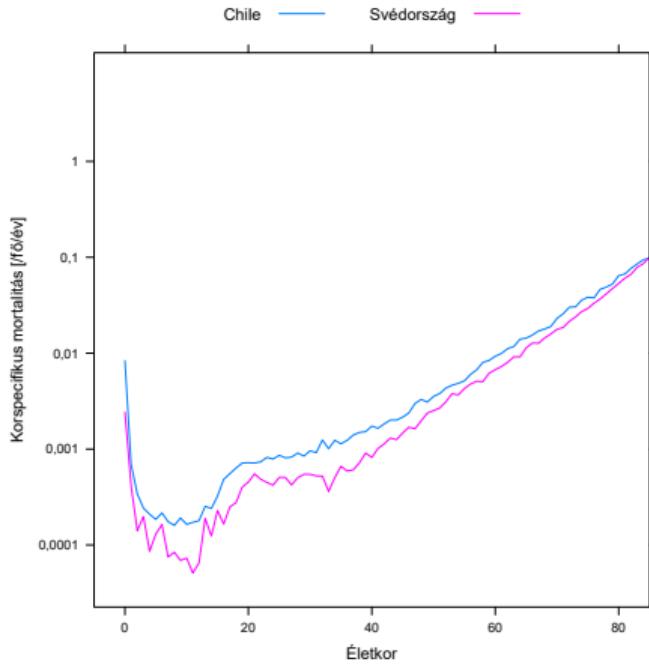
# A Simpson-paradoxon grafikus magyarázata: BK-ábra



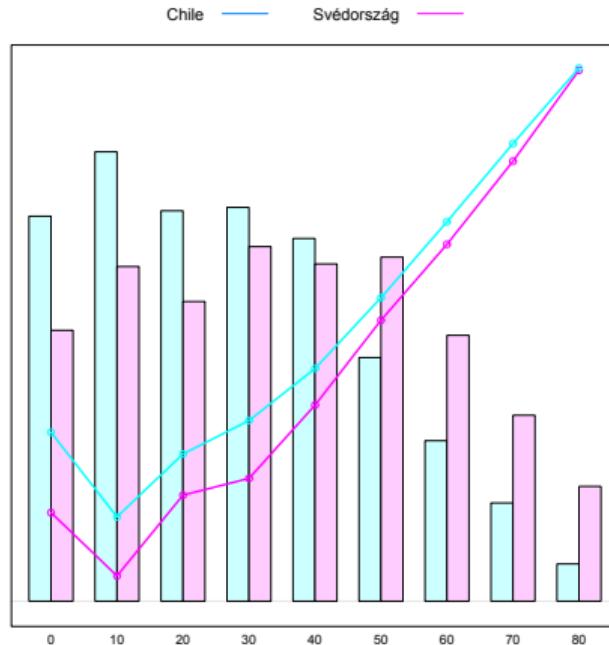
## Egy nagyon más(nak tűnő) példa, ami mégis confounding

- Svédországban 2005-ben 91 ezer 710 halálozás történt, a lakosságszám 9 millió 10 ezer 729, így a nyers halálozási ráta 10,2/ezer fő/év
- Chilében ugyanabban az évben 86 ezer 100 halálozás történt, a lakosságszám 15 millió 519 ezer 347, így a nyers halálozási ráta 5,5/ezer fő/éve
- Svédországban kétszer (???) nagyobb a halandóság?
- ???

# Egy nagyon más(nak tűnő) példa, ami mégis confounding



# Egy nagyon más(nak tűnő) példa, ami mégis confounding



## Négy megjegyzés

- ① Ez nem csak az orvosi szakkutatások problémája – érdemes nézni a laikus sajtót is az ilyen példák miatt... (sajnos sokszor nagyon csábító, és még jobban belegondolva nagyon egyértelmű helyzetekben is el lehet rontani, vagy – rosszabb esetben – félre lehet ezzel vezetni embereket)
- ② Az, hogy felismerjük a confounding-ot egy helyzetben, azt jelenti, hogy igazoltuk, hogy a korreláció nem bizonyítja (pláne nem erősen) az összefüggést – de ettől még lehet összefüggés! A bizonyíték cáfolata nem jelent ellenbizonyítékot!
- ③ Egy gyakori vélemény: „persze, értem, hogy fontosak ezek a buborékok, meg, hogy a nyilak merről-merre mennek, de én egy mezei klinikus vagyok, engem nem érdekelnek ezek a módszertani izék, a lényeg, hogy azok körében, akik X-et kapnak/szednek, kevesebb a baj, nyilván javasolni fogom” – teljesen fals! Az, hogy a nyilak „merről-merre” mennek, nem módszertani részletkérdés, hanem elsőrendű a klinikai döntéshozatalhoz! Miért? Ezen múlik, hogy mi történik, ha beavatkozunk a rendszerbe! Ennek figyelmen kívül hagyásával elképzelhető, hogy a beavatkozás nem éri el a kívánt hatást – vagy rosszabb esetben egyenesen az ellenkezőjét éri el
- ④ Egy szóhasználati kérdés: azt minden további nélkül is mondhatjuk, hogy „együtt jár” (korreláció), csak kauzális szót nem használhatunk (kauzális szó: okoz, megnövel, lecsökkent, gyakoribbá tesz stb.)

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés
- Az empirikus orvosi kutatások alapgondolata és a kauzalitás
- A confounding problémája
- **A confounding megoldásai: megfigyelés és kísérlet**
- Okozati következtetések levonása megfigyeléses adatokból
- A véletlen szerepéről
- Bizonyitekok hierarchiája és összessége
- A végpont és meghatározása

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

# A megoldási lehetőség

- A feladat tehát: elérni, hogy a csoportok csak a vizsgált expozícióban térjenek el! Hogyan biztosítható ez?
- Törekedni sokféleképp lehet rá...
  - Ha csak azonos életkorú alanyokat vonunk be, akkor ez nem lehet confounder (azzal nem tudunk mit kezdeni, ha hat a végpontra, de azt megakadályozhatjuk, hogy összefüggjön az expozícióval)
  - Ha a tervezési fázisban ezzel nem is törődtünk, akkor az elemzésben még mindig javíthatunk (pl. kiveszünk csak azonos életkorúakat – ennél majd eljárhatunk ügyesebben is)
- De ez mind csak azon confounderek ellen nyújt védelmet, amiről van információnk: tudunk információ szerezni, és egyáltalán, eszünkbe jutott, hogy confounder
- Hogyhogy, lehet védekezni olyan confounder ellen is, amiről eszünkbe sem jut, hogy confounder?!
- Igen!

# Az első példa



James Burns Amberson (1890-1979)

## Az első példa

Obviously, the matching could not be precise, but it was as close as possible, each patient having previously been studied independently by two of us. Then, by a flip of the coin, one group became identified as group I (sanocrysin-treated) and the other as group II (control). The members of the separate groups were known only to the nurse in charge of the ward and to two of us. The patients themselves were not aware of any distinction in the treatment administered.

Amberson JB, McMahon BT, Pinner M (1931). A clinical trial of sanocrysin in pulmonary tuberculosis. American Review of Tuberculosis 24:401-435.

# A megoldási lehetőség

- Az egyetlen univerzális megoldás: ha az alanyaink *véletlenszerűen* választott része kapja meg az expozíciót
- Ezt hajtva végre ugyanis *semmilyen* szempont szerint nem lesz szisztematikus eltérés a két csoport között a vizsgált expozición kívül
- Semmilyen, azaz azok a szempontok is ki lesznek egyensúlyozva, amikről nincs információnk (sőt, akár eszünkbe sem jutott, hogy confounder!)
- (Szisztematikus: a véletlen besorolás ingadozása miatt persze lesz eltérés, de ezt majd a későbbi statisztikai eszköztár kezeli)
- Ez a *randomizálás* gondolata (R. A. Fisher)
- Figyelem: ehhez szükséges, hogy mi *irányítsuk* az expozíciót, mi határozzuk meg, hogy ki kap ilyet
- Ezt a kutatási módszert, amelyben a kutatók *aktívan befolyásolják*, hogy ki kap expozíciót *kísérletes (experimentális) vizsgálatnak* nevezzük
- Így (és csak így!) biztosítható, hogy az összehasonlított csoportok *tényleg* csak az expozíció szerint térjenek el...
- ... emiatt a végponton talált esetleges különbség *tényleg* az expozíció tényének (és esetleg a véletlen ingadozásnak) tudható be

# A megfigyelés

- Amennyiben a kutatást végzők nem befolyásolják az expozíciót, csak *passzíve* feljegyzik, hogy ki részesült benne (ahogy amúgy is történt volna), akkor *megfigyeléses (obszervációs) vizsgálatról* beszélünk
- Megfigyeléses vizsgálatnál *mindig* ott lebeg Damoklész kardjaként a fejünk felett a confounding: vajon *tényleg* csak az expozíció tényében térnek el az expozíció szerint képezett csoportok...?
- Mert ha nem, és véletlenül valami hat a végpontra is...
- Tenni lehetünk a confounding ellen, de tökéletesen nyugodtak soha sem lehetünk
- Ugyanis ahhoz, hogy tegyük valamit, a minimum, hogy (a) eszünkbe jusson, hogy mi a confounder és (b) le is tudjuk mérni azt
- Az ilyeneket majd ki tudjuk statisztikai úton is szűrni
- A kísérlet előnye lényegében az, hogy *automatikusan* védelmet jelent *minden* confounder ellen: azok ellen is, amikre nem is gondolunk, vagy nem tudjuk (jól) lemérni
- Ez *csak* kísérlettel érhető el!

# Megfigyelés és kísérlet

- Igazából ezért szeretjük a kísérleteket!
- Így – és csak így! – biztosítható, hogy a csoportok között ne legyen más szisztematikus eltérés mint az expozíció
- Ezért ha találunk különbséget, az tényleg az expozíciónak (és a véletlen ingadozásnak) tudható be
- Csakhogy a kísérletek néha kivihetetlenek (T1DM vs. császármetszés, távvezeték vs. rákkockázat, vagy ejtőernyő vs. gravitáció okozta trauma), de még ha végezhetők is, akkor is
  - sokszor drágák, lassúak, szervezésigényesek stb., ebből adódóan egyrészt korlátozott a mintanagyság (gyakorlatban már a néhány ezer fő is sok, tíz ezer fő fölé igen ritka esetben mennek) és a követés időbeli hossza (jellemzően legfeljebb néhány év) – ebből adódóan a ritka, illetve lassan kialakuló végpontok nehezen vizsgálhatóak (intuitíve érezhető, a pontos magyarázatot lásd kicsit később)
  - megszervezésük kötöttségeiből adódóan még a legjóhiszeműbb szervezés esetén is egy „steril” csoportot jelent a kísérletbe bekerülők halmaza, kérdés, hogy az így nyert ismeretek mennyire általánosíthatóak az összes alanyra – ez azonban csak egy félig jogos kritika, erre még később visszatérünk
- Néha ezek miatt még akkor is végeznek megfigyelést, ha lehetne kísérlet csinálni, sőt, akár történt is

# Egy gyógyszeres példa erre

Characteristics	NOAC			
	Apixaban	Dabigatran	Rivaroxaban	Warfarin
No in group				35 436
Women	39.7 (2522)	33.9 (4304)	43.1 (3100)	41.2 (14 598)
Median (interquartile range) age (years)	71.3 (65.8-77.2)	67.6 (62.0-72.4)	71.8 (65.7-78.9)	72.4 (64.7-79.8)
Age >65	78.2 (4967)	64.4 (8180)	77.7 (5590)	74.2 (26 295)
Age >75	33.7 (2140)	13.9 (1766)	38.1 (2737)	41.4 (14 655)
Previous atrial fibrillation diagnose	68.9 (4374)	70.0 (8889)	60.2 (4333)	51.5 (18 243)
Mean (SD) CHA <sub>2</sub> DS <sub>2</sub> VASc score†	2.8 (1.6)	2.2 (1.4)	2.8 (1.6)	2.8 (1.7)
Mean (SD) HAS-BLED score‡	2.3 (1.2)	2.0 (1.1)	2.2 (1.2)	2.2 (1.2)

Larsen TB, Skjøth F, Nielsen PB, Kjældgaard JN, Lip GY. Comparative effectiveness and safety of non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study. BMJ. 2016 Jun 16;353:i3189.

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés
- Az empirikus orvosi kutatások alapgondolata és a kauzalitás
- A confounding problémája
- A confounding megoldásai: megfigyelés és kísérlet
- **Okozati következtetések levonása megfigyeléses adatokból**
- A véletlen szerepéről
- Bizonyitekok hierarchiája és összessége
- A végpont és meghatározása

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

# Kauzalitás vizsgálata megfigyeléses adatokból

- Annyit mondunk, hogy „tenni tehetünk a confounding ellen” meg, hogy „statisztikai úton szűrhető” (hiába megfigyelésesek az adataink); de mit jelent ez?
- 5 fő módszer (számos variánssal):
  - Rétegzés
  - Standardizáció
  - Illesztés (matching)
  - Többváltozós – regressziós – modellezés
  - Propensity score eljárások
- De *bármelyiket* is használjuk, azt nem lehet megkerülni, hogy csak azokat tudjuk szűrni, amikről eszünkbe jutott, hogy confounderek, és le is tudtuk mérni őket! (Ez volt a kísérletek nagy előnye!)
- Plusz, a módszertől függően különböző további feltevések jöhetnek be, tehát ezek sem univerzálisak
- Fontos, hogy ezek a confounding analitikai fázisban történő korrigálásnak eszközei – tehetünk tervezési fázisban is lépéseket
- A manapság legfontosabb eljárás a többváltozós modellezés, azzal külön is fogunk foglalkozni; a cél most inkább az, hogy kiderüljön: egyáltalán megoldható ez a probléma

# A rétegzés alapötlete

- Tulajdonképpen már láttuk:

	Nem szed HRT-t	Szed HRT-t
Összességében	2,6% (290/11000)	2,3% (72/3200)

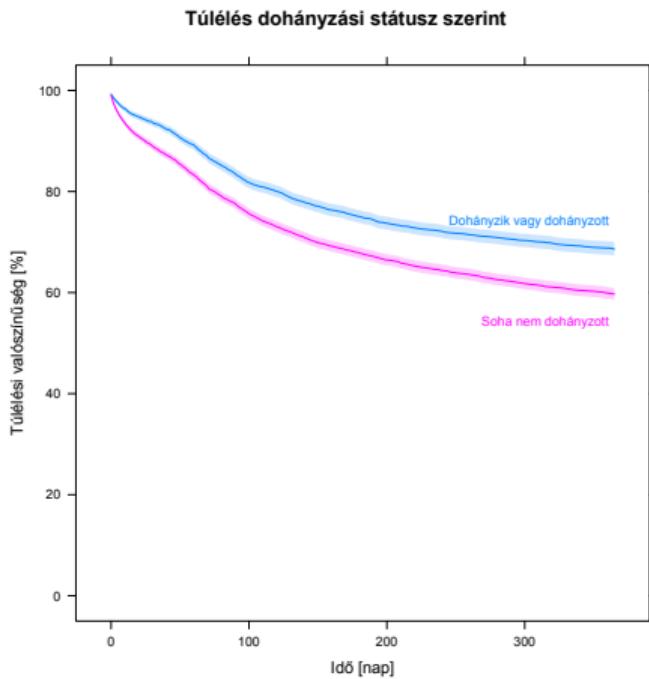
- Rétegzésnek hívják a statisztikusok, amikor az adatokat valamilyen szempont szerint hézag- és átfedélesmentes csoportokra bontjuk, és mindegyikben külön-külön végrehajtjuk az elemzést:

	Nem szed HRT-t	Szed HRT-t
Alacsony szocioökonómiai státusz	4% (240/6000)	6% (12/200)
Magas szocioökonómiai státusz	1% (50/5000)	2% (60/3000)
Összességében	2,6% (290/11000)	2,3% (72/3200)

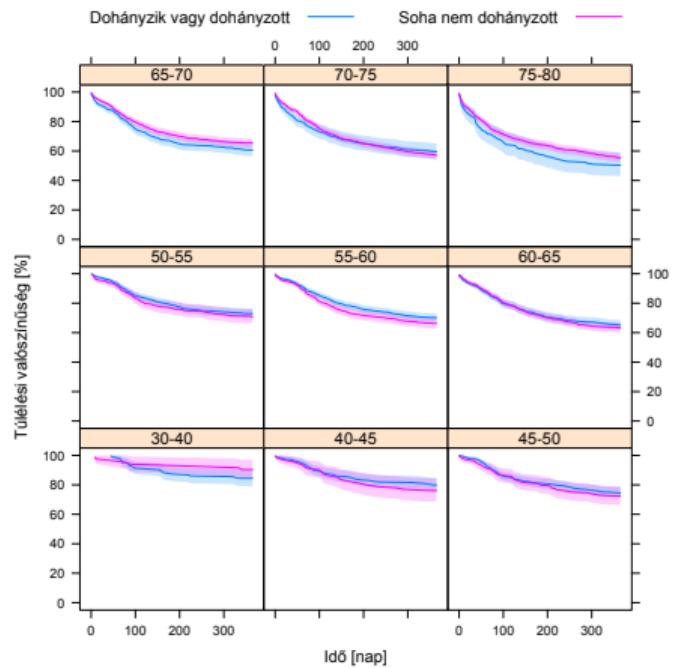
# A rétegzés előnyei és hátrányai

- Előny: semmilyen feltételezéssel nem kell élni, univerzális választ ad
- Hátrány: nem egy választ ad (hanem annyit, ahány réteg van)
- Ez részint nehezen áttekinthető, részint bizonytalan (minél több réteg van, annál kevesebb alany jut egybe)
- A gyakorlatban simán lehet 10-20, vagy akár annál is több confounderünk (ne felejtsük el, hogy ezek igazából potenciális confounderek, a valóságban mi sem tudhatjuk, hogy mik az igaziak, ezért az összes potenciálisat be kell rakkunk)
- Ezek egyáltalán nem biztos, hogy kétkimenetűek, de a legnagyobb gond, hogy ha teljesen általánosak akarunk lenni, akkor ezek kombinatorikusak
- Külön problémát jelentenek a folytonos változók: szét kell vágni, de azt meg nem lehet jól csinálni (ha túl széles a sáv, akkor különböző dolgokat mosunk egybe, ha túl szűk, akkor kevés alany lesz egy rétegben)

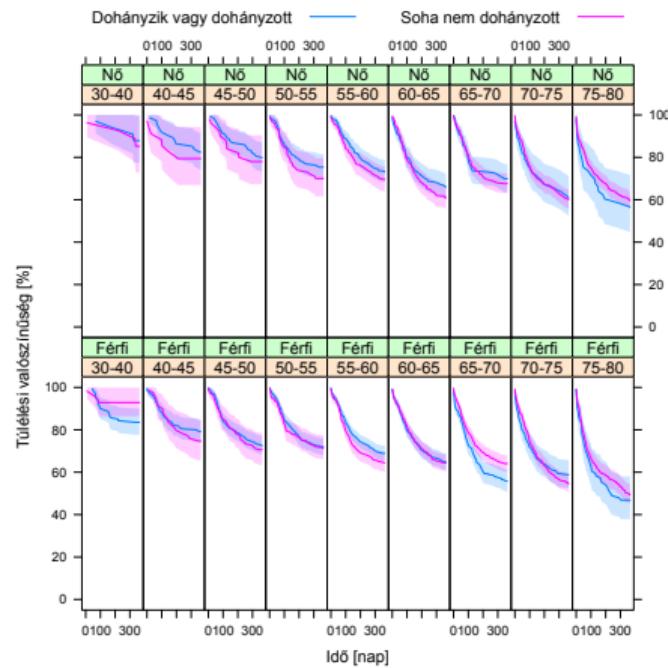
# A rétegzés illusztrálása



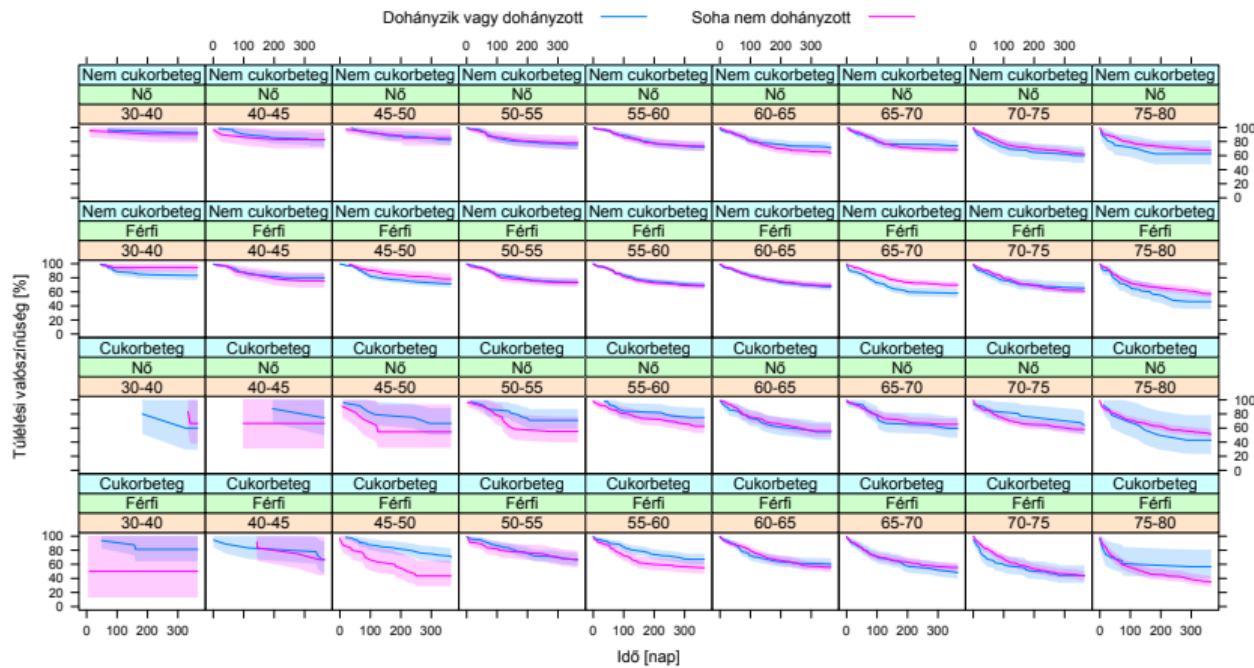
# A rétegzés illusztrálása



# A rétegzés illusztrálása



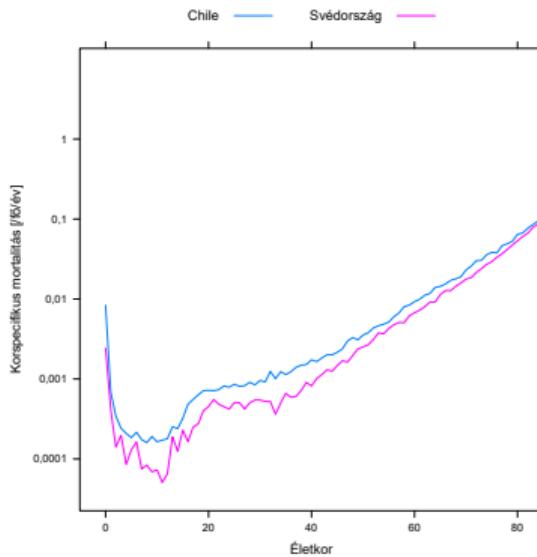
# A rétegzés illusztrálása



# A standardizálás illusztrálása egy problémán

- Vegyük elő a svéd-chilei példát!
- Emlékeztetőül:
- Svédországban 2005-ben 91 ezer 710 halálozás történt, a lakosságszám 9 millió 10 ezer 729, így a nyers halálozási ráta (CDR)  $10,2/ \text{ezer fő}/\text{év}$
- Chilében ugyanabban az évben 86 ezer 100 halálozás történt, a lakosságszám 15 millió 519 ezer 347, így a nyers halálozási ráta  $5,5/ \text{ezer fő}/\text{éve}$
- Svédországban kétszer (???) nagyobb a halandóság?

# A rétegzés illusztrálása



(Igen, igazából ez egy rétegzés volt! Több eredmény van, sőt, nagyon sok, csak ügyes ábrázolással ezt mégis jól áttekinthetővé tettük!)

# A számítás menete (Svédország példáján)

Korcsoport	Halálozások száma	Létszám	Létszám megoszlás	Korspecifikus mortalitás
0	378	965477	0,1071	0,0004
10	215	1192801	0,1324	0,0002
20	518	1068031	0,1185	0,0005
⋮	⋮	⋮	⋮	⋮
100	637	1228	0,0001	0,5188
	91710	9010729	1	

$$\text{CDR}_{\text{Svédország}} = \frac{91710}{9010729} = \frac{378 + 215 + 518 + \dots + 637}{9010729} = \\ = \frac{965477 \cdot 0,0004 + 1192801 \cdot 0,0002 + 1068031 \cdot 0,0005 + \dots + 1228 \cdot 0,5188}{9010729} = \dots$$

## A számítás tehát (Svédország példáján)

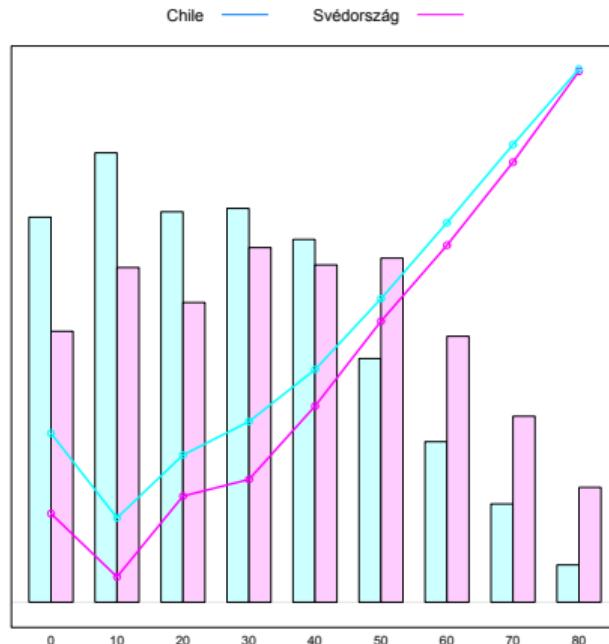
$$\begin{aligned}
 \text{CDR}_{\text{Svédország}} &= \dots = \frac{965477 \cdot 0,0004 + 1192801 \cdot 0,0002 + 1068031 \cdot 0,0005 + \dots + 1228 \cdot 0,5180}{9010729} = \\
 &= \frac{965477 \cdot 0,0004}{9010729} + \frac{1192801 \cdot 0,0002}{9010729} + \frac{1068031 \cdot 0,0005}{9010729} + \dots + \frac{1228 \cdot 0,5180}{9010729} = \\
 &= \frac{965477}{9010729} \cdot 0,0004 + \frac{1192801}{9010729} \cdot 0,0002 + \frac{1068031}{9010729} \cdot 0,0005 + \dots + \frac{1228}{9010729} \cdot 0,5180 = \\
 &= 0,1071 \cdot 0,0004 + 0,1324 \cdot 0,0002 + 0,1185 \cdot 0,0005 + \dots + 0,0001 \cdot 0,5180 = 0,010178
 \end{aligned}$$

Azaz azt is mondhattuk volna, hogy a korspecifikus mortalitásokat szorozzuk a korosztály létszámarányával, és ezeket összeadjuk – tehát lényegében egy súlyozott átlag: a korspecifikus mortalitások súlyozva a korcsoportok létszám szerinti megoszlásával, tehát a korfával

## Emlékeztetőül a magyarázat

- A kutya ott van elásva, hogy nagyon mások a „súlyozó függvények”, tehát a korfák!
- Igaz ugyan, hogy minden életkorban rosszabbak a chilei adatok, csak épp közülük azok esnek nagy súllyal latba, amik jobbak, míg a svédeknél azok, amik rosszabbak (mert azért a chilei 20 éves mortalitás még mindig jobb, mint a svéd 70 éves – ezen múlik a dolog)

# A magyarázat szemléltetve



# A standardizálás alapötlete

- Először is vegyük észre, hogy amikor korspecifikus rátákat használtunk, akkor lényegében egy confounding-ot oldottunk meg rétegzéssel
- Most egy módszert fogunk látni, mely a rétegzésnek legalább azt a problémáját megoldja, hogy nagyon sok eredményt kapunk
- (Persze a dolognak ára lesz!)
- A gond tehát az, hogy a korspecifikus mortalitásokat *különböző* korfákkal súlyozzuk
- Elég kézenfekvő ötlet: súlyozzuk *mindkét* országot *ugyanazzal* a korfával!

# A standardizálás alapötlete

Korcsoport	Korspecifikus mortalitás		Létszám megoszlás (korfa)		
	Svédország	Chile	Svédország	Chile	WHO standard
0	0,0004	0,0010	0,1071	0,1523	0,1754
10	0,0002	0,0004	0,1324	0,1778	0,1706
20	0,0005	0,0008	0,1185	0,1544	0,1614
30	0,0006	0,0012	0,1403	0,1558	0,1475
40	0,0015	0,0023	0,1334	0,1435	0,1263
50	0,0041	0,0054	0,1361	0,0964	0,0992
60	0,0105	0,0138	0,1052	0,0635	0,0668
70	0,0290	0,0360	0,0736	0,0389	0,0373
80	0,0885	0,0909	0,0454	0,0147	0,0135
90	0,2386	0,1849	0,0079	0,0028	0,0019
100	0,5188	0,4440	0,0001	0,0001	0,0000

A svéd számolás akkor, ha chilei korfát használunk:

$$0,1523 \cdot 0,0004 + 0,1778 \cdot 0,0002 + \dots + 0,0001 \cdot 0,5180 = 0,04659$$

Így már tényleg kisebb, mint a chilei érték!

# A standardizálás alapötlete

Korcsoport	Korspecifikus mortalitás		Létszám megoszlás (korfa)		
	Svédország	Chile	Svédország	Chile	WHO standard
0	0,0004	0,0010	0,1071	0,1523	0,1754
10	0,0002	0,0004	0,1324	0,1778	0,1706
20	0,0005	0,0008	0,1185	0,1544	0,1614
30	0,0006	0,0012	0,1403	0,1558	0,1475
40	0,0015	0,0023	0,1334	0,1435	0,1263
50	0,0041	0,0054	0,1361	0,0964	0,0992
60	0,0105	0,0138	0,1052	0,0635	0,0668
70	0,0290	0,0360	0,0736	0,0389	0,0373
80	0,0885	0,0909	0,0454	0,0147	0,0135
90	0,2386	0,1849	0,0079	0,0028	0,0019
100	0,5188	0,4440	0,0001	0,0001	0,0000

A chilei számolás akkor, ha svéd korfát használunk:

$$0,1071 \cdot 0,0010 + 0,1324 \cdot 0,0004 + \dots + 0,0001 \cdot 0,4440 = 0,01123$$

Így már tényleg nagyobb, mint a svéd érték!

# A standardizálás alapötlete

- Hogy most az a svéd korfa, a chilei korfa, magyar korfa, vagy valami más, az első körben nem fontos kérdés, a lényeg, hogy *ugyanaz* legyen a korfa
- A kapott eredmény egy fiktív halálozási ráta lesz...
- ...viszont összehető a két populáció között!
- Értelme: mi *lenne* egy országban a halálozás, ha a korösszetétele olyan *lenne* mint a megadott korfa (de közben a korspecifikus mortalitások ugyanazok maradnának)
- Ha mindkettőnél ugyanazt adjuk meg, akkor értelemszerű, hogy eltüntettük az eltérő korösszetétel hatását
- Lényegében két lépést végzünk: rétegzünk, majd újra összerakunk – ahogy a rossz számítás is elmondható lenne, csak épp most az utóbbi lépésben kiküszöböljük azt, ami a rossz számításnál a problémát okozta
- A számérték függ a választott közös korfától, és akár a sorrendet is befolyásolhatja (de: ha az egyik korspecifikus görbe végig a másik felett húzódik, akkor lehetetlen, hogy megforduljon a sorrend, bármilyen korfát is választunk)

## Referencia (vagy standard) populáció

- A fenti módszerrel mindenkorban csak két országot tudunk összehasonlítani
- Gyakorlati okokból célszerű, ha lehetőleg mindenki ugyanahhoz a korfához standardizál, hiszen így az eredmények egymással is összehasonlíthatók lesznek
- Nem csak adott párokat vizsgálhatunk, hanem egyszerűen minden ország közli a sajátját, de az egyezményes korfához standardizálva, így elég minden országnak egyetlen számot megadni, és mégis bármelyik ország bármelyikkel összehasonlíthatóvé válik
- Ráadásul így kevésbé lehet játszani azzal, hogy olyan korfát választunk, amivel az jön ki, amit látni szeretnénk
- Ezért van néhány, nemzetközileg elfogadott ún. referencia, vagy standard populáció, melyeket általában használnak: pl. US Standard, Segi, ESP, WHO
- Ha standardizált eredményt közlünk, akkor odaírva, hogy mihez standardizáltuk, azonnal összehasonlítható lesz az eredmény

# A standardizálás menete

Korcsoport	Korspecifikus mortalitás		Létszám megoszlás (korfa)		
	Svédország	Chile	Svédország	Chile	WHO standard
0	0,0004	0,0010	0,1071	0,1523	0,1754
10	0,0002	0,0004	0,1324	0,1778	0,1706
20	0,0005	0,0008	0,1185	0,1544	0,1614
30	0,0006	0,0012	0,1403	0,1558	0,1475
40	0,0015	0,0023	0,1334	0,1435	0,1263
50	0,0041	0,0054	0,1361	0,0964	0,0992
60	0,0105	0,0138	0,1052	0,0635	0,0668
70	0,0290	0,0360	0,0736	0,0389	0,0373
80	0,0885	0,0909	0,0454	0,0147	0,0135
90	0,2386	0,1849	0,0079	0,0028	0,0019
100	0,5188	0,4440	0,0001	0,0001	0,0000

$$\text{CDR}_{\text{Svédország}} = 0,1071 \cdot 0,0004 + 0,1324 \cdot 0,0002 + \dots + 0,0001 \cdot 0,5188 = 0,010178$$

$$\text{CDR}_{\text{Chile}} = 0,1523 \cdot 0,0010 + 0,1778 \cdot 0,0004 + \dots + 0,0001 \cdot 0,4440 = 0,005548$$

$$\text{DSR}_{\text{Svédország}} = 0,1754 \cdot 0,0004 + 0,1706 \cdot 0,0002 + \dots + 0,0000 \cdot 0,5188 = 0,004316$$

$$\text{DSR}_{\text{Chile}} = 0,1754 \cdot 0,0010 + 0,1706 \cdot 0,0004 + \dots + 0,0000 \cdot 0,4440 = 0,005251$$

# Amit elértünk (és amit nem)

- Egyetlen számban kaptunk confounding-ra szűrt eredményt
- (A fenti példa esetében a számértéknek önmagában nincs semmilyen értelme, csak összehasonlításban értelmezhető)
- De az esetlegesen eltérő rétegspecifikus hatásokat elfedi (és ez esetben a standard-választás is számíthat) – bár ilyenkor semmilyen egyetlen számba sűrítő index nem lesz az igazi
- További probléma, hogy bár a végeredmény csak egyetlen szám, valójában megjelenik a sok réteg problémája:
  - Ha több millió réteg van, akkor a standardnak annyi számot kell tartalmaznia
  - (Széleskörű megállapodás pedig csak néhány tényezőre van: nem, életkor)
  - Ráadásul igazából a kis létszámú rétegek problémája is megmarad: a kapott standardizált számérték nagyon bizonytalan lesz

# Alkalmazási területek

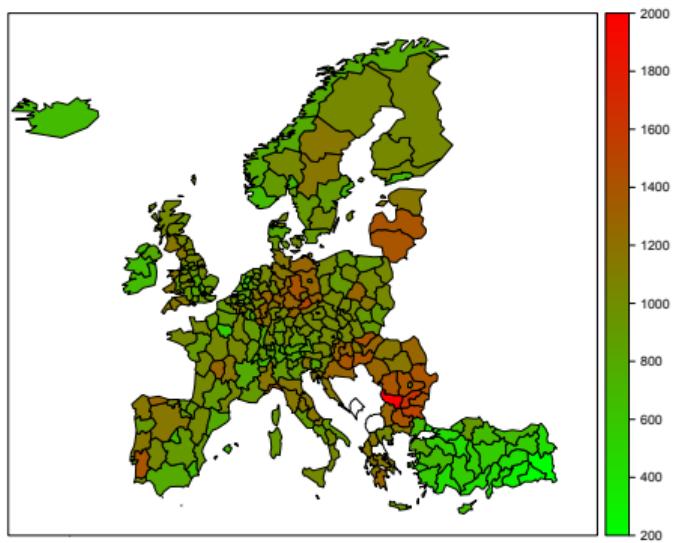
- Elvileg nem csak ilyen problémákra alkalmazható

	Nem szed HRT-t	Szed HRT-t
Alacsony szocioökonómiai státusz	4% (240/6000)	6% (12/200)
Magas szocioökonómiai státusz	1% (50/5000)	2% (60/3000)
Összességében	2,6% (290/11000)	2,3% (72/3200)

- Megtehetnénk például, hogy definiálunk egy „referencia szocioökonómiai státusz-fát” (pl. 55% magas, 45% alacsony) és ezzel súlyozva sűrítjük vissza egy számba a CV eseményeket
- Ekkor a szedők körében  $6\% \cdot 0,45 + 2\% \cdot 0,55 = 3,8\%$  az arány, a nem szedők körében  $4\% \cdot 0,45 + 1\% \cdot 0,55 = 2,35\%$
- Ez a fenti referenciára standardizált infarktus-rizikó (megoldódott a confounding!)
- Nagyon ritka, az alkalmazása lényegében a morbiditási/mortalitási adatokra korlátozott (erős történeti hagyományai vannak)

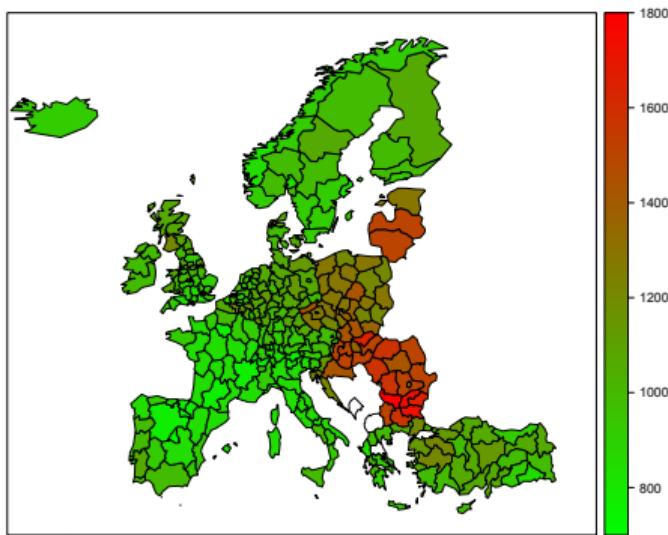
# A standardizálás szemléltetése: térbeli különbségek

Nyers halálozási ráták:



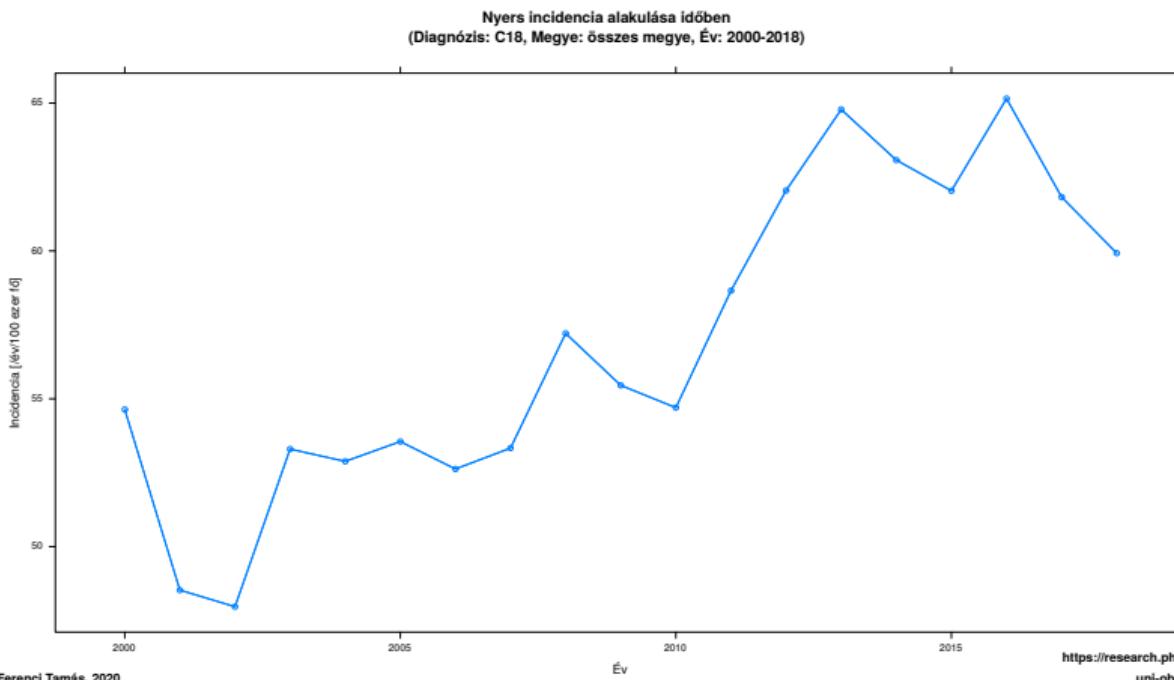
# A standardizálás szemléltetése: térbeli különbségek

Standardizált halálozási ráták:



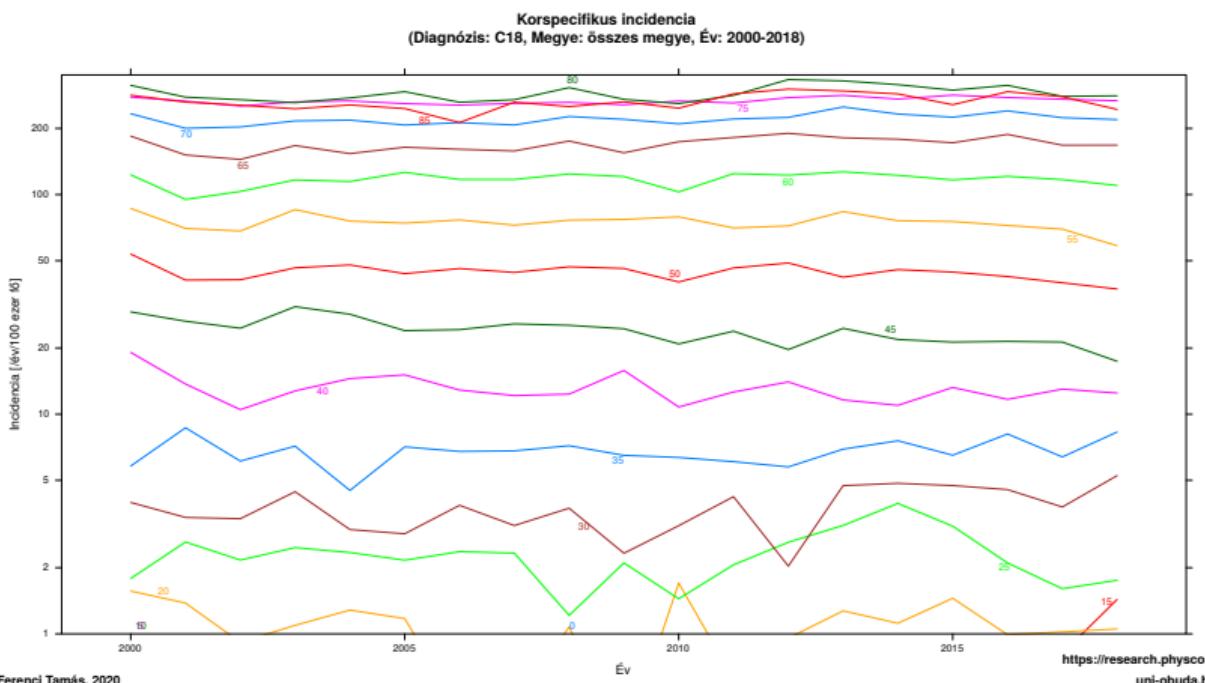
# A standardizálás szemléltetése: időbeli különbségek

Nyers halálozási ráták:



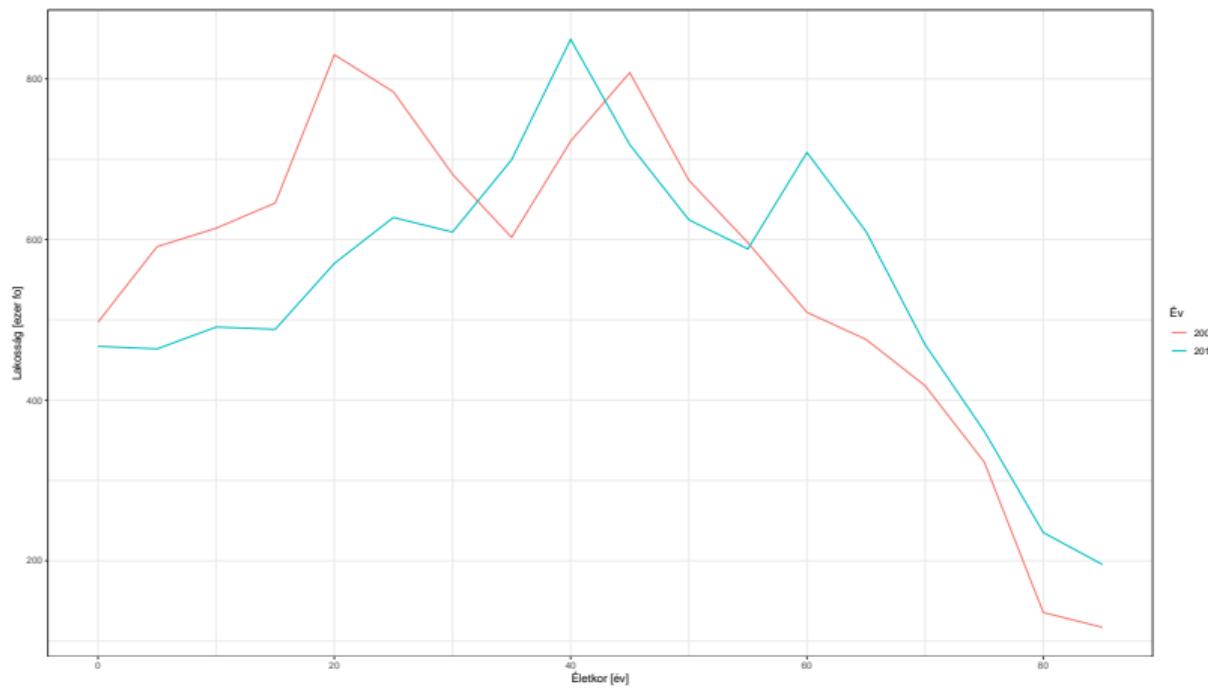
# A standardizálás szemléltetése: időbeli különbségek

Korspecifikus incidenciák változása időben:



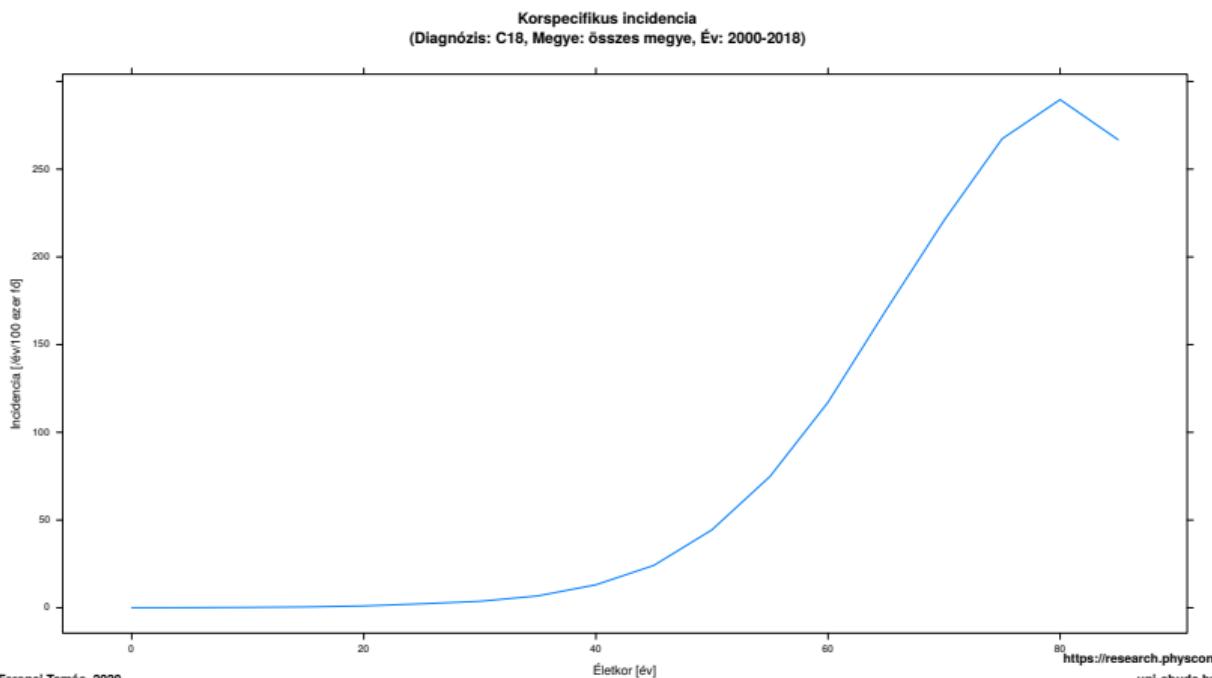
# A standardizálás szemléltetése: időbeli különbségek

A társadalom öregedése:



# A standardizálás szemléltetése: időbeli különbségek

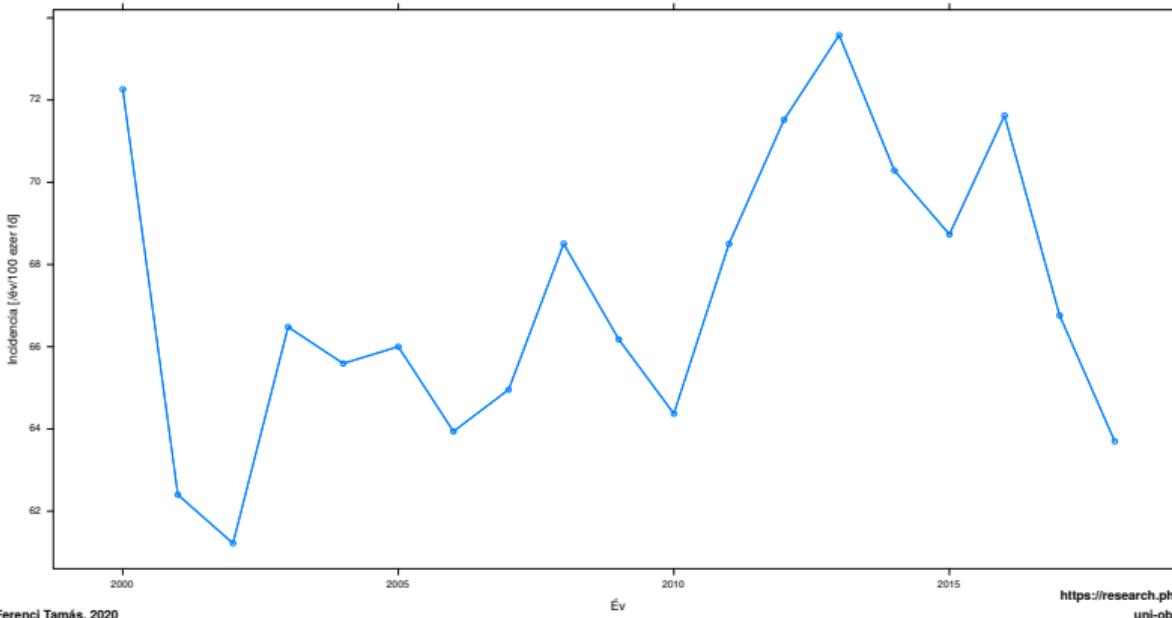
A rák korfüggése:



# A standardizálás szemléltetése: időbeli különbségek

Standardizált halálozási ráták:

Standardizált incidencia alakulása időben  
(Diagnózis: C18, Megye: összes megye, Év: 2000-2018, Standard: ESP, 2013)



Ferenc Tamás, 2020

[https://research.physcon.  
uni-obuda.hu](https://research.physcon.uni-obuda.hu)

# Illesztés (matching)

- Lényege: az összehasonlítandó csoportok alanyai között párokat képezünk, úgy, hogy azok tagjai lehetőleg minél hasonlóbbak legyenek, és így párosítva vizsgáljuk a végpontbeli eltéréseket, vagy más módon, de igyekszünk kiegyensúlyozni a csoportok confounderek szerint eltérő összetételét
- Előnyök:
  - ① Ha nem is tökéletesen univerzális, de elég enyhe feltételek mellett nyújt jó választ, miközben általában a korábbiaknál több confounder is figyelembe vehető
- Hátrányok:
  - ① Problémás lehet a „hasonlóság” meghatározása
  - ② Minél több szempont szerint kell hasonlóságot számolni, tehát minél több confounder van, annál nehezebb/bizonytalanabb lesz a feladat (még ha nem is annyira, mint a korábbi módszereknél)
  - ③ Nem feltétlenül használja gazdaságosan fel a mintát
  - ④ Gond van, ha kettőnél több expozíció szerinti csoport van, amiket össze kell hasonlítani

# Többváltozós – regressziós – modellezés

- Lényege: egy matematikai modellt tételezünk fel, melyben a végpontot mint változót valamilyen függvényforma szerint leírják a confounderek és az expozíció; ezt megbecsülve az expozíció hatása elkülöníthető (*ceteris paribus* értelmezés)
- Előnyök:
  - ① Jól képes számos confounder felhasználni, melyek lehetnek folytonosak és kategoriálisak (akár vegyesen is), rendkívül flexibilis, ugyan sok feltevésre épít, de azok többsége jól ellenőrizhető
- Hátrányok:
  - ① Nagyon sok confounder itt sem kezelhető
  - ② A modellfeltevések teljesülése mindig kérdéses

# Propensity score eljárások

- Lényege: meghatározzuk annak valószínűségét, hogy egy alany adott expozícióban részesüljön, a különböző ismert változói alapján, majd ezen ún. propensity score-ok alapján vagy hasonló párokat képezünk és körükben vizsgálódunk (PS matching), vagy ez alapján rétegzünk (PS stratification) vagy hagyományos elemzést végzünk, de úgy, hogy ez alapján meghatározott súlyokkal súlyozzuk az alanyokat (IPTW)
- Előnyök:
  - ① A többváltozós modellezéshez nagyon hasonló, talán kevesebb feltételezéssel a változók eloszlásáról és a köztük lévő kapcsolatokról (jelenleg is aktív vita tárgya)
- Hátrányok:
  - ④ A regressziós modellhez nagyon hasonló ( pontos összevetésük jelenleg is aktív vita tárgya)

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés
- Az empirikus orvosi kutatások alapgondolata és a kauzalitás
- A confounding problémája
- A confounding megoldásai: megfigyelés és kísérlet
- Okozati következtetések levonása megfigyeléses adatokból
- **A véletlen szerepéről**
- Bizonyitekok hierarchiája és összessége
- A végpont és meghatározása

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

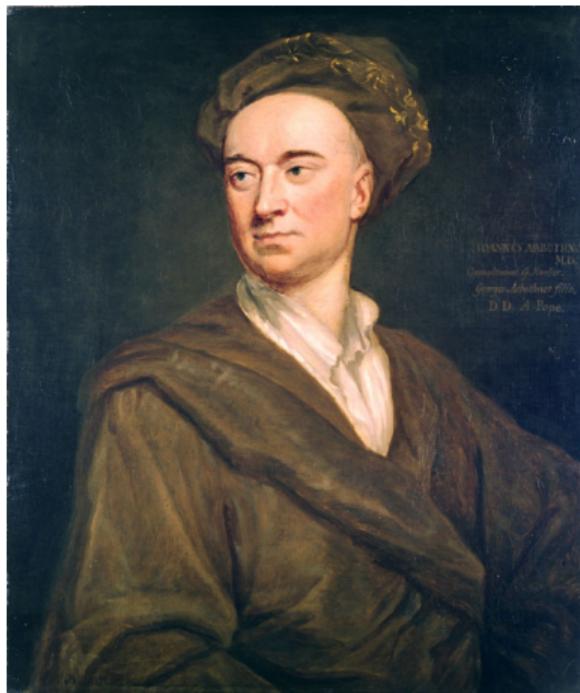
## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

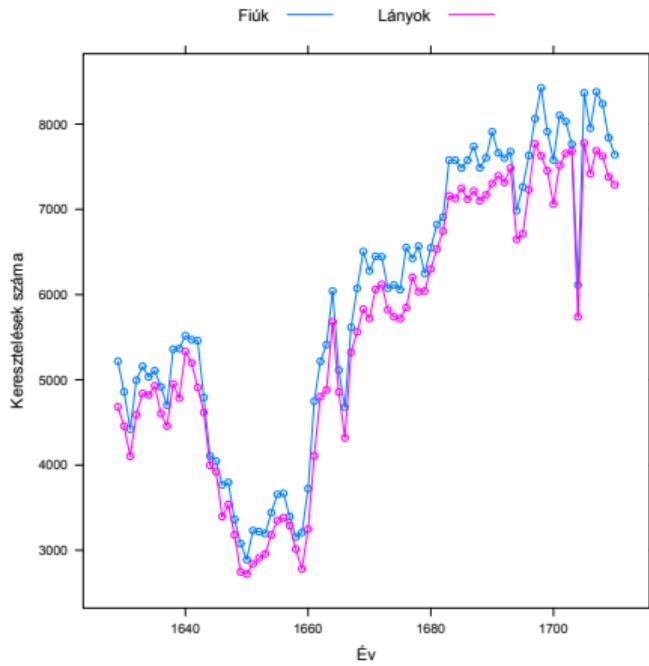
# Az első példa



# Az első példa

Christened.			Christened.		
Anno.	Males.	Females.	Anno.	Males.	Females.
1667	5616	5322	1689	7604	7167
68	6073	5560	90	7959	7302
69	6506	5829	91	7662	7392
70	6278	5719	92	7602	7316
71	6449	6061	93	7676	7483
72	6443	6120	94	6985	6647
73	6073	5822	95	7263	6713
74	6113	5738	96	7632	7229
75	6058	5717	97	8062	7767
76	6552	5847	98	8426	7626
77	6423	6203	99	7911	7452
78	6568	6033	1700	7578	7061
79	6247	6041	1701	8102	7514
80	6548	6299	1702	8031	7656
81	6822	6533	1703	7765	7683
82	6909	6744	1704	6113	5738
83	7577	7158	1705	8366	7779
84	7575	7127	1706	7952	7417
85	7484	7246	1707	8379	7687
86	7575	7119	1708	8239	7623
87	7737	7214	1709	7840	7380
88	7487	7101	1710	7640	7288

# Az első példa



# Arbuthnot gondolatmenete

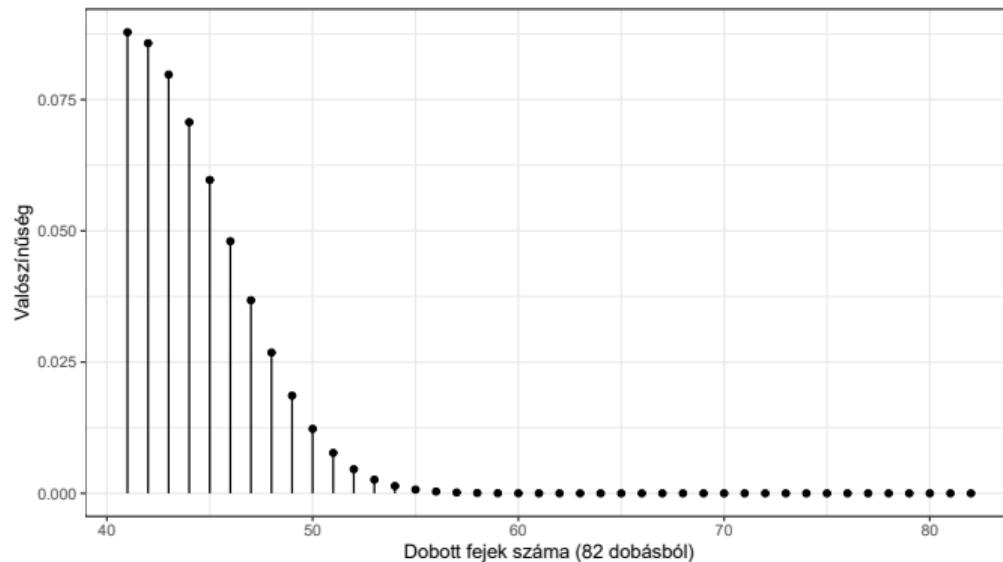
- Fiú/lány arány nem fele-fele: létezik Isten (teológiai kérdés, most elfogadjuk igaznak)
- De van egy statisztikai kérdés is: tényleg nem fele-fele?
- Analógia: minden év egy pénzfeldobás, fej – fiútöbbség, írás – lánytöbbség
- Arbuthnot okfejtése:
  - ① Az, hogy 82-ből 82-ször fejet dobtunk *nem jelenti biztosan*, hogy a pénzérme nem szabályos (=létezik Isten), *mert* szabályos pénzérmével is lehet 82-ből 82-ször fejet dobni
  - ② *De:* ennek a valószínűsége extrém kicsi ( $2,1 \cdot 10^{-23} \%$ )
- Alapvető észrevétel: ha ragaszkodunk hozzá, hogy *biztos* döntést hozzunk, akkor nem tudunk válaszolni...
- ...és ez *mindig* így van! Ha 820-ból 820 fej lett volna, *biztosan* akkor sem tudunk dönteni
- Az *egyetlen* biztos válasz, hogy „nem tudom”, minden máshoz *muszáj* valamennyi hibát vállalni

# Arbuthnot gondolatmenete

- Hiba vállalása tehát: akkor is azt mondjuk, hogy szabálytalan a pénzérme, ha valójában szabályos
- E nélkül ugyanis soha nem tudunk semmit mondani: ha csak akkor mondjuk, hogy szabálytalan, ha tényleg biztosan az, akkor soha nem mondhatjuk, mert szabályosság mellett is kaphatunk, nem nulla valószínűsséggel, csupa fejet, akárhányszor is dobuk fel
- Ami igazán fontos: ezt a hiba-vállalást tudjuk számszerűsíteni
- Ha csak akkor mondjuk, hogy szabálytalan a pénzérme, ha 82-ből 82 fejet dobunk (de már 81-nél is azt mondjuk, hogy „á, ez lehetett a véletlen miatt!”) akkor a fent említett nagyon kis hibavalószínűséget vállaljuk
- Ha már 81-nél is azt mondjuk, hogy „nem hisszük, hogy ez a véletlen miatt volt” (mert ha szabályos lenne, akkor ez annyira valószínűtlen lenne), akkor kicsit nagyobb a hiba ( $1,7 \cdot 10^{-21} \%$ )
- Ez mind számszerűen meghatározható, hiszen a szabályosság esetén tudjuk, hogy 50% valószínűsséggel dobunk fejet, ezért – ha még azt is hozzá tesszük, hogy függetlenek a dobások – ebből minden kiszámolható

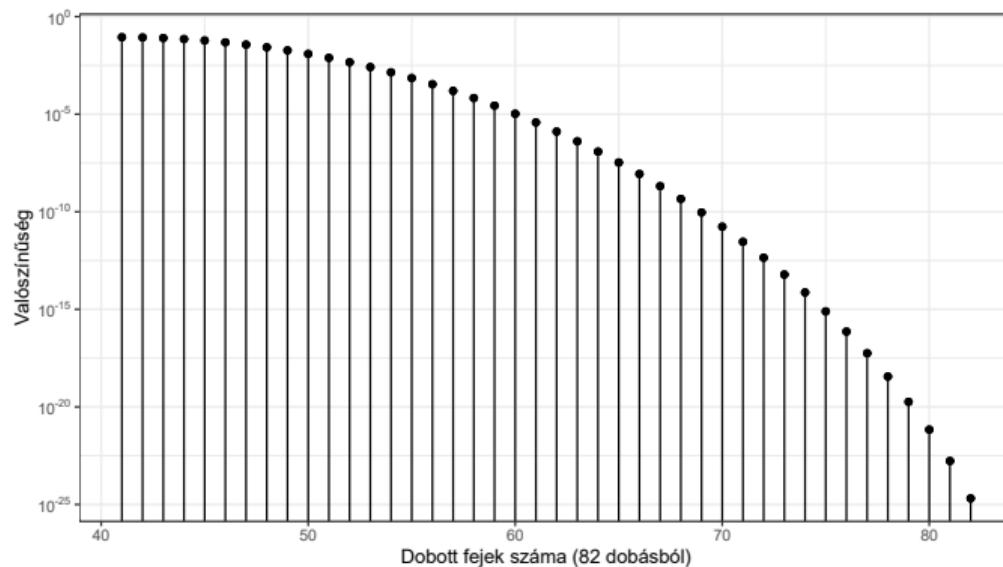
# Arbuthnot gondolatmenete

Számoljuk is ki ténylegesen: ha szabályos a pénzérme, ekkora valószínűséggel dobunk mégis, pusztán a véletlen ingadozás szeszélye révén adott számú fejet 82 dobásból



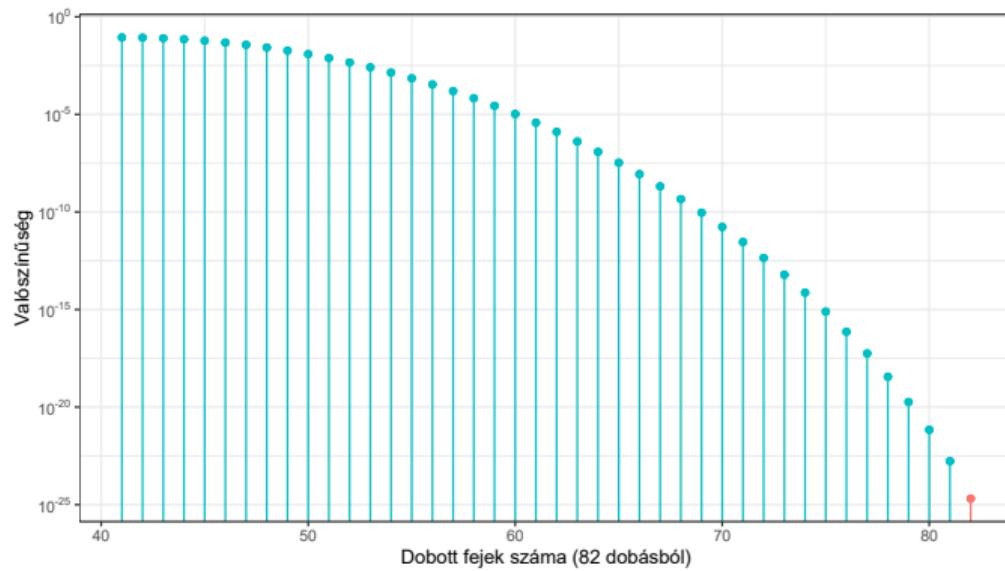
# Arbuthnot gondolatmenete

Számoljuk is ki ténylegesen: ha szabályos a pénzérme, ekkora valószínűséggel dobunk mégis, pusztán a véletlen ingadozás szeszélye révén adott számú fejet 82 dobásból



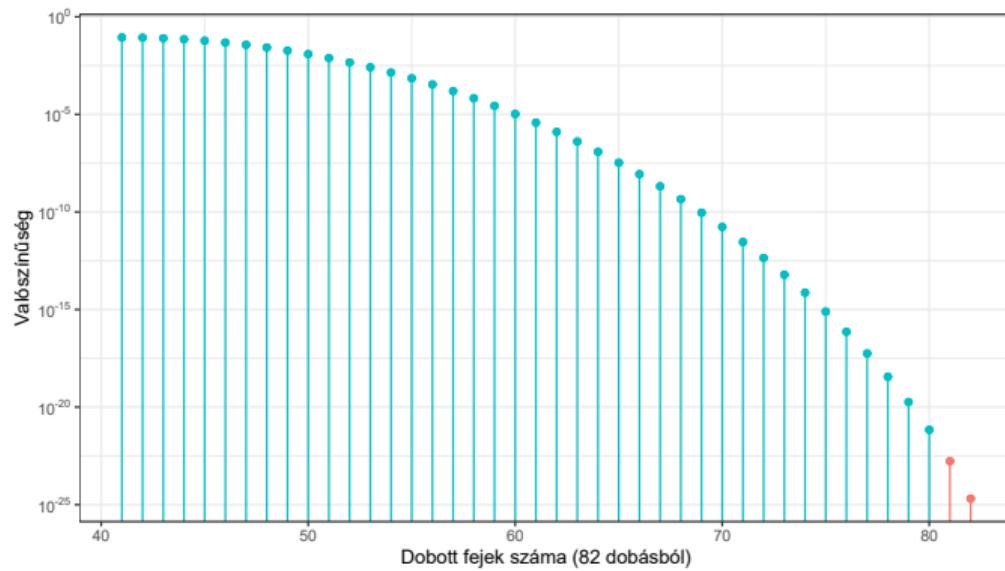
# Arbuthnot gondolatmenete

Ha csak 82-nél mondjuk, hogy szabálytalan:  $2,1 \cdot 10^{-23} \%$  – ilyen értelmű – hibavalószínűség



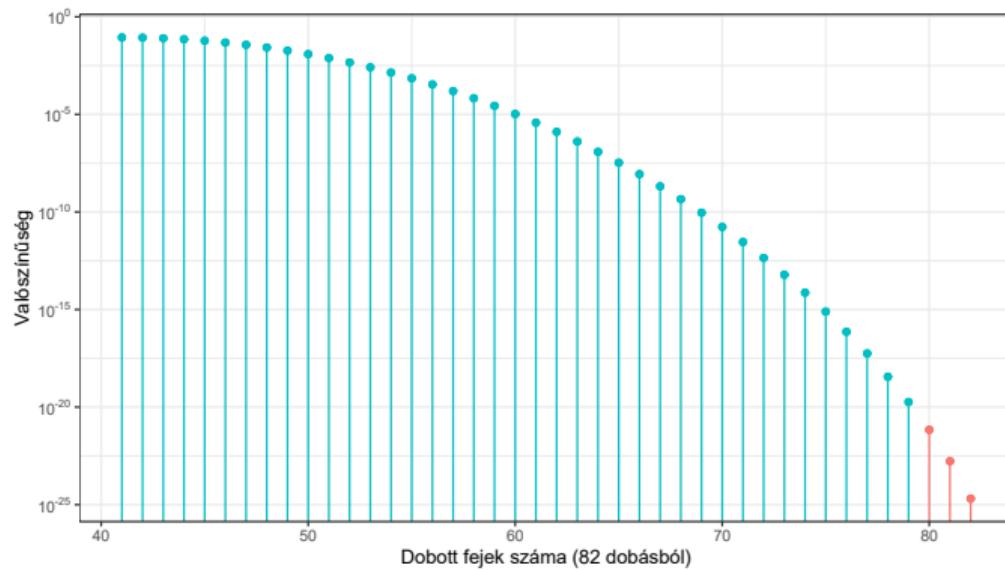
# Arbuthnot gondolatmenete

Ha 81-nél és 82-nél mondjuk, hogy szabálytalan:  $1,7 \cdot 10^{-21} \%$  – ilyen értelmű – hibavalószínűség



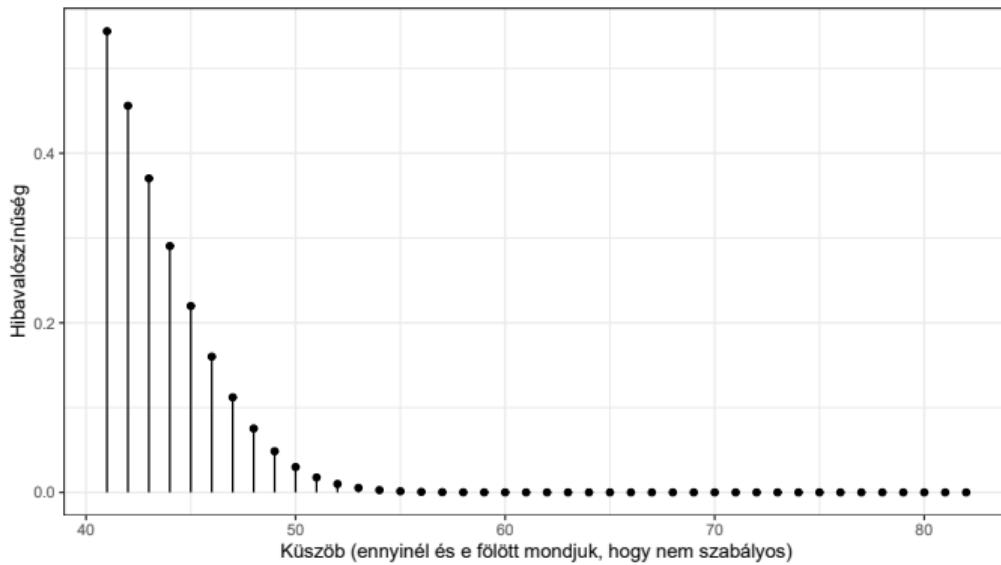
# Arbuthnot gondolatmenete

Ha 80-nál, 81-nél és 82-nél mondjuk, hogy szabálytalan:  $7,0 \cdot 10^{-19} \%$  – ilyen értelmű – hibavalószínűség



# Arbuthnot gondolatmenete

Mekkora hibát vállalunk (a „tévesen mondjuk, hogy szabálytalan a pénzérme (van Isten) miközben igazából nem” értelemben), ha adott küszöböt választunk



# Arbuthnot gondolatmenete

- Például úgy szeretnénk dönten, hogy 5% hibát vállalunk ebben az értelemben (tehát ha szabályos a pénzérme, 5% valószínűséggel mondjuk mégis, tévesen, a véletlen ingadozás szeszélye folytán, hogy nem szabályos)
- Kiszámolható, hogy ehhez mi lesz a megfelelő határ, lásd előző ábra
- A válasz 49 (kb. 5% hibavalószínűség)
- („kb.”: a diszkrétség miatt nem lehet itt most teljesen pontosan beállítani)
- Ha abban az esetben mondjuk, hogy nem szabályos a pénzérme, ha 49-szer vagy annál többször dobunk fejet 82-ből, akkor e stratégia mellett igazolhatóan 5% lesz a valószínűsége, hogy egy szabályosra azt mondjuk, hogy szabálytalan
- (Más szóval: ha nagyon sokszor kell ilyen helyzetben döntenünk, akkor a fenti szabályt követve hosszú távon az esetek 5%-ában fogunk ilyen értelmű hibát véteni)
- Ha nem 82-ször dobjuk fel a pénzérmét, akkor természetesen más lesz a döntési szabály, de a fenti módon meghatározható minden esetben a szükséges küszöb

# Mi köze ennek az orvosi vizsgálatok kiértékeléséhez?

Arbuthnot fenti modellje szinte tökéletesen megfelel a mai orvosi vizsgálatok kiértékeléséhez túlnyomó részében alkalmazott modellnek, csak a szavakat kell lecserélni:

- Pénzérme feldobása: adott beteg meggyógyul-e – fej: igen, írás: nem
- Pénzérme fej-dobási valószínűsége: gyógyulási valószínűség a gyógyszeres kezeléssel
- Tegyük fel, hogy gyógyszer nélkül a betegek fele meghal, és hogy a gyógyszer csak javítani tud (vagy nem változtat, ha hatástalan)
- (Valójában a gyógyszer nélküli arányt sem tudjuk kőbevésetten – ezért kell két csoport – és az sem biztos, hogy csak javítani tud, de most egyszerűsítsünk, hogy a lényegre koncentráljunk)
- 82 betegem van, akinek gyógyszert adok, 82 meggyógyul – hat-e a gyógyszer?
- Tökéletes a megfeleltetés: hatástalan a gyógyszer, azaz 50% a kezelés melletti gyógyulási arány – szabályos a pénzérme, hatásos a gyógyszer, azaz 50%-nál nagyobb a gyógyulási arány – szabálytalan a pénzérme (figyelem, *bármennyivel* nagyobb)
- Mintanagyság – érmedobások száma
- Ugyanúgy csak hibával terhelt választ lehet hozni, *de* úgy, hogy a hiba mértéke maga is kontrollálható

# Mi köze ennek az orvosi vizsgálatok kiértékeléséhez?

- Ilyen értelmű hiba: a hatástalan gyógyszert tévesen hatásosnak hisszük
- Adja magát egy kérdés: ha ez hiba (ráadásul elég drámai, egy hatástalan gyógyszert törzskönyvezünk), és a mértékét mi szabjuk meg, akkor miért 5%-ot választottunk?? Miért nem 4-et? 1-et? Egymilliomodot??
- Nem beugratós a kérdés: ha ezt az értéket csökkentjük akkor *tényleg* csökken annak a valószínűsége, hogy hatástalan gyógyszert hatásosnak hiszünk...
- ...csak épp ezzel a hibázással szemben áll egy másik: hogy egy hatásost is hatástatlanak hiszünk!
- Ha csökkentjük ezt a valószínűséget (nem elég 50 betegnek meggyógyulnia, 55-nek kell, 60-nak, stb.), akkor csakúgyan ritkábban fogunk hatástalan gyógyszereket törzskönyvezni, de egyre gyakrabban fogjuk még a hatásosakra is nagy valószínűsséggel azt mondani, hogy „á, ez lehetett a véletlen miatt” – erre a gondolatra mindenki visszatérünk
- Vegyük észre, hogy ebben a gondolkodási keretben egy fordított logika érvényesül: nem azt nézzük, hogy ezen eredmény alapján mennyire valószínű, hogy hatástalan a gyógyszer, hanem azt, hogy ha hatástalan lenne, akkor mennyire valószínű, hogy ilyen eredményt kapnánk

# Precíz terminológia

- Nullhipotézis (null hypothesis)
  - Kiinduló, „alapértelmezett” állításunk, amit további információ nélkül elhiszünk
  - Tipikusan azt fogalmazza meg, hogy nem találtunk semmi újat (nincs eltérés, nincs hatás, nincs különbség)
  - Szabályos a pénzérme, nincs különbség a fiú-lány arányban, nem hat a gyógyszer
  - Lényegében van egy kiindulópontunk és nézzük, hogy a begyűjtött adat mennyire mond neki ellent
- Ellenhipotézis (vagy alternatív hipotézis, alternative hypothesis)
  - A nullhipotézis tagadása
  - Újdonság, heuréka, találtunk valamit (van eltérés, van hatás, van különbség)
  - Nem szabályos a pénzérme, eltér a fiúk és a lányok aránya, hat a gyógyszer
- A nullhipotézis általában csak egyféleképp tud teljesülni, az ellenhipotézis több módon is (a szabályosság 50% fejdobási valószínűség és csak az, szabálytalan sokféleképp lehet; a hatástalanság csak a 0 hatás, hatásos sokféleképp lehet, hathat kicsit is, meg nagyon is stb.)

# Precíz terminológia

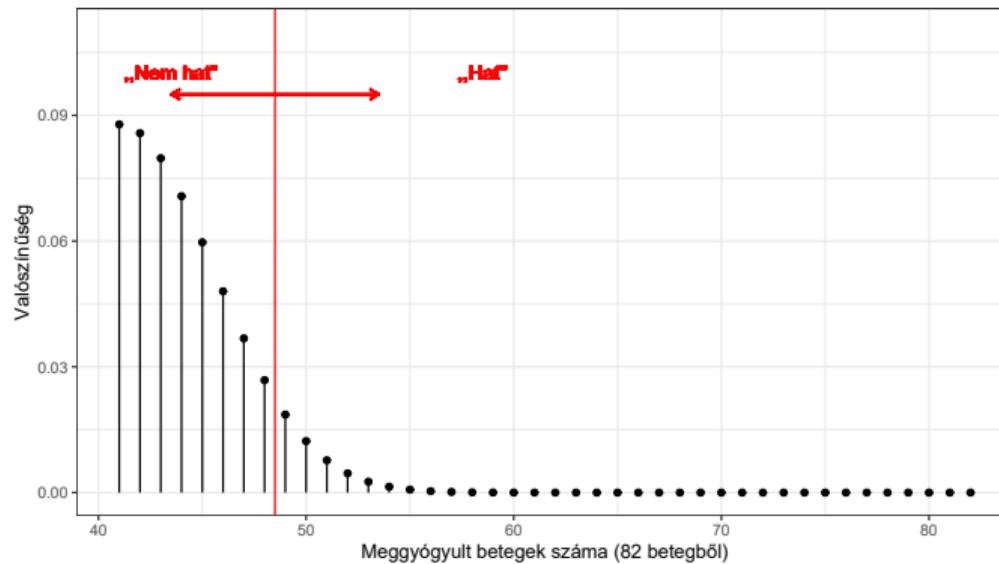
- Szignifikanciaszint (significance level): a célvalószínűség, ami alapján meghatározzuk a küszöb helyét; jele:  $\alpha$
- Elsőfajú hiba (Type I error)
  - Elvetjük a nullhipotézist, pedig igaz
  - Tévesen minősítünk hatásosnak egy hatástalan gyógyszert
  - Ha teljesülnek a próba előfeltevései, akkor a valószínűsége épp az általunk választott szignifikanciaszint,  $\alpha$
  - Azaz: a valószínűsége felett erős kontrollunk van (semmi egyébtől nem függ – természetesen a döntési szabály függhet például a mintanagyságtól, de az elsőfajú hiba valószínűsége nem)
- Másodfajú hiba (Type II error)
  - Nem vetjük el a nullhipotézist, pedig nem igaz
  - Tévesen minősítünk hatástartannak egy hatásos gyógyszert
  - A valószínűségének a jele:  $\beta$  (annak a valószínűsége, hogy nem vetjük el a nullhipotézist, feltéve, hogy nem igaz)
  - $1 - \beta$  az erő: feltéve, hogy hat a gyógyszer, mekkora valószínűsséggel mutatjuk is ezt ki
  - Nincs felette erős kontrollunk

# Precíz terminológia

- Szignifikáns hatás: adott szignifikanciaszinten elvetjük a nullhipotézist
  - Tartalmilag: a tapasztalt hatás – 42-nél többen gyógyultak meg – betudható-e egyszerűen a véletlen ingadozásnak, tehát a valóságban nem hat a gyógyszer és ami hatást láttunk az csak ebből a véletlenségből fakadt, vagy olyan nagy, hogy azt feltételezzük, hogy valódi különbség van mögötte (tényleg hat a gyógyszer)
- „Elfogadunk”: elfogadjuk a nullhipotézist (azaz azt mondjuk, hogy nem találtunk hatást, pl. nem hat a gyógyszer) – bár jobb lenne úgy fogalmazni, hogy „nem tudunk elutasítani”, később majd kiderül, hogy miért
- „Elutasítunk”: elutasítjuk a nullhipotézist (azaz azt mondjuk, hogy találtunk hatást, pl. hat a gyógyszer)

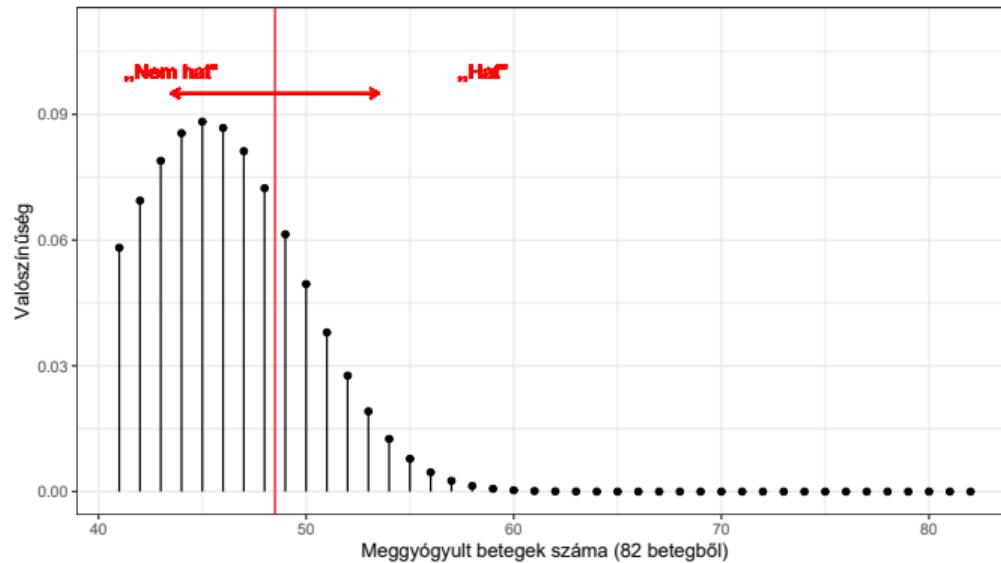
# Az erő és a hatásosság összefüggése

Mi az, hogy erő? Vegyük elő az ábrát hatástaralanság esetén (ez ugye még nem jó az erőhöz!):



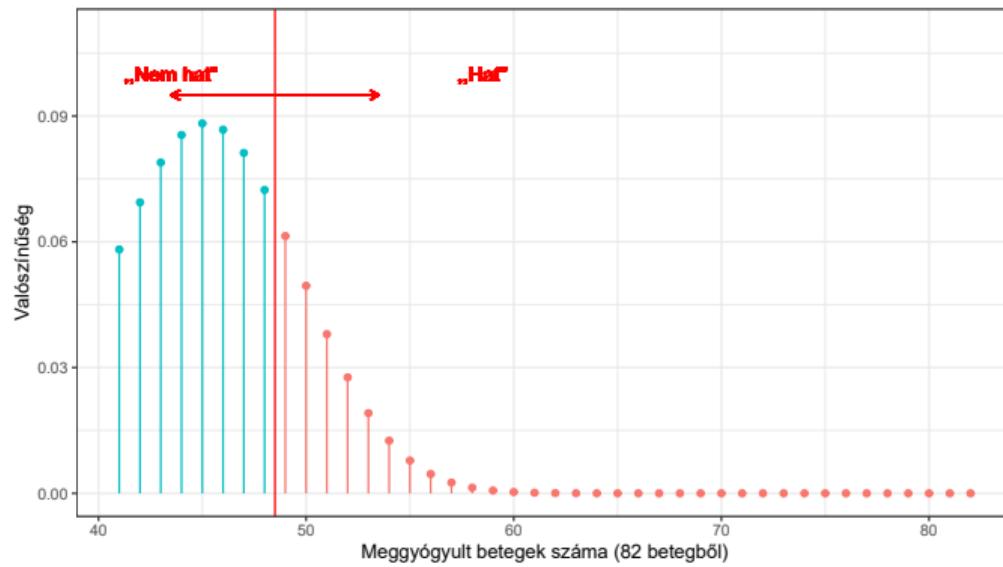
# Az erő és a hatásosság összefüggése

Mi az, hogy erő? Feltéve, hogy hat a gyógyszer (50-ről 55%-ra emeli a gyógyulási arányt)



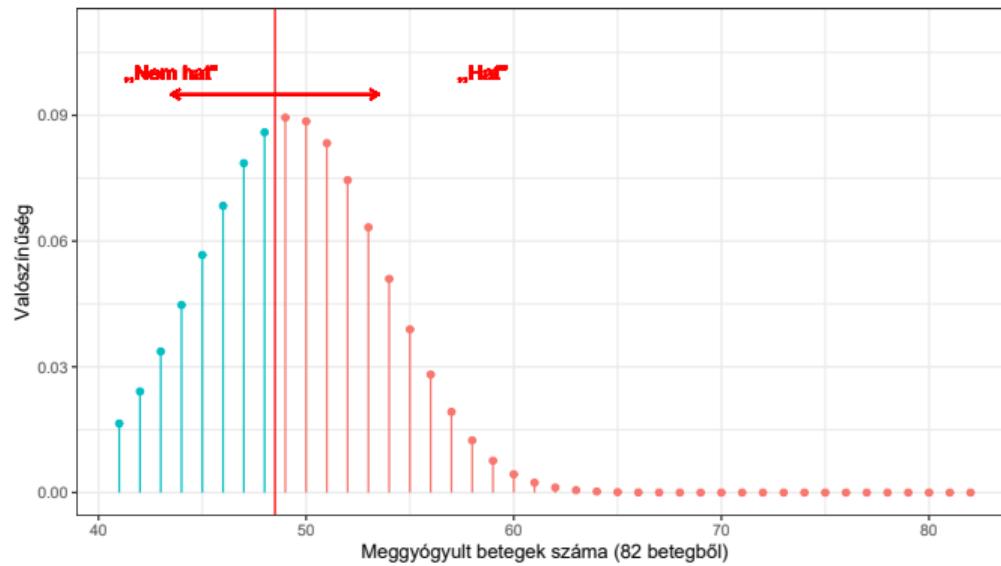
# Az erő és a hatásosság összefüggése

Mi az, hogy erő? Feltéve, hogy hat a gyógyszer (50-ről 55%-ra emeli a gyógyulási arányt), mekkora valószínűséggel mondjuk mi is azt, hogy hat



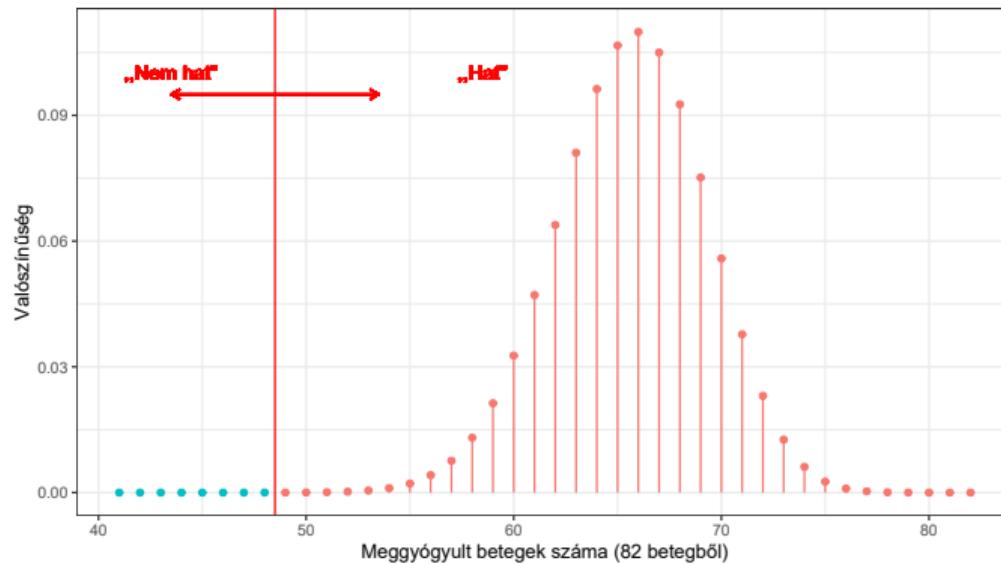
# Az erő és a hatásosság összefüggése

Kicsit jobban ható gyógyszer (60%)...

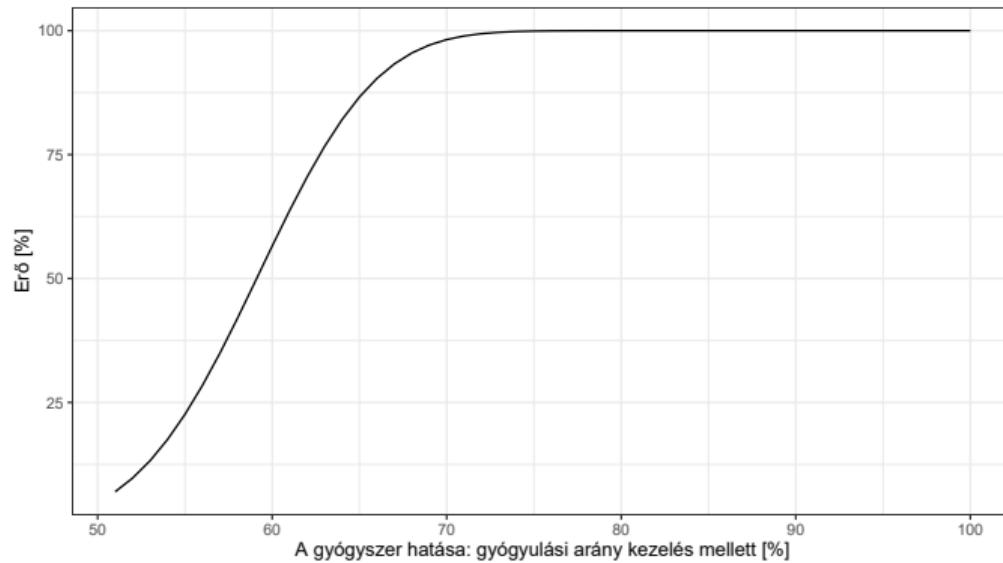


# Az erő és a hatásosság összefüggése

Még jobban ható gyógyszer (80%)...



# Az erő és a hatásosság összefüggése



# Az erő és a hatásosság összefüggése

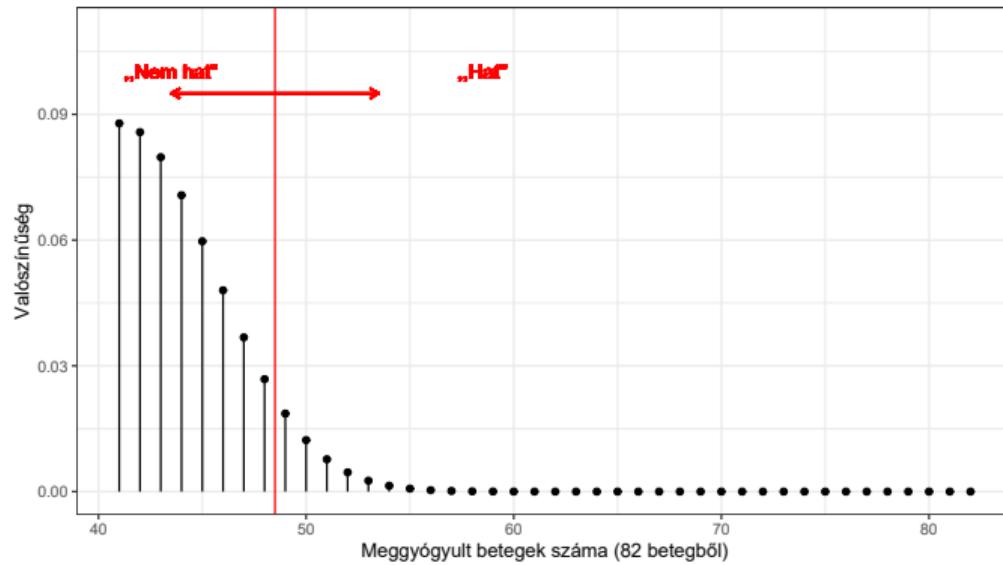
- Természetesen a döntési határ minden esetben ugyanott van (hiszen azt a hatástaralanság feltételezése mellett a eloszlás alapján határoztuk meg, nem függ a gyógyszer hatásától!)
- Tanulság: az erő függ a gyógyszer valódi hatásától – minél hatásosabb a gyógyszer, annál nagyobb lesz ( minden más változatlanul tartva) az erő

# Az erő és a mintanagyság összefüggése

- Hogyan tudjuk növelni az erőt? (hiszen a gyógyszer hatásosságára nincs ráhatásunk)
- Intuitív ötlet: növeljük meg a mintanagyságot! („kiátlagolódnak az eltérések”)
- Csakugyan így van, nézzük meg miért

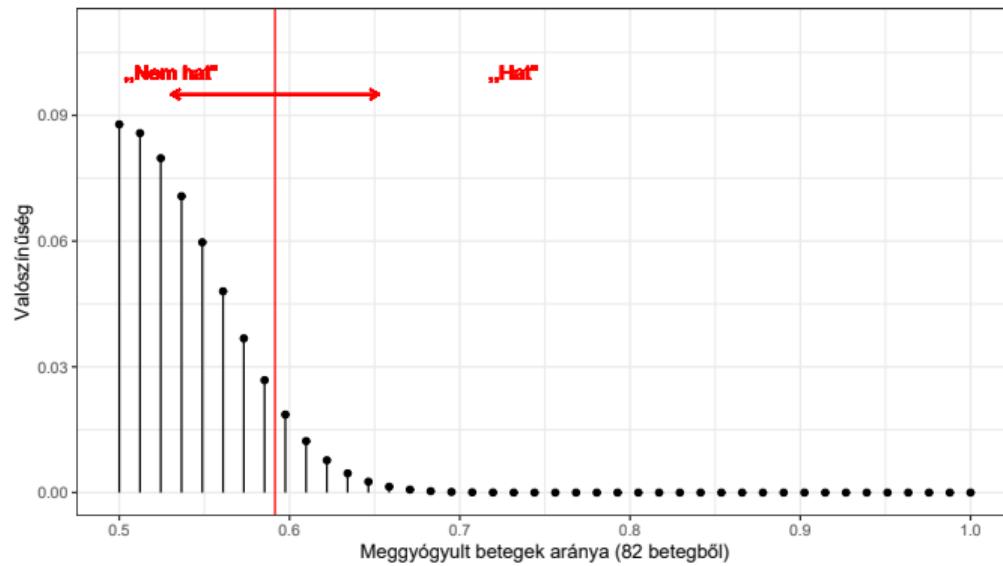
# Az erő és a mintanagyság összefüggése

Az eredeti ábránk (82 beteggel)



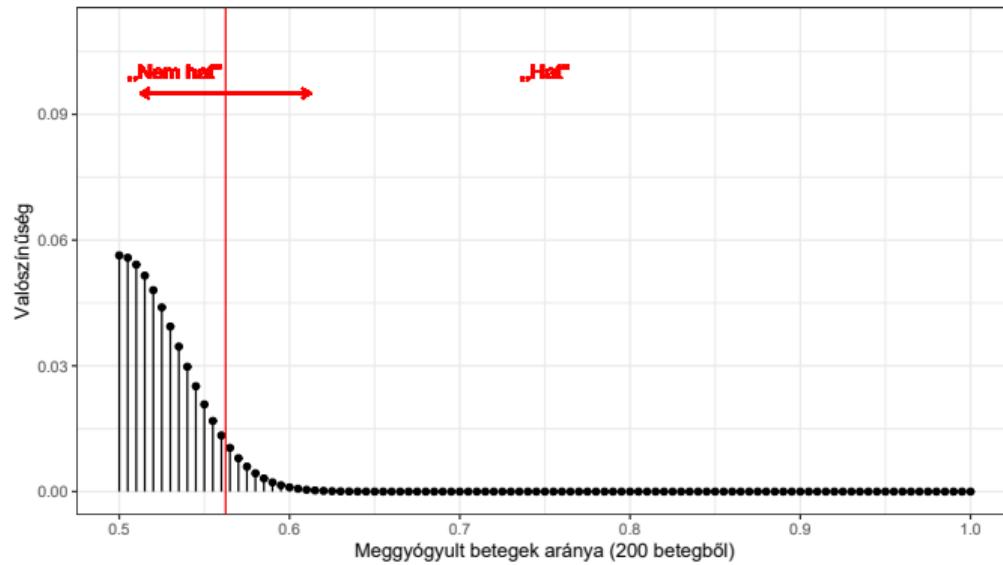
# Az erő és a mintanagyság összefüggése

A vízszintes tengelyt skálázzuk át (csak hogy összehasonlítható legyen később)



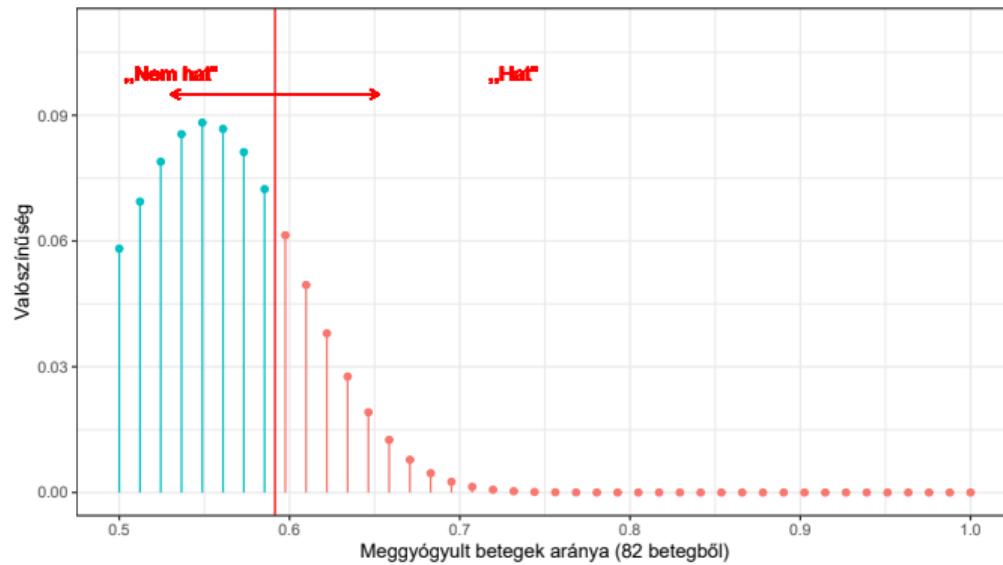
# Az erő és a mintanagyság összefüggése

Ugyanez az ábra 200 beteggel (mi változott?)



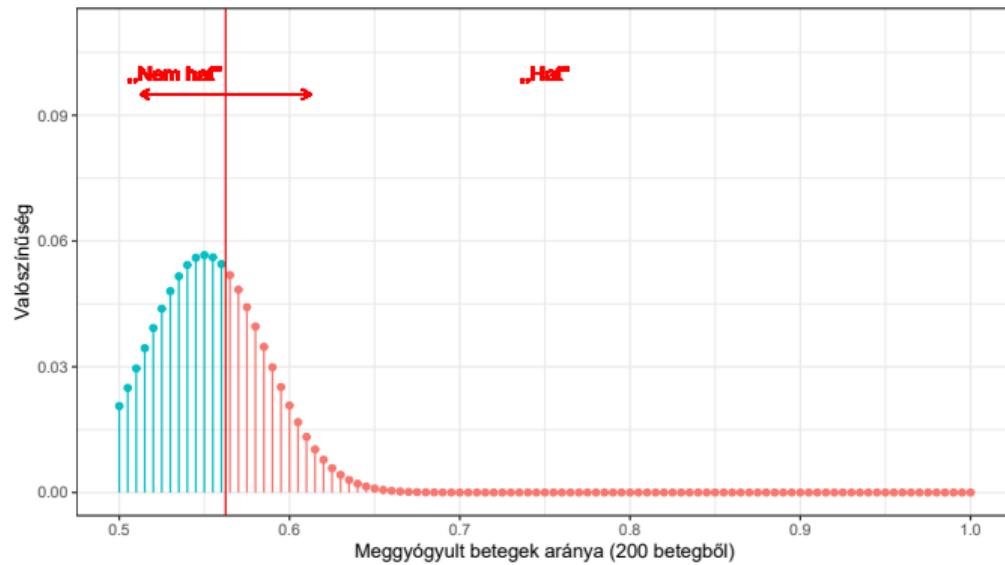
# Az erő és a mintanagyság összefüggése

Kicsit ható gyógyszer, 82 beteg



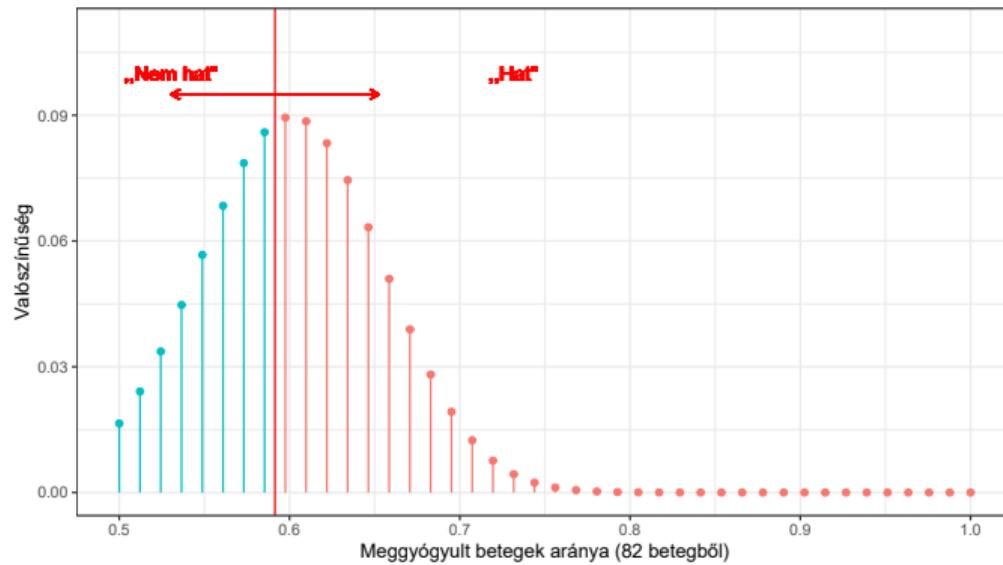
# Az erő és a mintanagyság összefüggése

Kicsit ható gyógyszer, 200 beteg



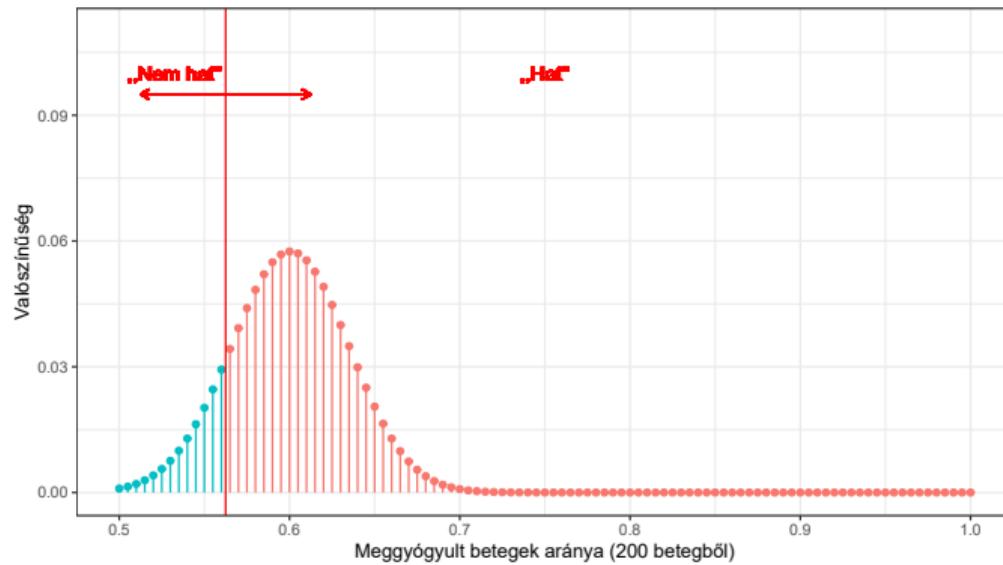
# Az erő és a mintanagyság összefüggése

Közepesen ható gyógyszer, 82 beteg



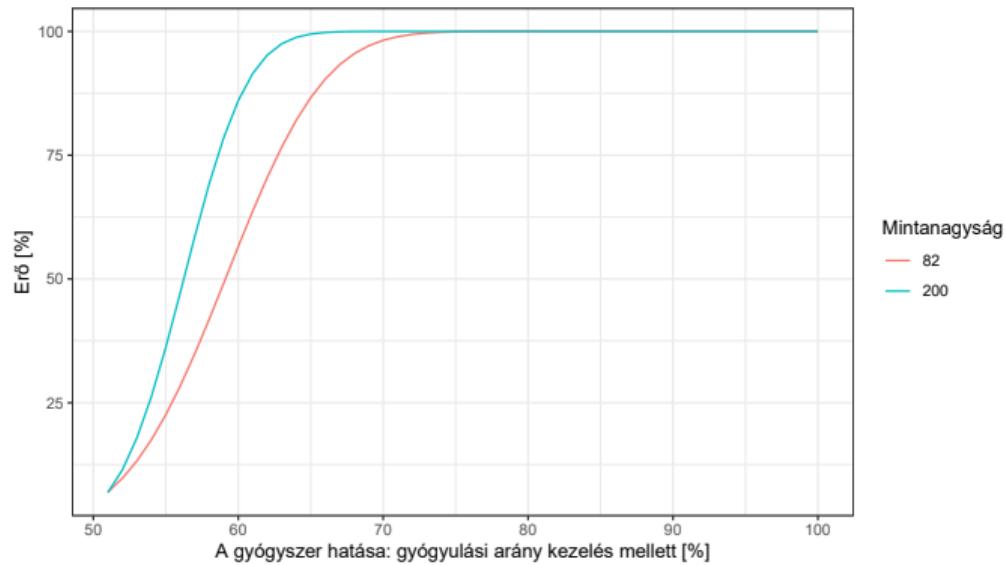
# Az erő és a mintanagyság összefüggése

Közepesen ható gyógyszer, 200 beteg



# Az erő és a mintanagyság összefüggése

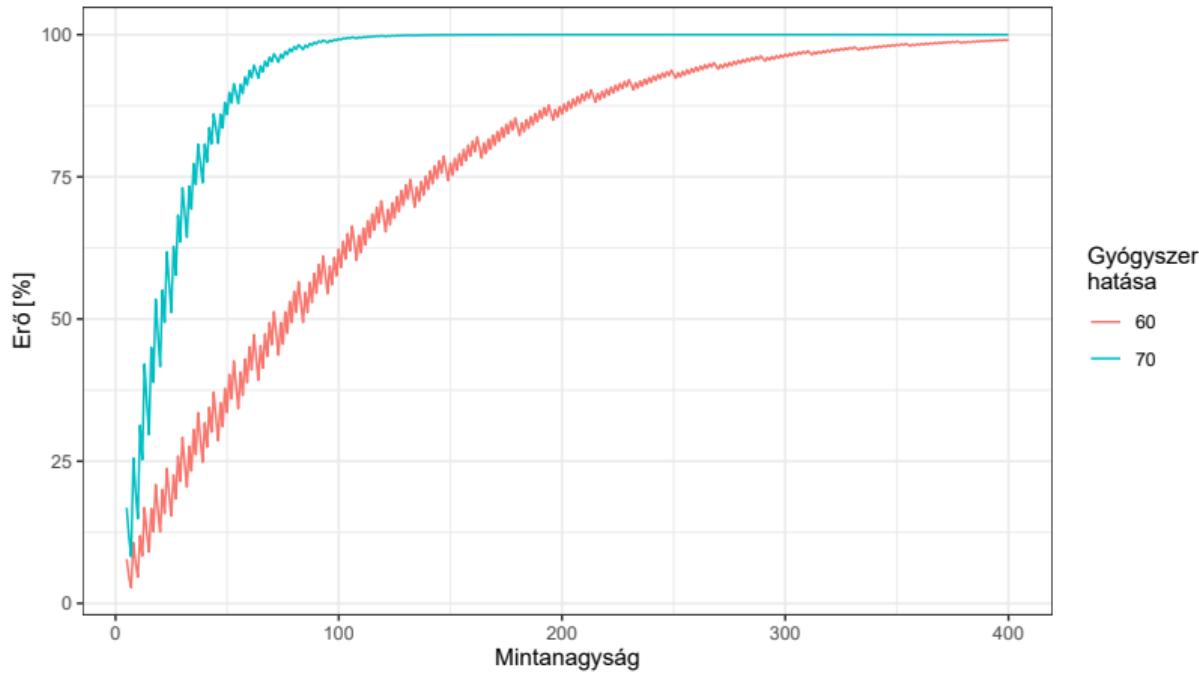
A végeredmény:



# Az erőt meghatározó tényezők

- Végeredményben tehát azt látjuk, hogy igazából két tényező hat az erőre: a gyógyszer hatásossága (hatásnagyság) és a mintanagyság
- (Sőt, valójában három, mert a szignifikanciaszint is)
- Kétféleképp is elmondható: ugyanazon hatásosság mellett nagyobb az erő, ha nagyobb a mintanagyság, avagy kisebb hatás is kimutatható ugyanazon erővel, ha nagyobb a mintanagyság
- „Kis hatáshoz nagy minta kell, nagy hatáshoz kis minta is elég”
- (Mármint a kimutatásához, értsd: hogy igazolni is tudjuk, hogy az nem csak a véletlen ingadozás miatt van, tehát, hogy szignifikáns)
- Nézzük meg úgy is ezt a függvényt, hogy a hatást rögzítjük, és a mintanagyságot variáljuk

# Az erő és a mintanagyság összefüggése



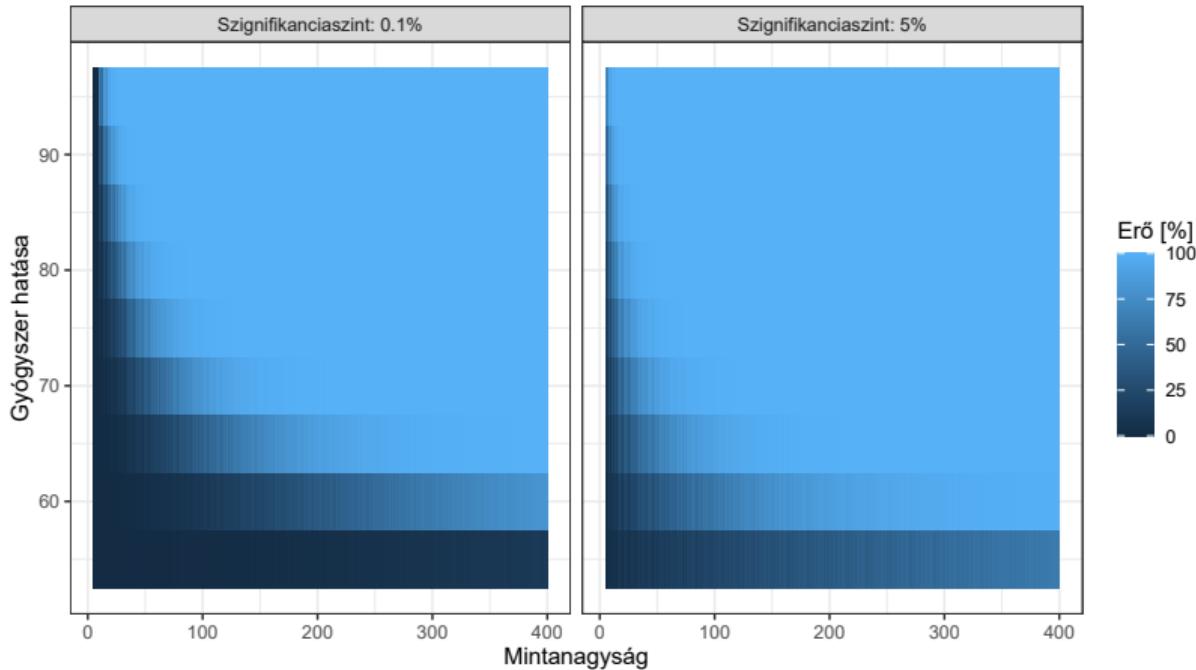
## Racionális mintanagyság-tervezés, erőelemzés

- (A furcsa szaggatás a diszkrétség miatt van – nem tud 70,2 beteg meggyógyulni – ne törődjünk vele)
- Ez egy nagyon fontos dolgot lehetővé tesz: a mintanagyság racionális tervezését!
- Csak az előbbi ábrára „fordítva” nézünk rá: nem mintanagysághoz olvassuk le az erőt, hanem adott, megkívánt erőhöz – tipikusan 80% – keressük ki a mintanagyságot
- (Ellenőrző kérdés: miért 80? Miért nem 90? 99? – hiszen akkor jobb eséllyel igazolni is tudjuk, hogy hat a gyógyszer, ha tényleg hat...!)
- Figyelem: minden ilyen számítás tehát feltételes azon, hogy mekkora a gyógyszer hatása!
- Furcsa lehet, hogy előre meg kell adnom a hatást, hiszen a kutatás célja épp ennek kiszámolása, de ez olyan mint a csillagászat: minél halványabb csillagot akarok kimutatni, annál nagyobb távcső kell – úgyhogy amikor távcsövet veszünk, kell valamilyen feltételezéssel élni a csillag fényességéről

# Racionális mintanagyság-tervezés, erőelemzés

- Kétféleképp is elmondható: ugyanazon mintanagyság mellett nagyobb az erő, ha hatásosabb a gyógyszer, avagy ugyanazon erőhöz kisebb minta is elég, ha hatásosabb a gyógyszer
- (Látszott az előző ábrán is: igazából ugyanazt a háromdimenziós függvényt mondjuk el, csak kérdés, hogy hogyan nézünk rá, milyen metszetet készítünk belőle)

# Az erőt meghatározó tényezők



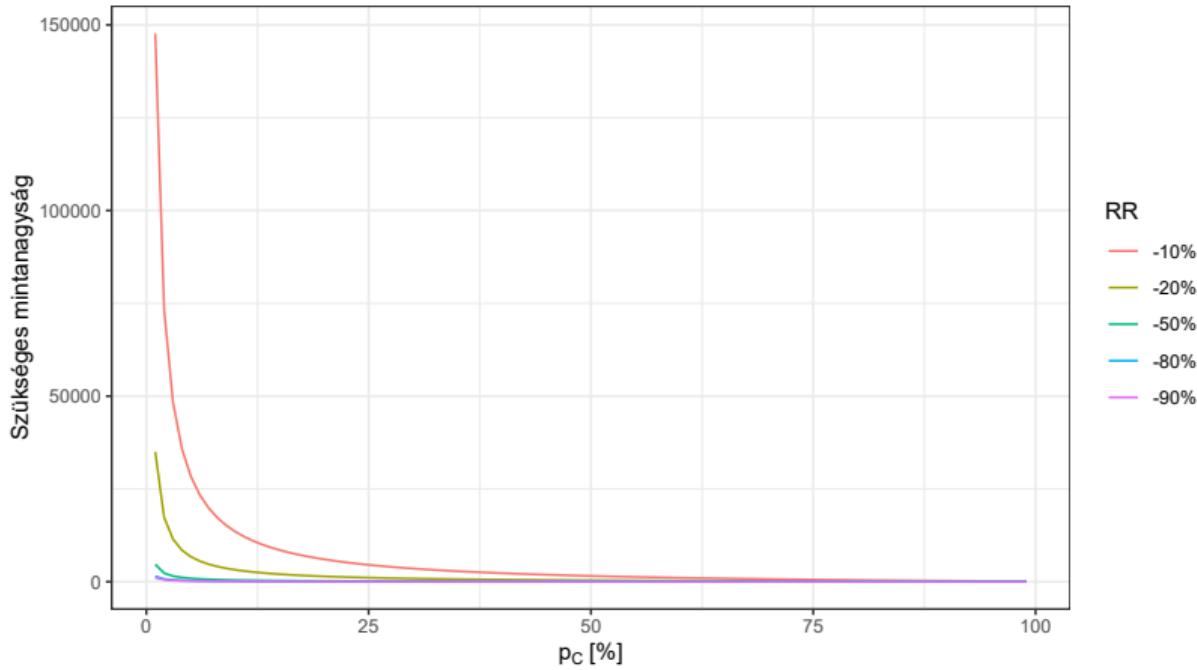
# Az erőt meghatározó tényezők

- Itt most egyszerre látjuk az összes tényezőtől való függést
- Bejött a szignifikanciaszint is
- Tanulság: ha lecsökkentjük, leromlik az erő is
- (Emlékezzünk vissza mit mondtunk már a legelején: ha nagyon szigorúak vagyunk, akkor a valódi hatást is nehéz lesz észrevenni...)

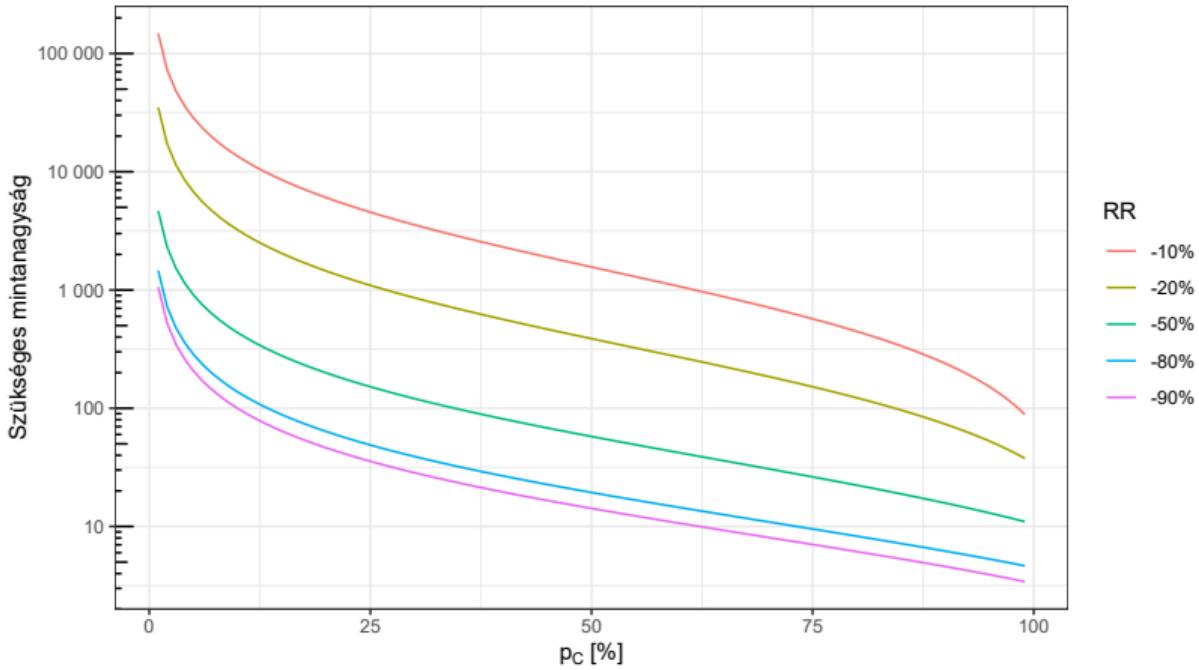
# Egy valódi esettanulmány

- Igazi példa: most már két csoport van, gyógyszer nem csak javítani tud
- Profilaktikus példa: a kontrollcsoportban sem biztos, hogy mindenki elszennvedi a végpontot
- A kontrollcsoportban  $p_C$  a kockázat, a kezelés ezt 10, 20, 50, 80 illetve 90%-kal csökkenti
- Mekkora mintanagyság kell ennek a kimutatásához?
- (Szokásosan  $\alpha = 0,05$  szignifikanciaszint,  $1 - \beta = 0,8$  erő)

# Egy valódi esettanulmány



# Egy valódi esettanulmány



# Egy valódi esettanulmány

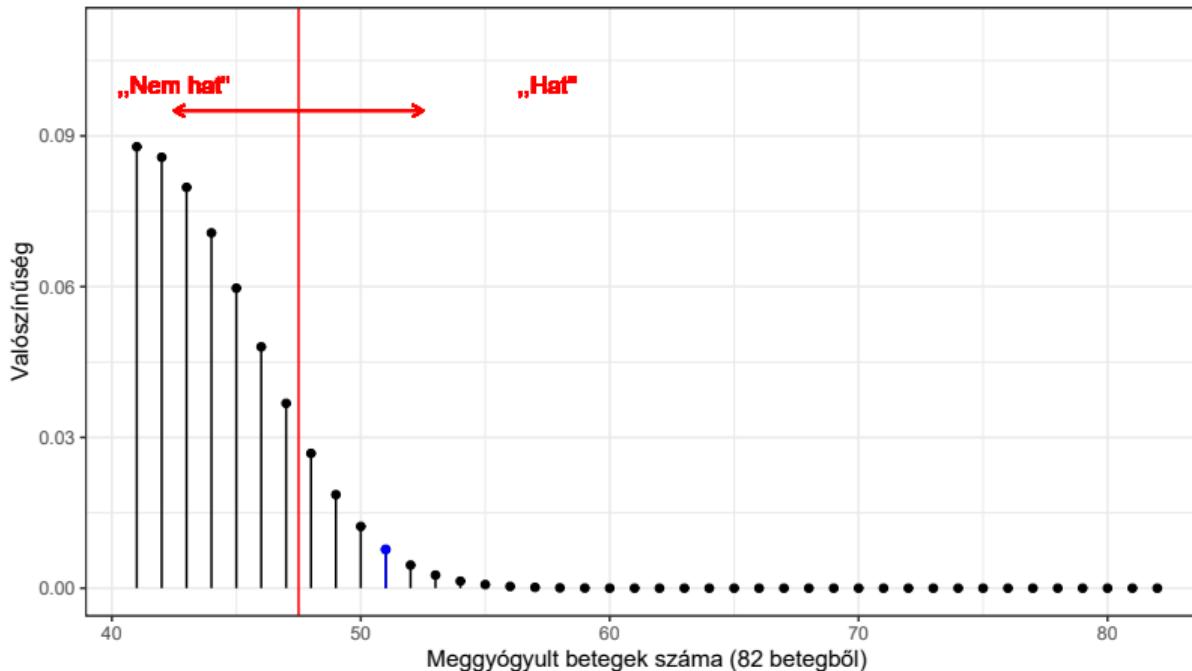
- Gondoljuk végig az eddigieket az ábrán!
- Egy plusz tanulság: annál könnyebb szignifikánsnak kimutatni valamit, minél nagyobb az abszolút különbség is a csoportok között (még ha a relatív különbség állandó is!)
- Lásd vakcinás példa:
  - Ezért hatalmasak az ilyen kísérletek
  - Persze kölcsönhat az utánkövetés hosszával (abszolút húsbavágó kérdés: lehet, hogy kevesebb alannal is jók lennének, ha hosszabb a vizsgálat, vagy ugyanannyi alannal pontosabb eredményt kapnánk, de így később lenne csak eredmény...!)
  - Challenge kísérlet?

# A p-érték

- A fentiek azt sugallják, hogy a szignifikanciaszinttel egy bináris döntést adjunk meg („hat” – „nem hat”)
- De így semmi nem derül ki abból, hogy mennyire erős az a döntés
- Ötlet: ha nem csak azt adjuk meg, hogy 5%-on hat-e, hanem azt is, hogy 5%-on hat-e és 1%-on hat-e, akkor gazdagítjuk az információt
- És 0,1%-on? Vagy ha 5-ön sem hatott akkor 10-en...? És így tovább
- Például: 82-ből 51 gyógyult meg

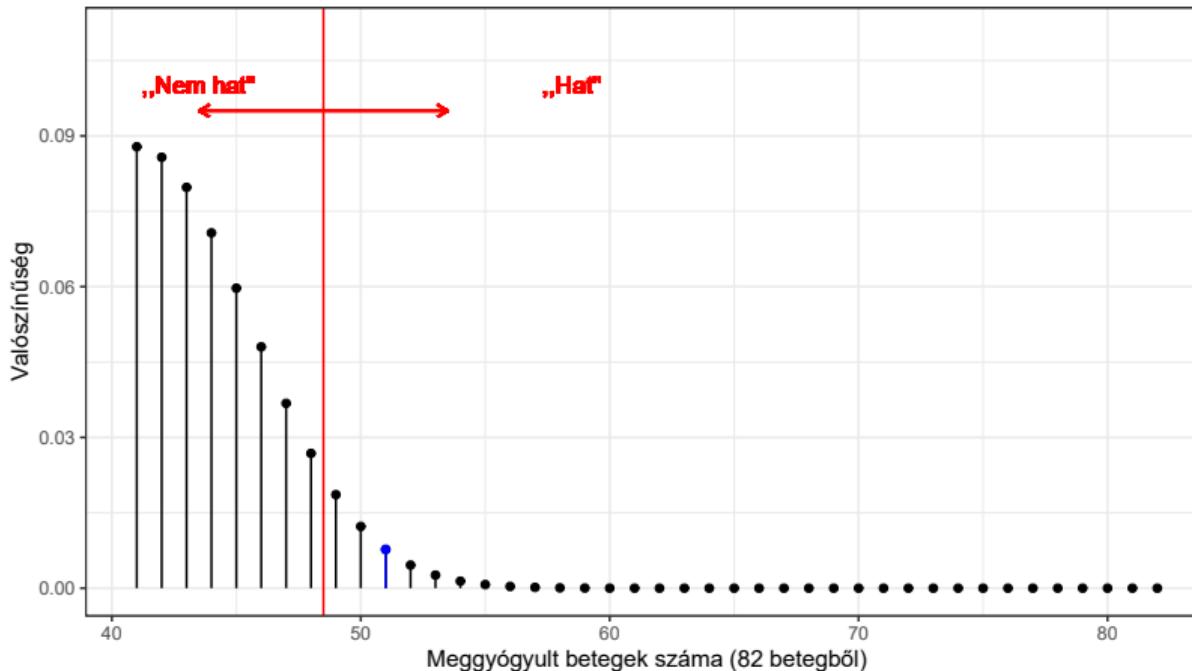
# A *p*-érték

10%-os szignifikanciaszint: túl magas



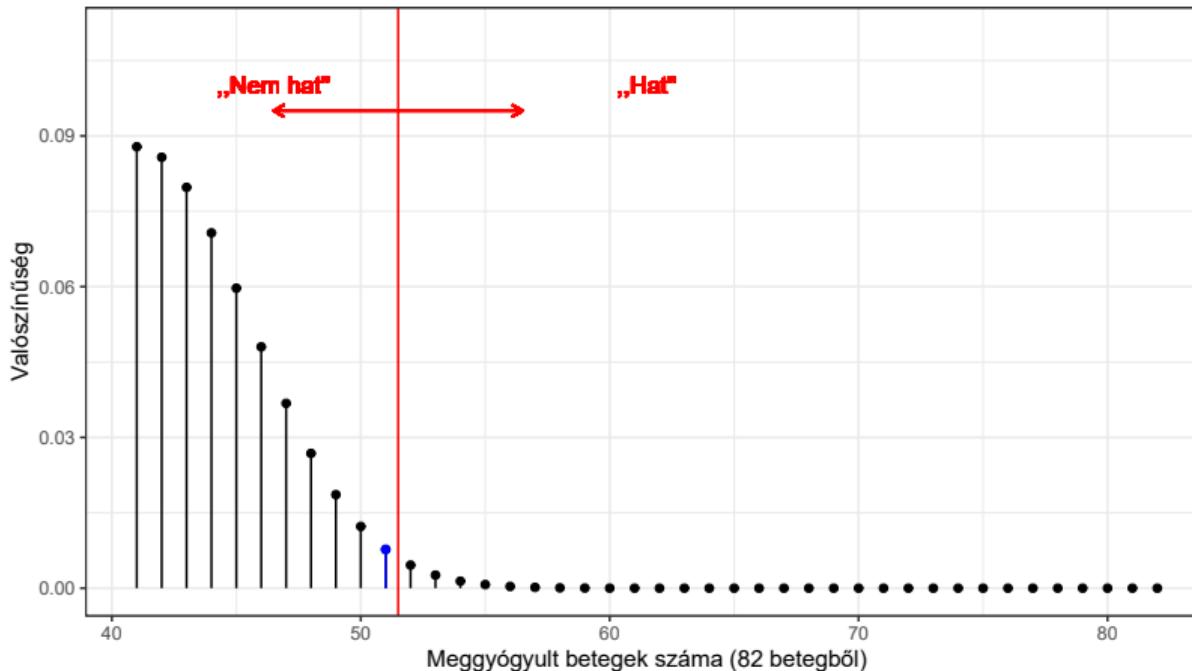
# A *p*-érték

5%-os szignifikanciaszint: még mindig kicsit magas



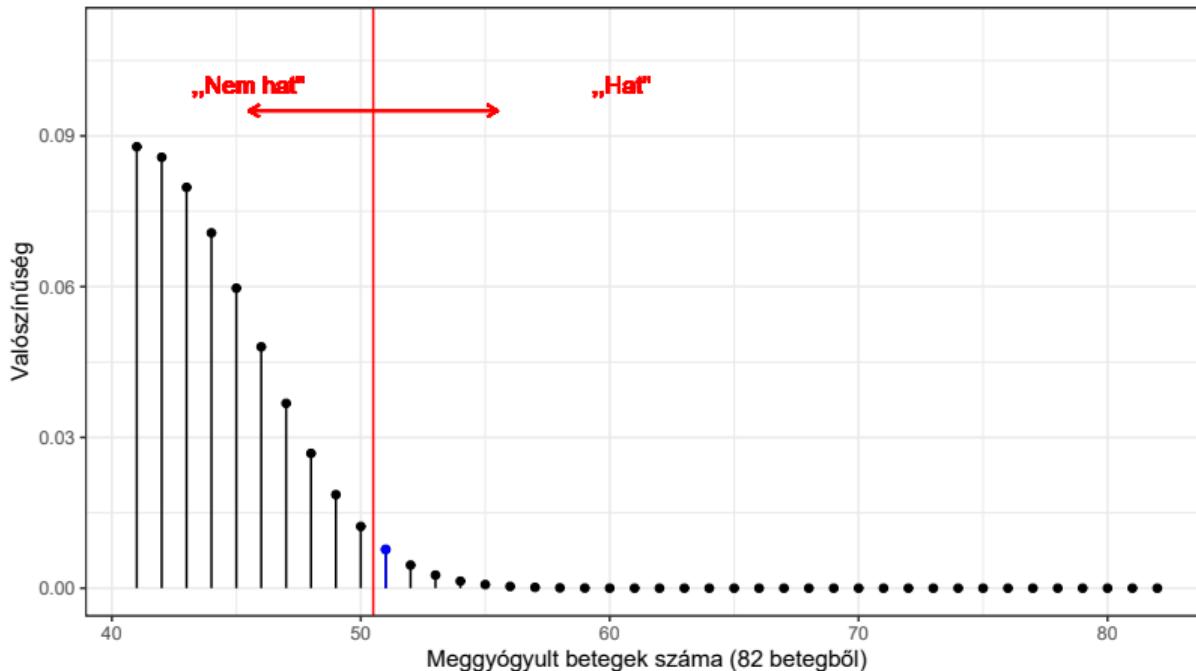
# A *p*-érték

1%-os szignifikanciaszint: hopp, ezzel már túllőttünk



# A *p*-érték

2%-os szignifikanciaszint: ez megint picit sok



# A *p*-érték

- A pontos érték (egy lehetséges, népszerű módszerrel – a diszkrétség miatt nem egyértelmű)  
 $p = 0,01762$
- Ez tehát az a szignifikanciaszint, ahol épp az elfogadás és az elutasítás határán vagyunk
- Azaz: ha ennél nagyobb szignifikanciaszintet választunk, akkor elutasítunk (hatásosnak minősítjük a gyógyszert), hiszen ekkor „ennyítettünk”, ha kisebbet, akkor elfogadunk (hatástalannak minősítjük a gyógyszert), mert ekkor „szigorítottunk”
- (Ezért szokták minden azt nézni, hogy a  $p$  kisebb-e mint 0,05, hiszen 5% a szignifikanciaszint szokásos értéke)
- Lényegében lehetővé teszi, hogy az olvasó tetszőleges, általa szimpatikusnak talált szignifikanciaszinten döntsön (ne csak a szerző által használton, amit jelentene az, ha csak a döntés van leírva adott szignifikanciaszinten)
- Egyúttal egyfajta mérőszáma is annak, hogy az adatok mennyire mondanak ellent a nullhipotézisnek (mennyire erősen bizonyítják, hogy a hatás nem pusztán a véletlen ingadozás miatti – természetesen a fordított logika értelmében!)
- Sajnos itt két dolog keveredik, a manapság használt kiértékelési módszertan több iskola nem teljesen konzisztens keveréke

# A mintavételeles helyzet

- A fentiek mind értelmezhetőek a mintavétel keretében
- Van egy nagy halmazunk, a sokaság, ez az amire a kérdésünk irányul
- Csak épp mi nem tudjuk az egészet megfigyelni, csak egy részhalmazát, ez a minta
- A statisztikának azt az ágát, ami azzal foglalkozik, hogy hogyan tudunk egyáltalán bármit mondani az egész sokaságról, ha csak egy részét ismerjük, szokás következtető statisztikának nevezni
- A valóságban igazából a sokaság végtelen lesz (és így nem is elemeivel adjuk meg, hanem eloszlásával), de szemléletesebb, és a lényeget ugyanúgy mutatja, hogy ha egy véges sok elemből álló halmazként gondolunk rá

# Egy szemléletes kép

- Egy kényelmes elképzelési mód az urnamodell:
  - Van egy nagyon-nagyon-nagyon nagy urnánk, benne a golyók valahány, de legkevesebb 50%-ára ráírva, hogy „túlélt”, a többire az, hogy „meghalt”
  - A klinikai kutatás az, hogy a jól megkevert urnából vakon kihúzunk 82 golyót
- Képzeletben megismételjük az egészet: visszadobjuk a golyókat, jól átkeverjük az urnát, és kezdjük előlről
- Ezt jó sokszor megismételjük: ekkor kapjuk az ábrákat, amik mutatták, hogy mekkora valószínűséggel lesz épp adott számú gyógyult

# Egy szemléletes kép

- Azt fogjuk mondani, hogy...
  - ...az urna a *sokaság* („a valóság”), amire kíváncsiak vagyunk, csak épp nem tudjuk közvetlenül megnézni – például 50%, ha hatástan a gyógyszer, 55, ha kicsit hat stb.
  - ...a kivett 82 golyó a *minta*, a szemüveg amin keresztül a sokaságot látjuk, csak épp ez nem mindig a valódi képet mutatja – jelen esetben az előbbi ábrán látott eloszlás
- A képzeletbeli újra-mintavétel szemlélete épp azért jó, mert érthetővé teszi ez utóbbit, hogy a mintából kapott érték ingadozik (noha a valódi érték adott, állandó – csak mi nem tudjuk, hogy mi)
- Úgy is fogjuk mondani: a mintából kapott érték mintavételeiről-mintavételre ingadozik – egyrészt attól függ, hogy mi a valódi érték, másrészt attól, hogy pont hogyan vettük a mintát (nem úgy értve, hogy bármilyen értelemben „rosszul” vettük a mintát)
- Ez utóbbit úgy hívjuk, hogy mintavéti ingadozás, az ebből fakadó hiba a mintavéti hiba
- Kikerülhetetlen, de a jellegéről tudunk nyilatkozni

# A következtető statisztika és ágai

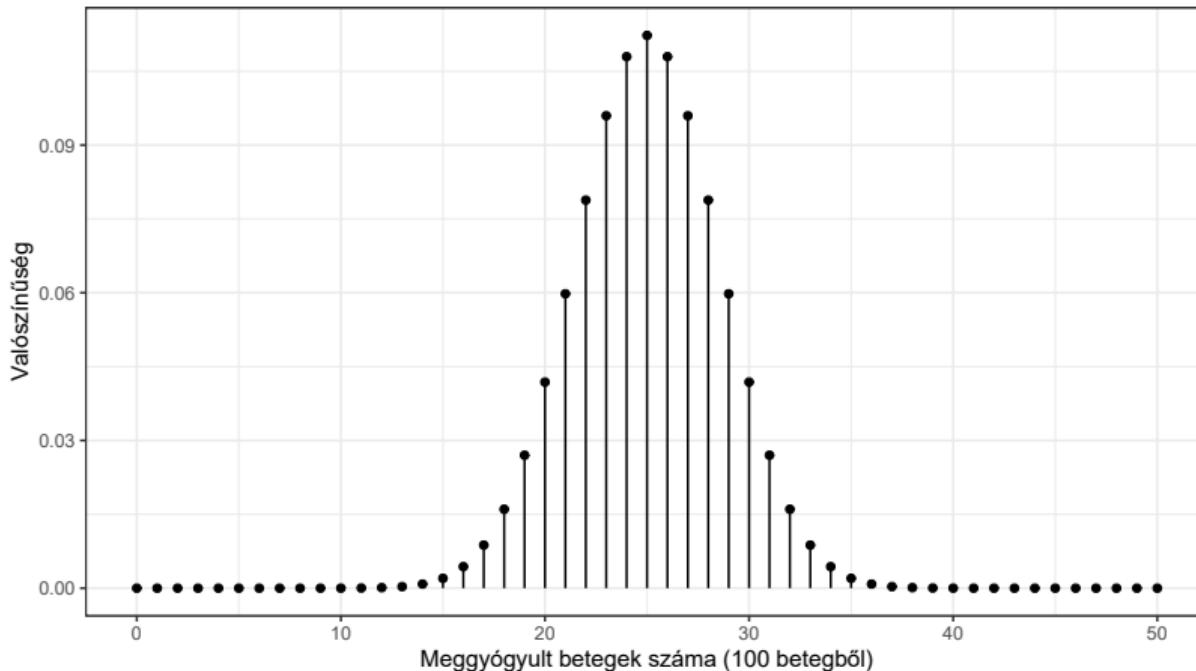
- Amivel eddig foglalkoztunk az „igaz-e, hogy hatástanlan a gyógyszer?” (értsd: „igaz-e, hogy a gyógyszer melletti gyógyulási arány 50%?”) jellegű kérdések megválaszolása volt
- Ezt az ágát a következtető statisztikának szokás hipotézisvizsgálatnak is nevezni
- Érdekes még a „mennyi a gyógyszer melletti gyógyulási arány?” típusú kérdések megválaszolása
- Ezt hívjuk becsléselméletnek
- (Persze a kettő szorosan összefügg)

# Becsléselmélet

- Egyrészt igyekszünk a legjobb „tippek” (becslést) adni
- Arra, hogy mit értünk legjobb alatt, adhatóak formális matematikai kritériumok
- Itt is igaz, hogy biztos válasz nincs
- Például, ha 50-ből 30-an gyógyultak meg, akkor csakúgyan jó becslés lesz a 60% a gyógyulási arányra
- (Jó az előbb említett precíz matematikai értelemben is!)
- De nem lesz biztos válasz, hiszen...

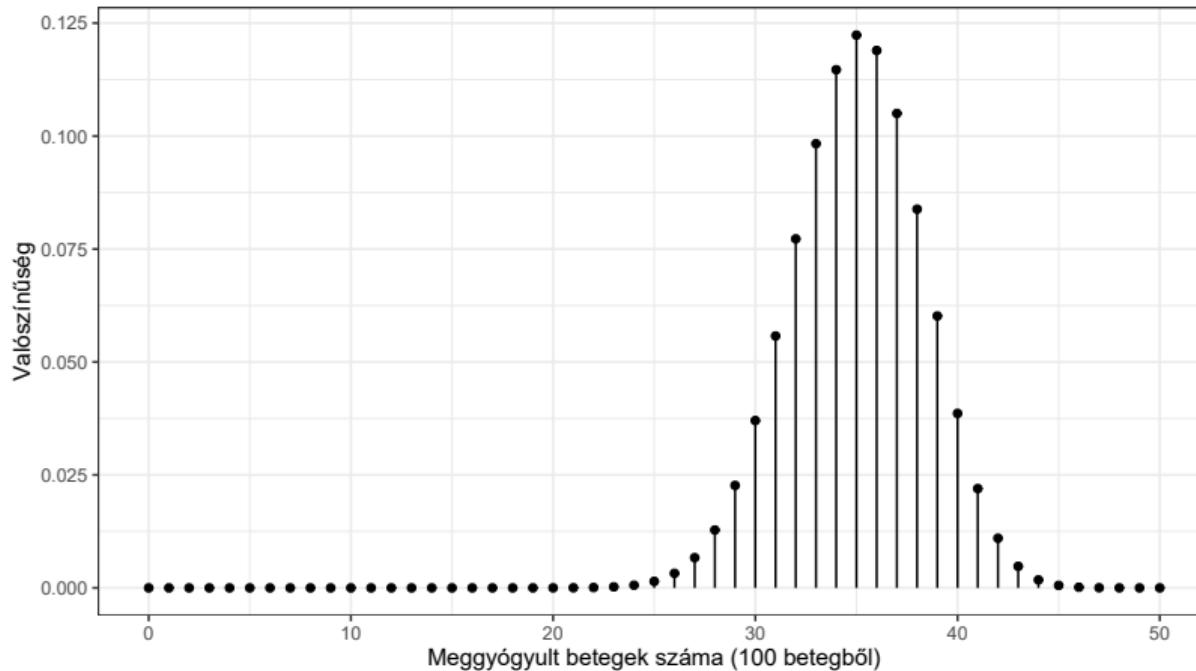
# Becsléselmélet

...a 30 kijöhét akkor is, ha a valódi gyógyulási arány 50%



# Becsléselmélet

...de akkor is, ha épp 70%



# Konfidenciaintervallum

- E bizonytalanságot tipikusan a konfidenciaintervallummal (CI) adjuk meg
- Mik azok a valódi értékek, amitől a mintában tapasztalt érték eltérése nem szignifikáns
- Itt is érvényesül a fordított logika: nem azt jelenti az intervallum, hogy ha ezt tapasztaltuk a mintában, akkor nagy valószínűsséggel hol van a valódi érték, hanem, hogy mely valódi értékek esetén kaphatnánk még elfogadható valószínűsséggel azt, amit ténylegesen kaptunk is
- Ezt fejezi ki a megbízhatósági szint (tipikusan 95%)
- „Mi kompatibilis a mintával”

# Mintavételi hiba (variancia)

- A mintavételi hibára azt szokták mondani: *variancia* típusú hiba
- A becsült érték eltérhet a valóságtól, de ha sokszor megismételnénk a mintavételt, akkor *átlagosan* nem lenne eltérés: ingadozik ugyan az eredményünk, de legalább a jó érték körül
- ...és minél nagyobb a mintanagyság, annál kevésbé
- Tehát ez a fajta hibázás nem bináris, nem arról van szó, hogy van ilyen vagy nincs, hanem számszerűen lemérhető: *mennyire* ingadozik a mintából kapott érték annak átlaga körül (pl. mekkora a szórása: ez az ún. standard hiba)
- Azt, hogy milyen kicsi ez a variancia szokták a vizsgálat precizitásának is nevezni
- Jól bevált apparátussal – következtető statisztika: *p*-értékek, konfidenciaintervallumok stb. – vizsgálható jellemző

# Nem-mintavételi hiba (torzítás)

- minden, ami nem a mintavételi ingadozásból fakad
- ...úgyhogy egyet már nagyon is láttunk eddig is: épp ilyen a confounding is!
- Egy másik tipikus példa: mi van, ha a mintát nem olyan tökéletesen választjuk (azaz nem lehet belőle következtetni a sokaságra)?
- Általános jellemzője az ilyen hibáknak, hogy még ismételt mintavétellel sem tartanak a nullába, azaz a mintából kapott értékeknek az átlaga sem a valódi érték
- Tehát: nem is a jó érték körül ingadozik az eredmény!
- Az ilyen típusú hibákat nevezzük *torzításnak* (lemérhető az átlag és a valódi érték különbségével, ennek neve az ellenség megtévesztése végett szintén torzítás)
- Sokszor keveset beszélnek róla az orvosi gyakorlatban, pedig rettentő fontos tud lenni...

# Variancia és torzítás

- Nagyon nagy általánosságban beszélve a torzítástól félünk jobban
- Lehet a kettőt együtt is mérni: mennyire ingadozunk a jó érték körül
- Ebben benne van mindkét komponens: mennyire ingadozunk az átlag körül (variancia) és az átlag hol van a jó értékhez képest (torzítás)!
- Neve átlagos négyzetes hiba (belátható, hogy értéke a torzítás és a variancia négyzetének összege)

# Hibák vizsgálata

- Mindezek persze elméleti kérdések, olyan értelemben, hogy egyetlen minta alapján úgysem tudjuk megmondani, hogy a hibázás variancia vagy torzítás-e
- Ezért is mondtuk, hogy „képzeletbeli ismételt mintavételek” kontextusában értelmezhető
- Szebben szólva ezek vizsgálatához mind valamilyen modellt kell feltételeznünk

# A nem-mintavételi hibák szerepe

- Ki szokta úgy érezni, hogy egy adott gyógyszergyár által fizetett kísérletben minden pont az Ő gyógyszere lesz a jobb? (Pl. lelkes kutatók végignézték a 4 vezető pszichiátriai újság 1992-2002 közötti összes cikkét: amikor a kísérlet szponzora a vizsgált gyógyszer gyártója volt, akkor 78% volt a sikerarány, függetlennél 48%, kompetítör által fizetett 28%...)
- De ez meg hogyan lehet egyáltalán? (hiszen ugyanaz a gyógyszer valakinek a saját terméke, valakinek meg a kompetítóré!)
- „Miért jobb az olanzapin mint a riszperidon, a riszperidon mint a kvetiapin és a kvetiapin mint az olanzapin?”
- Ezek tipikusan nem véletlen ingadozások, nem is direkte meghamisított eredmények, hanem a – többé vagy kevésbé jóhiszemű – nem-mintavételi hibák eredményei!

Kelly RE Jr, Cohen LJ, Semple RJ, et al. Relationship between drug company funding and outcomes of clinical psychiatric research. *Psychol Med.* 2006 Nov;36(11):1647-56. Heres S, Davis J, Maino K, et al. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry.* 2006 Feb;163(2):185-94.

# A mintavételi hibák kézbentartása következtető statisztikával

- Fontos már most hangsúlyozni, hogy a mintavételi hibák kezelésére bevált apparátus ( $p$ -érték, konfidenciaintervallumok stb.) az égegyadta világban semmit de semmit nem mond a nem-mintavételi hibákról!
- (Miközben simán lehet, hogy egy orvosi kutatás domináns hibaforrását ez jelenti!)

# Összefoglalva

Mintavételi hiba (elkerülhetetlen de kézbentartható, variancia típusú hiba, mintanagysággal csökken) vs. nem-mintavételi hiba (elkerülhető lehet de nehezen karakterizálható, torzítás típusú hiba, mintanagysággal nem feltétlenül csökken)

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés
- Az empirikus orvosi kutatások alapgondolata és a kauzalitás
- A confounding problémája
- A confounding megoldásai: megfigyelés és kísérlet
- Okozati következtetések levonása megfigyeléses adatokból
- A véletlen szerepéről
- **Bizonyitekok hierarchiája és összessége**
- A végpont és meghatározása

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

# A „bizonyítékok összessége” szemlélet, bizonyítékok aggregálása

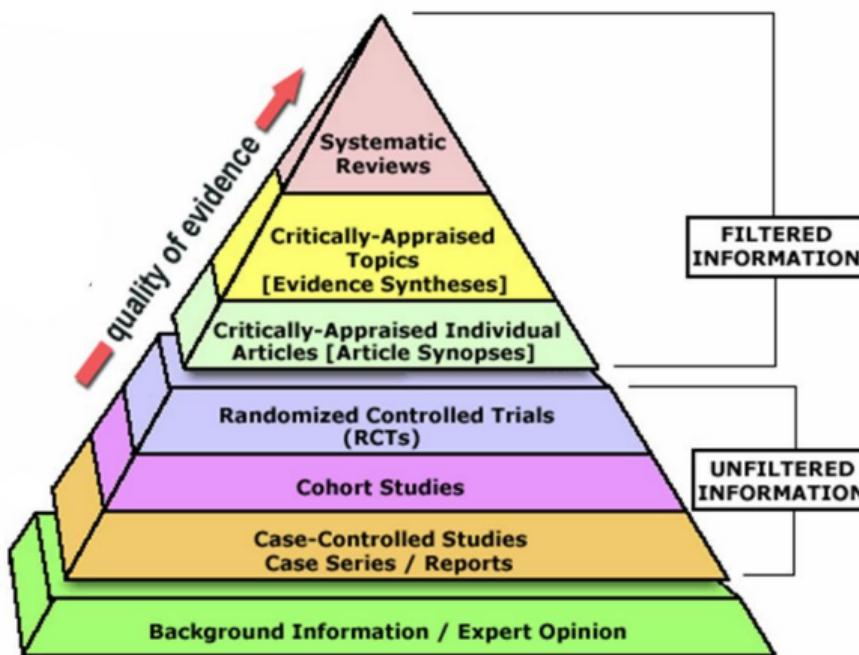
- Ahogy egy tanulmányban szereplő alanyok is ingadoznak a tanulmány valódi értéke körül, úgy a tanulmányok eredményei is ingadoznak „az” eredmény körül
- Különösen, ha figyelembe vesszük, hogy hány kutatás készül manapság, mondhatjuk, hogy *legalább* egy tanulmányt ma már mindenre és az ellenkezőjére is lehet két perc alatt találni...
- „New study shows” – nem az az érdekes, hogy az új kutatás mit mond, hanem az, hogy a kutatások összessége mit mond!
- Ezért a valós képhez a tanulmányokat össze kell – lehetőleg teljeskörűen – gyűjteni (*szisztematikus review*)...
- ...és az eredményeiket – matematikai úton – aggregálni, szintetizálni (*meta-analízis*)
- A gyakorlatban úgyis jórészt bizalmi kérdés, hogy mit hiszünk el, de legalább a nyomon követhetőség (*transzparenzia*) megvalósul így

# Bizonyítékok hierarchiája, a bizonyítóerő

- Ne feledjük a bizonyítékok hierarchiáját: nem arról van szó, hogy valami nem bizonyíték, valami meg igen, a bizonyítóerő tér el, nem bináris a dolog...
- Általában a kísérletes bizonyíték a nagyobb bizonyítóerejű (a confounding elkerülése miatt), de egyszerű kísérletet is lehet rosszul végezni, és megfigyelést is lehet jól, másrészt ott vannak a kísérlet hátrányai (pl. kisebb mintaméret) is, ami miatt ha olyan a kérdés, hogy ez nagy súllyal esik latba, simán lehet rosszabb is
- Ráadásul nem csak ez a két kategória létezik: mindegyiken belül is vannak fokozatok
- Adja magát egy kézenfekvő – de néha leegyszerűsítő – továbbfejlesztés: a bizonyítékok hierarchiája

# Az EBM-piramis

Ezek a gondolatok vezetnek el a gyakran látott EBM-piramishoz (avagy a bizonyítékok szintjei):

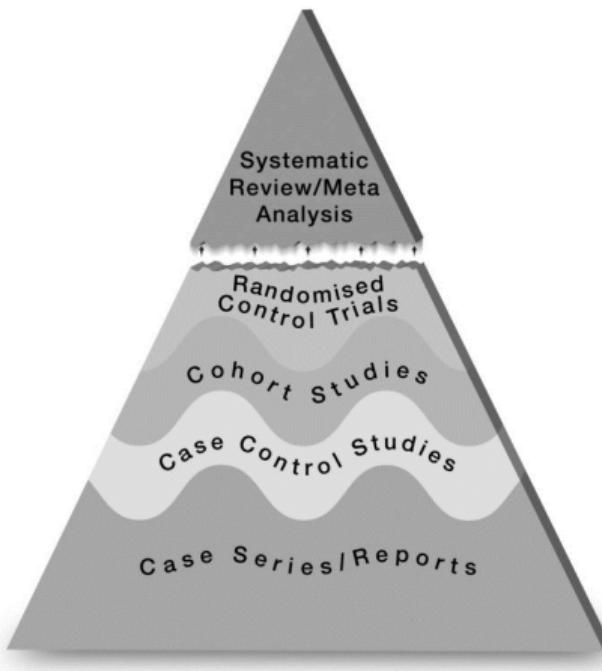


# Mik itt a problémák?

De az előbbi megjegyzések azt is mutatják, hogy ezt miért kezeljük fenntartásokkal/óvatosan!

- Főszabályként lehet, hogy igaz, de a bizonyítóerő akkor sem „osztálytulajdonság”
- minden egyes tanulmányt egyedileg kell értékelni, kritikusan, hiszen millióyi, piramisban meg sem jelenő tulajdonsága befolyásolja a bizonyítóerejét (a randomizálás megfelelő volt, milyen volt a kontrollálás, milyen volt a maszkolás stb. stb.)

# Javított piramisok



Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016 Aug;21(4):125-7. doi: 10.1136/ebmed-2016-110401.

# Javított piramisok



Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016 Aug;21(4):125-7. doi: 10.1136/ebmed-2016-110401.

# Tartalom

## 1 Az empirikus orvosi kutatások alapvonalai

- Pár történeti megjegyzés
- Az empirikus orvosi kutatások alapgondolata és a kauzalitás
- A confounding problémája
- A confounding megoldásai: megfigyelés és kísérlet
- Okozati következtetések levonása megfigyeléses adatokból
- A véletlen szerepéről
- Bizonyitekok hierarchiája és összessége
- **A végpont és meghatározása**

## 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek

## 3 Mintavételi hiba, következtető statisztika

## 4 A statisztikai modellek alapjai

## 5 Rendszerszintű problémák

## 6 Filozófiai kérdések

# Végpont

- Mi az, hogy végpont?
- Elsődleges vs. másodlagos
  - „Ami fontos” (klinikailag)
  - Amire a kísérlet ereje tervezve van!

# A végpontok jellegei

Mivel írható le? (feltételezzünk az egyszerűség kedvéért kétkarú vizsgálatot)

- Bináris (meghalt-e): arányt kapunk ( $p$ )
  - RR:  $p_T / p_C$
  - ARR:  $p_T - p_C$

# Bináris végpont egy problémája

- Még a legegyszerűbb esetekben is dilemmákhoz vezethet, egy példa: a kontrollcsoportban 2% kapott infarktust, a sztatinnal kezelt csoportban 1%, akkor most...
  - A fantasztikus gyógyszerünk 50%-kal csökkenti a kockázatot!
  - A fantasztikus gyógyszerünk nélkül 100%-kal nagyobb az infarktuskockázatod!
  - 100 emberből 98-at *feleslegesen* kezelünk, mert amúgy se kapna infarktust, 1-et *hiába* kezelünk, mert gyógyszerrel együtt is infarktust kap, és csak 1 az, akinél elérünk valamit (viszont közben minden a 100-at kitesszük a mellékhatás-kockázatoknak, minden a 100-zal kifizettetjük stb.)!
- Ugye mennyire máshogyan hangzik? Pedig csak osztani kell tudni, hogy lássuk: ez a három *igazából ugyanaz*!

## A „keretezési” hatás

- Malenka (1993): 10% eséllyel meghalsz kezelés nélkül, az A gyógyszer 80%-kal csökkenti ezt a kockázatot, a B gyógyszerrel 100 embert kezelve 8-at megment, melyiket választanád? → A-t 56,8%, B-t 14,7%, mindegy 15,5%, nem tudom 13,0% (a kutatásban betegek vettek részt)
- Bucher (1994): orvosok egy része relatív, más része abszolút formában kapta meg ugyanazon kutatás eredményeit; teljesen máshogy értékelték a gyógyszer jóságát
- Schwartz (1997): az abszolút és relatív információ megértése szorosan összefügg a matematikai alapkézségekkel
- Gigerenzer (2008): „100%-kal növeli a fogamzásgátló a trombóziskockázatot” (1 per 7000 kockázatnövekedés, csak akkora pánik lett belőle Angliában 1995-ben, hogy 13 ezer többlet-terhességmegszakítást hajtottak végre...)

# A „keretezési” hatás

- Gigerenzer (2008): „a mammográfia 25%-kal csökkenti az emlőrák-mortalitást” → laikusok több mint negyede szerint ez azt jelenti, hogy minden 1000 nőből 250-nel kevesebb hal meg emlőrákban, még nőgyágyasz szakorvosok 15%-a szerint is ez a helyzet

Malenka DJ, Baron JA, Johansen S, et al. The framing effect of relative and absolute risk. J Gen Intern Med. 1993 Oct;8(10):543-8. Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. BMJ. 1994 Sep 24;309(6957):761-4. Moynihan R, Bero L, Ross-Degnan D, et al. Coverage by the news media of the benefits and risks of medications. N Engl J Med. 2000 Jun 1;342(22):1645-50. Mason D, Prevost AT, Sutton S. Perceptions of absolute versus relative differences between personal and comparison health risk. Health Psychol. 2008 Jan;27(1):87-92. Schwartz LM, Woloshin S, Black WC, et al. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997 Dec 1;127(11):966-72. Gigerenzer G, Wegwarth O, Feufel M. Misleading communication of risk. BMJ. 2010 Oct 12;341:c4830. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, et al. Helping Doctors and Patients Make Sense of Health Statistics. Psychol Sci Public Interest. 2007 Nov;8(2):53-96.

## Az esély

- Esély (odds): „valószínűség osztva 1 minusz valószínűsséggel” avagy „bekövetkezés valószínűsége osztva a be nem következés valószínűségével” (pl. megbetegedők száma osztva a meg nem betegedők számával, nem az alanyok számával):

$$\text{odds} = \frac{p}{1 - p}$$

- Furcsának hathat, pedig a sportfogadásból ismert: tényleg ugyanarról van szó
- Az, hogy „3 az 1-hez” adják a hazai csapat győzelmét (3 az 1-hez az oddsa) ugyanaz, mint hogy 25% a valószínűsége
- És fordítva: ahelyett, hogy „10% holnap az eső valószínűsége” nyugodtan mondhatná azt az időjárásjelentés, hogy „9 az 1-hez az eső oddsa”
- Bármelyikből kiszámítható a másik ( $p = \frac{\text{odds}}{1+\text{odds}}$ ), a használatuk tehát teljes mértékben csak konvenció kérdése, hogy az időjárásjelentésnél ezt szoktuk meg, a sportfogadásnál meg azt

# Az esélyhányados

- A két esély hánnyadosa:

$$OR = \frac{p_T / (1 - p_T)}{p_C / (1 - p_C)}$$

- Tehát teljesen mint a relatív rizikó, csak „időjárásjelentős” szóhasználat helyett „sportfogadó” szóhasználattal kifejezve a valószínűséget
- Miért jó (azon túl, hogy bizonyos statisztikai modellek ezt szolgáltatják, és hogy eset-kontroll elrendezésnél csak ezt lehet kiszámítani)?
- Az RR nem szimmetrikus, sem arra nézve, hogy melyik irányból nézzük az adatainkat, sem arra nézve, hogy mi a kimenet

# Az RR aszimmetriája – hogy nézünk rá az adatokra

Ekcéma és szénanátha kapcsolata:

	Szénanáthás	Nem szénanáthás	Összesen
Ekcémás	141	420	561
Nem ekcémás	928	13525	14453
Összesen	1069	13945	15522

- Szénanátha relatív rizikója ekcéma szerint (ha ekcémás vagy, hányszorosára nő a szénanátha valószínűsége)  $\frac{141/561}{928/14453} = 3,91$ , ekcéma relatív rizikója szénanátha szerint (ha szénanáthás vagy, hányszorosára nő az ekcéma valószínűsége)  $\frac{141/1069}{420/13945} = 4,38$
- Viszont: szénanátha esélyhányadosa ekcéma szerint (ha ekcémás vagy, hányszorosára nő a szénanátha esélye)  $\frac{141/420}{928/13525} = 4,89$ , ekcéma esélyhányadosa szénanátha szerint (ha szénanáthás vagy, hányszorosára nő az ekcéma esélye)  $\frac{141/928}{420/13945} = 4,89!$
- ...miközben a két betegség ebből a szempontból nyilván teljesen szimmetrikus, egyik sem kitüntetett, csak a kapcsolatukra vagyunk kíváncsiak

Bland JM, Altman DG. Statistics notes. The odds ratio. BMJ. 2000 May 27;320(7247):1468.

## Az RR aszimmetriája – mit nevezünk kimenetnek

Tekintsük a következő egyszerű példát:

	Meghalt	Túlélő
Kezelt	25	75
Kontroll	50	50

- Ha a halálozás rizikóját nézzük:  $RR = \frac{25/100}{50/100} = 0,5$ , de ha a túlélés „rizikóját”:  $RR = \frac{75/100}{50/100} = 1,5$
- Halálozás esélyhányadosa  $OR = \frac{25/75}{50/50} = 1/3$ , a túlélésé:  $OR = \frac{75/25}{50/50} = 3$
- Tehát az esélyhányados szimmetrikus (ha megfordítjuk a kimenetet, megfordul – reciproka lesz – az OR), de a relatív rizikó *nem*:  $\frac{1}{0,5} = 2 \neq 1,5!$
- Nyilván teljesen mindegy, hogy a halálozást vagy a túlélést vesszük kimenetnek, hiszen az egyik úgyis meghatározza a másikat, de a relatív rizikó *mégis* függ ettől
- Nagyon zavaró módon *nem ugyanazt* kapom, ha azt számolom, hogy hányan haltak meg és ha azt, hogy hányan éltek túl (noha ez nyilván teljesen mindegy kellene legyen)!

Cummings P. The relative merits of risk ratios and odds ratios. Arch Pediatr Adolesc Med. 2009 May;163(5):438-45.

# További megfontolások ehhez

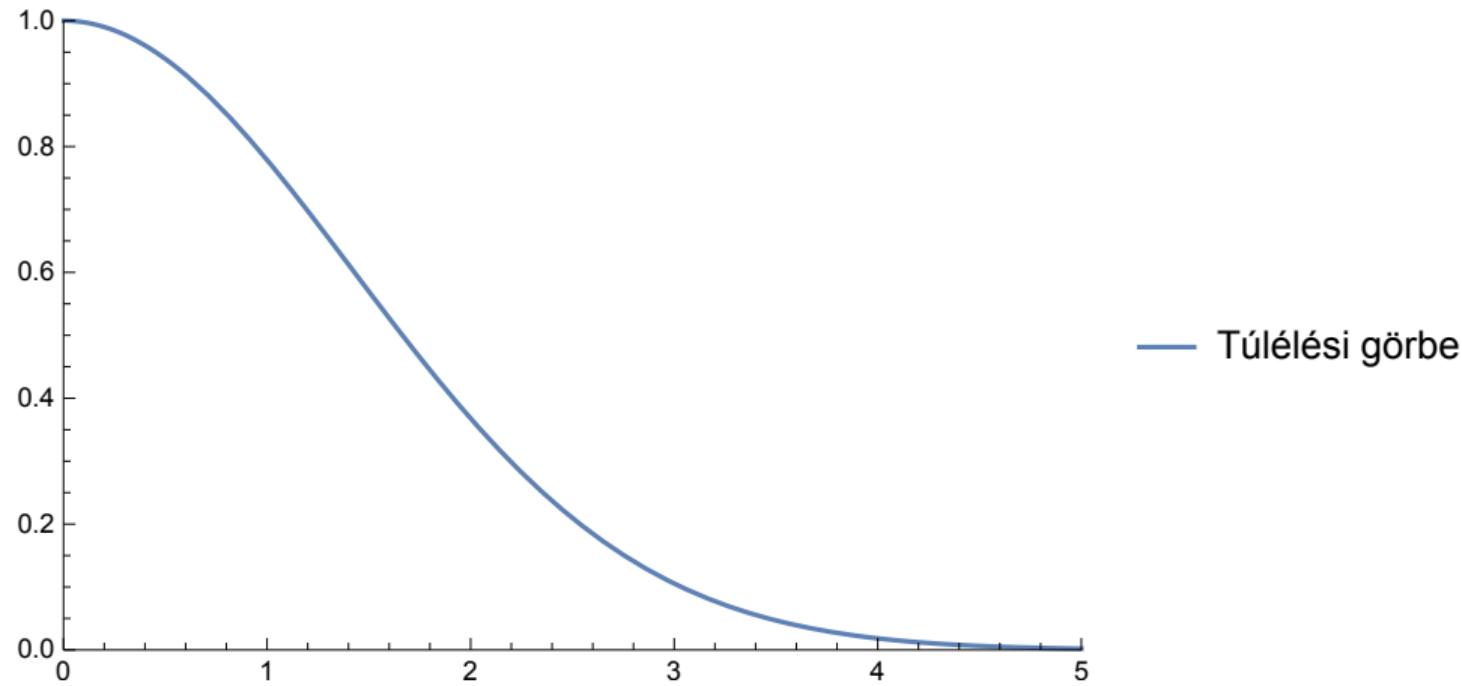
- Az OR nehezebben értelmezhető
  - „The only way, we are told, that physicians can understand probabilities: odds being a difficult concept only comprehensible to statisticians, bookies, punters and readers of the sports pages of popular newspapers.” (Stephen Senn)
- Az OR konstansabb (homogénebb) lehet
  - Egy 10%-os baseline rizikójú populáción valamilyen káros tényező hatása  $RR = 3$  – mi történik, ha ez a tényező egy 40%-os baseline rizikójú populációban hat?!
  - Az 1-nél nagyobb RR-ek matematikailag lehetetlen, hogy minden érvényesek tudjanak lenni! (Ellenérv:  $RR < 1$ -nél nincs ilyen baj, és az minden elérhető, legfeljebb megcseréljük a kimenetet)
  - OR-nél nincs ilyen baj, hiszen minden további nélkül lehet 1 fölé (a visszaszámolt valószínűség minden 0 és 1 között lesz)
- Az OR nem „összeönthető” (collapsible)
- A kettő közel van egymáshoz, ha a végpontok ritkák (mert ha  $p \rightarrow 0$ , akkor  $\frac{p}{1-p} \rightarrow p$ )

# A végpontok jellegei

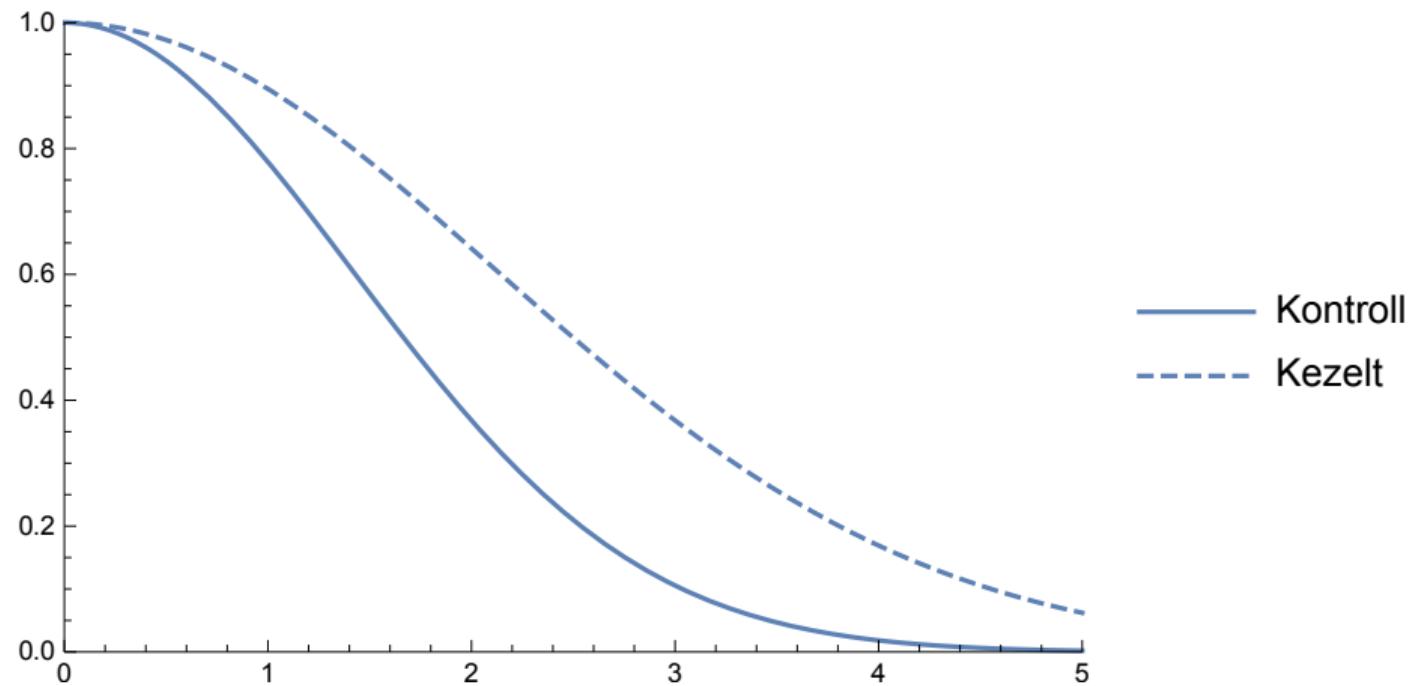
Mivel írható le? (feltételezzünk az egyszerűség kedvéért kétkarú vizsgálatot)

- Bináris (meghalt-e): arányt kapunk ( $p$ )
  - RR:  $p_T/p_C$
  - ARR:  $p_T - p_C$ , NNT:  $1/\text{ARR}$
  - OR:  $\frac{p_T/(1-p_T)}{p_C/(1-p_C)}$ , log OR, ésatöbbi
- Folytonos: eloszlásokat kapunk
  - Átlag/medián változott-e
  - Nagyon óvatosan a binarizálással (dichotomizálással)!
- Folytonos, de idő: túlélési görbéket kapunk
  - Adott idejű, például 1 éves, túlélés (ezzel lényegében binarizáljuk a végpontot)
  - Medián túlélés (szintén levágjuk, csak most „vízszintesen”)
  - Vagy: hazárd változott-e
  - Alternatív mérőszámok (RMST, cure modellek)

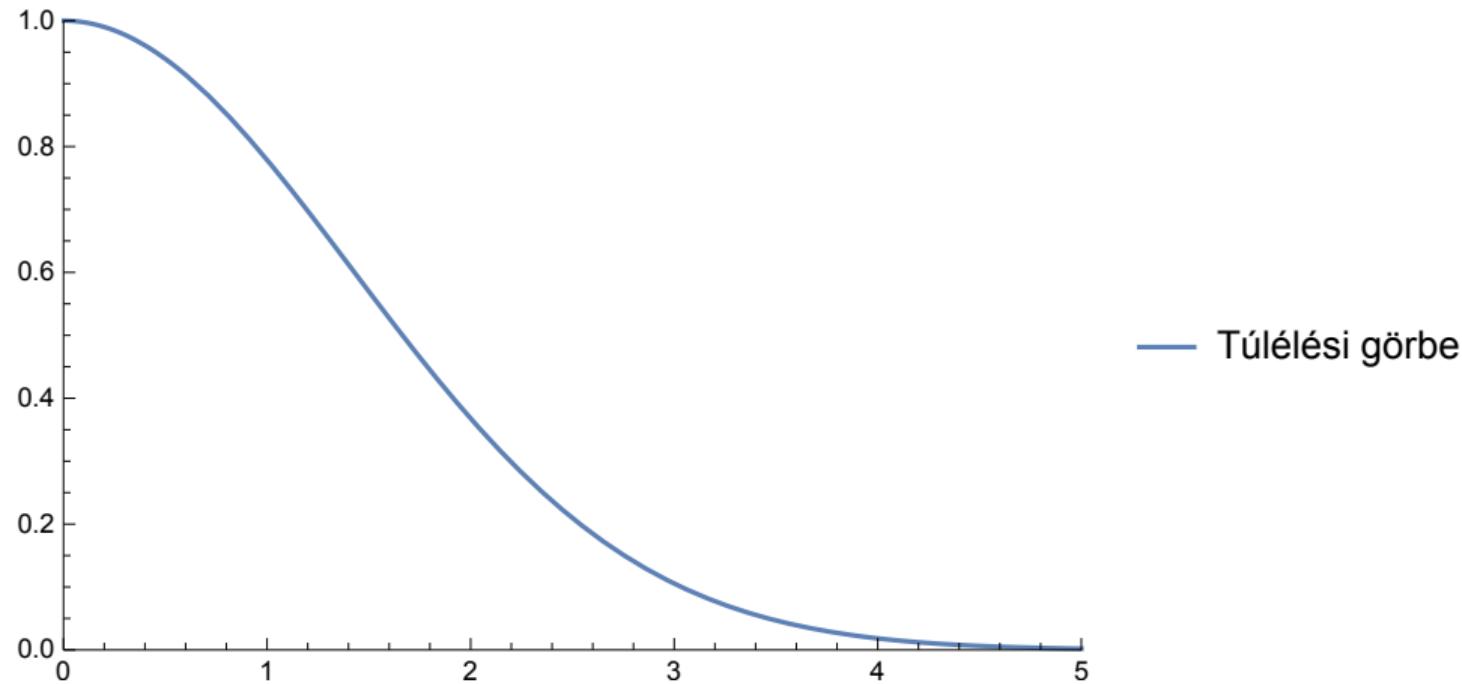
# Eltelt idő jellegű végpont: a hazárd fogalmának illusztrálása



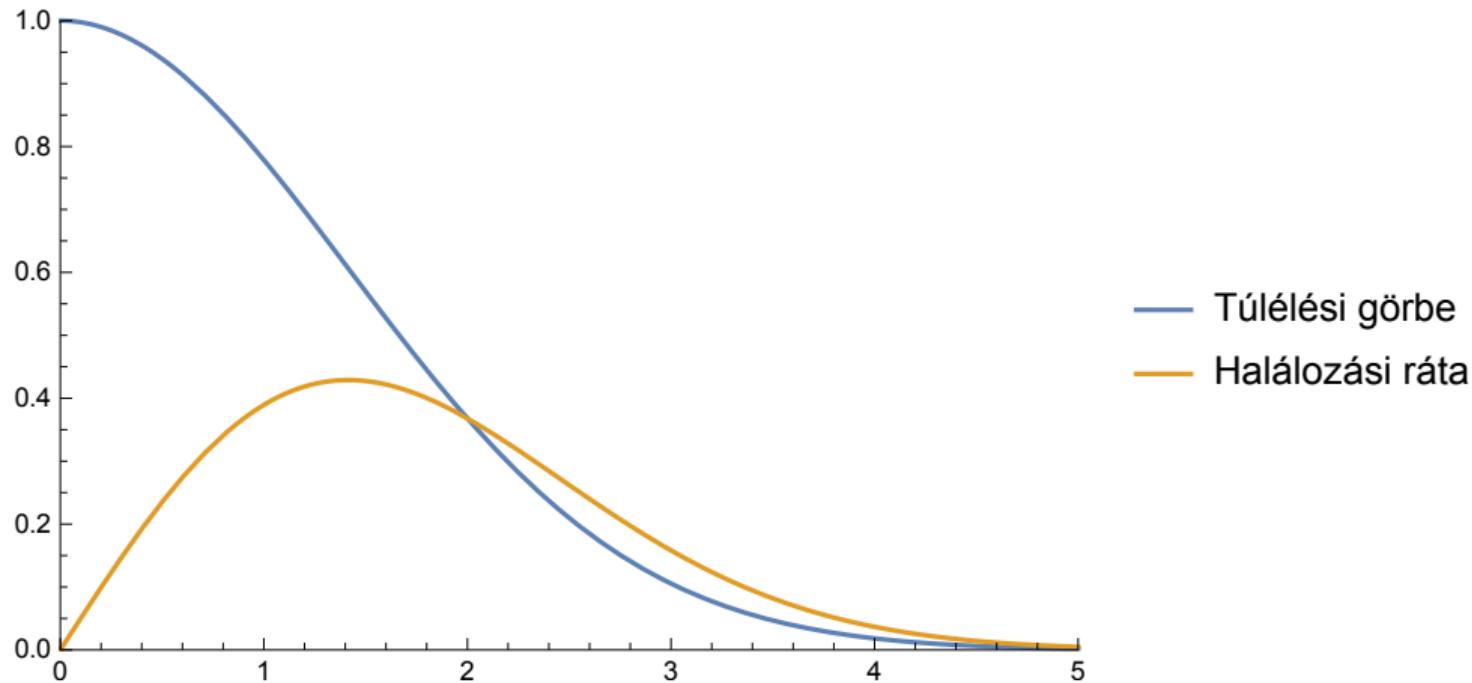
# Eltelt idő jellegű végpont: a hazárd fogalmának illusztrálása



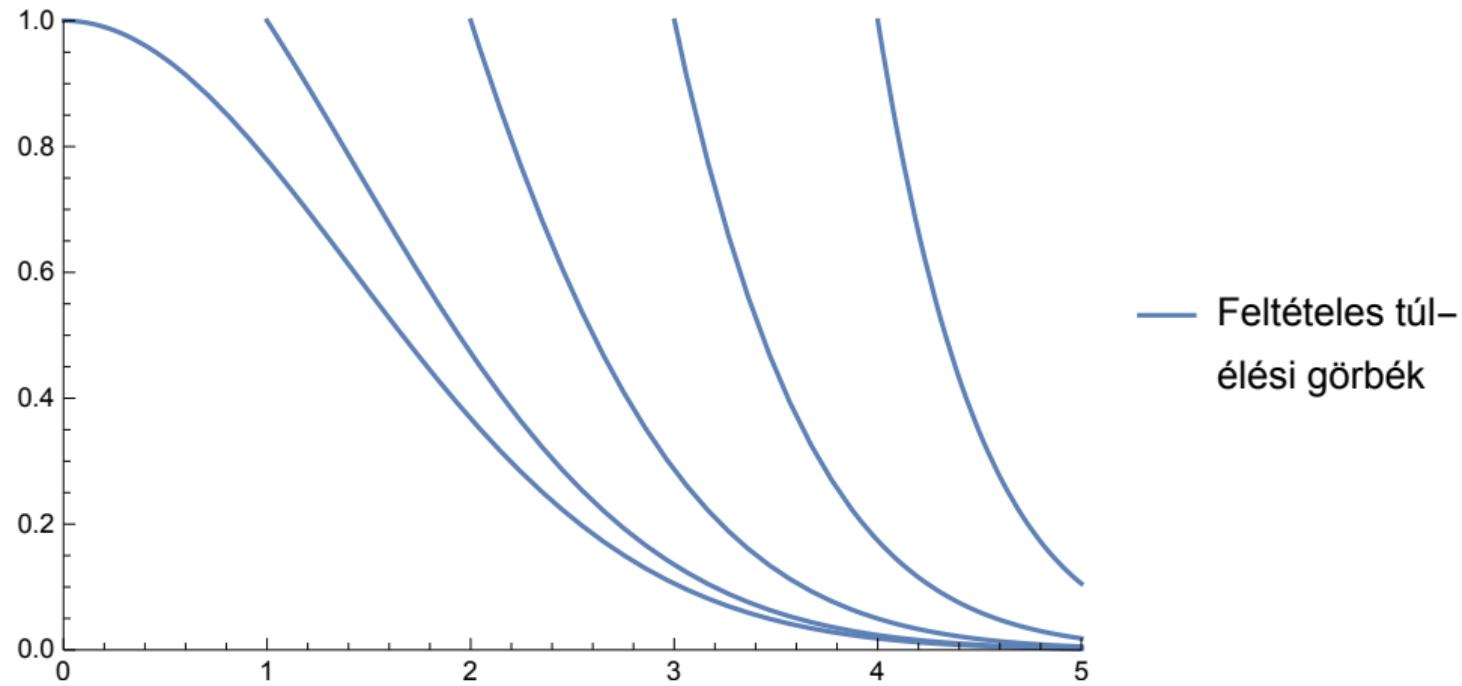
# Eltelt idő jellegű végpont: a hazárd fogalmának illusztrálása



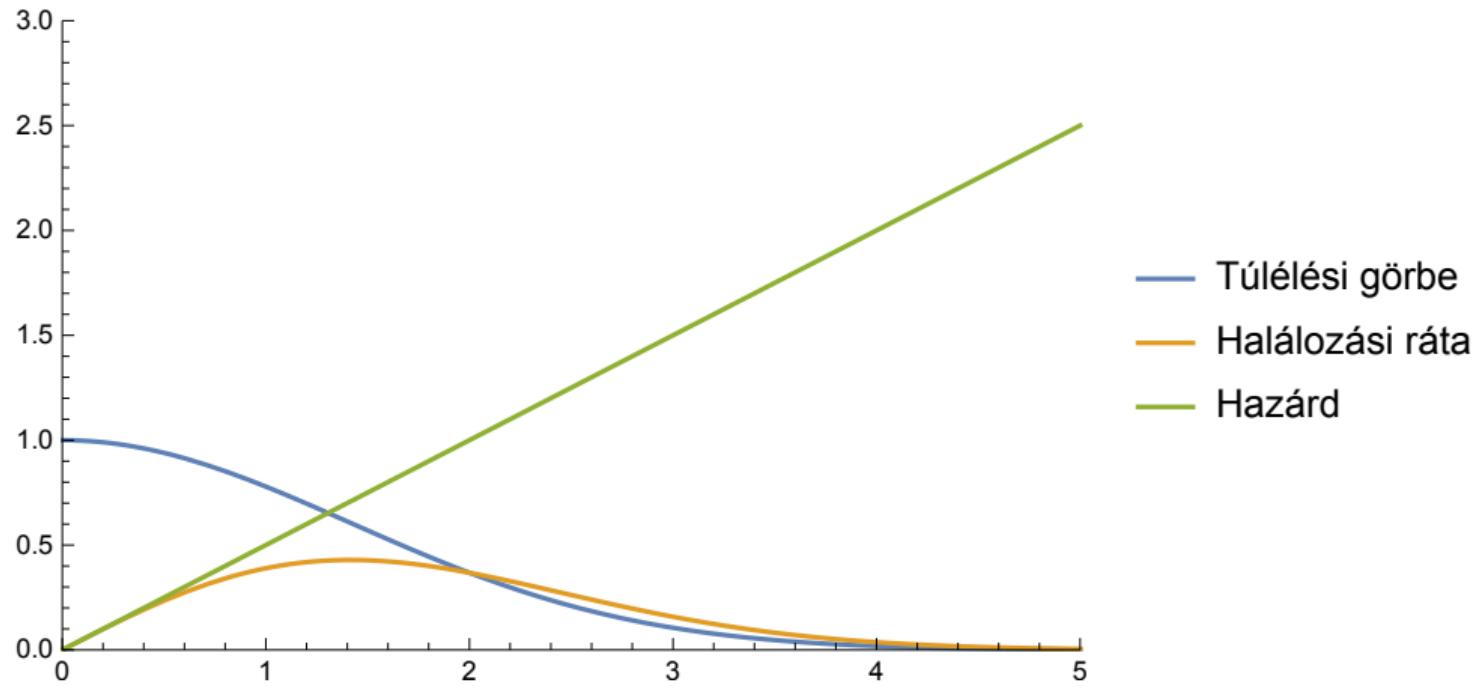
# Eltelt idő jellegű végpont: a hazárd fogalmának illusztrálása



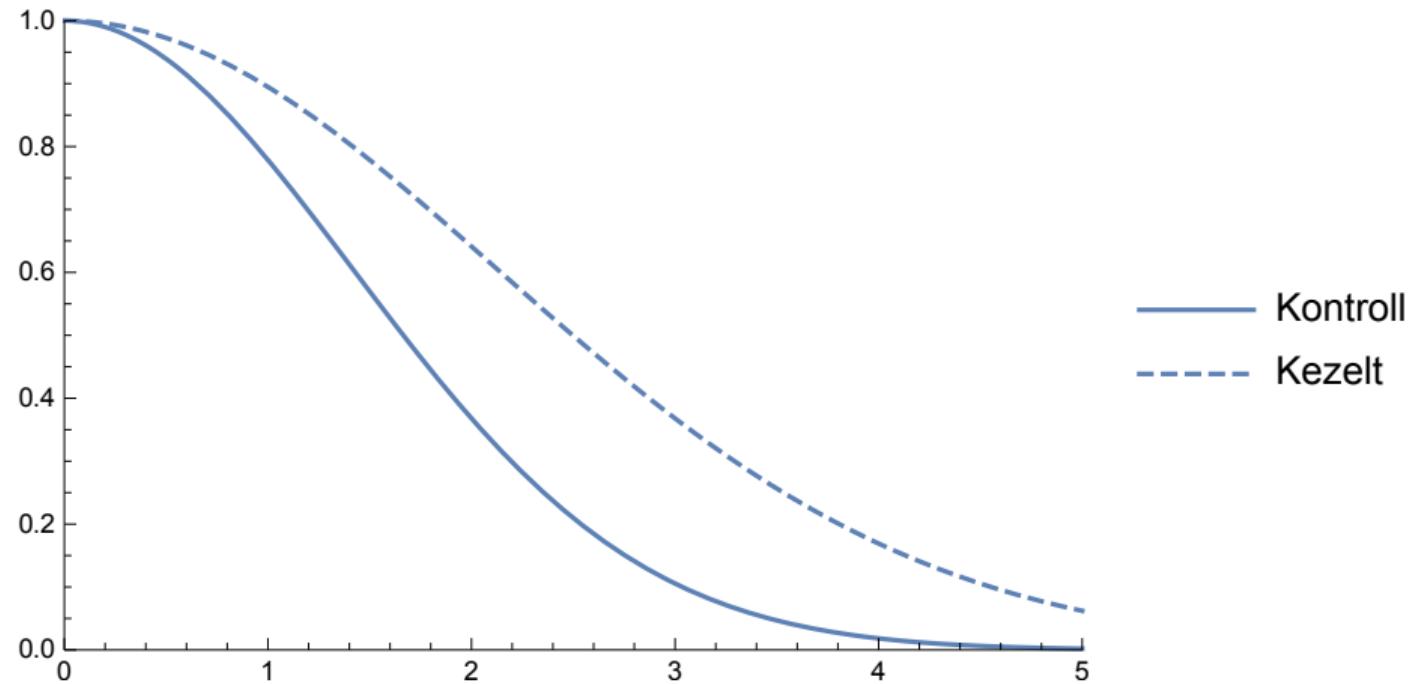
# Eltelt idő jellegű végpont: a hazárd fogalmának illusztrálása



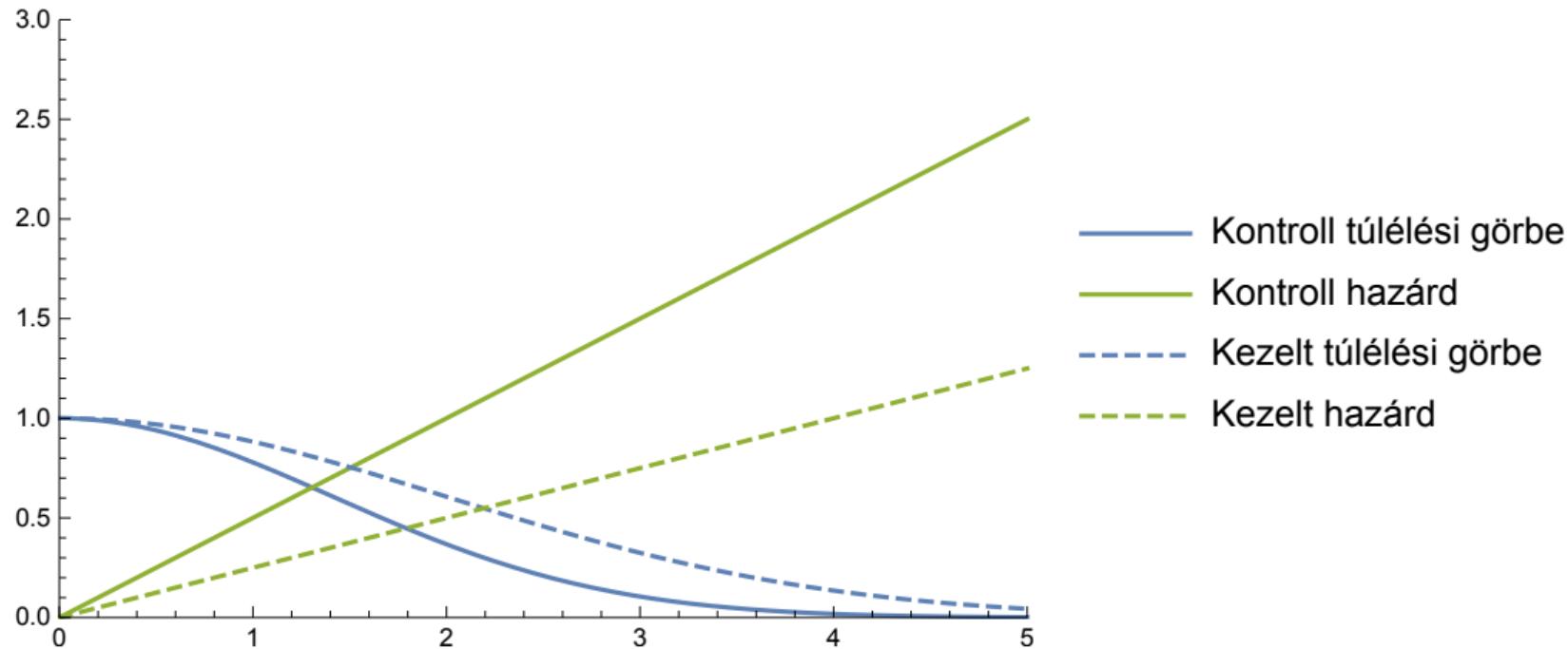
# Eltelt idő jellegű végpont: és akkor most jöjjön az összehasonlítás



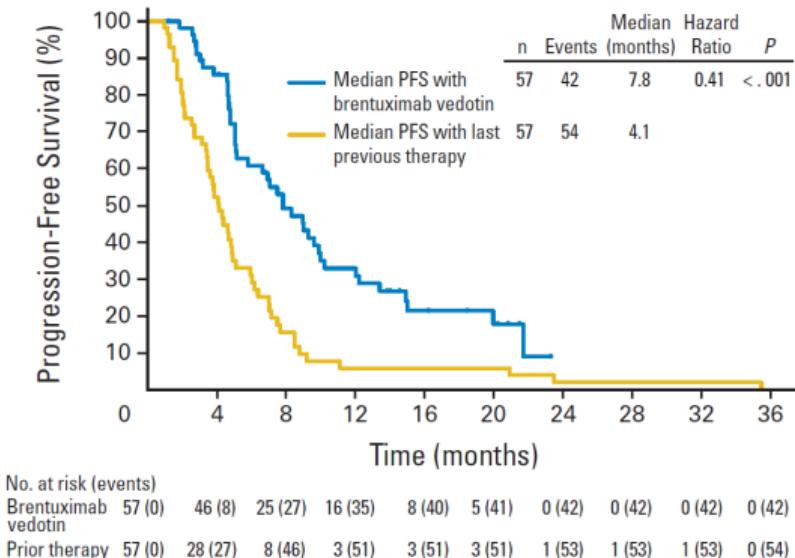
## Eltelt idő jellegű végpont: és akkor most jöjjön az összehasonlítás



## Eltelt idő jellegű végpont: és akkor most jöjjön az összehasonlítás



# Eltelt idő jellegű végpont: végül minden a gyakorlatban



Younes A, Gopal AK, Smith SE, et al. Results of a pivotal phase II study of brentuximab vedotin for patients with relapsed or refractory Hodgkin's lymphoma. J Clin Oncol. 2012 Jun 20;30(18):2183-9.

# A HR definíciója

- Hányszorosa a kezelt csoport hazárdja a kontrollcsoporténak?
- Na de melyik időpontban?
- Mindegy! (Mert minden időpontban ugyanannyiszorosa)
- Ezt hívjuk úgy, hogy a hazárdok proporcionálisak, a HR csak ilyen esetben értelmezhető egyértelműen
- (Az nem kötelező, hogy egyenes legyen a hazárd görbéje, lehet bármilyen alakú, csak a két csoport hánnyadosa legyen azonos minden időpillanatban)

# A HR interpretációja

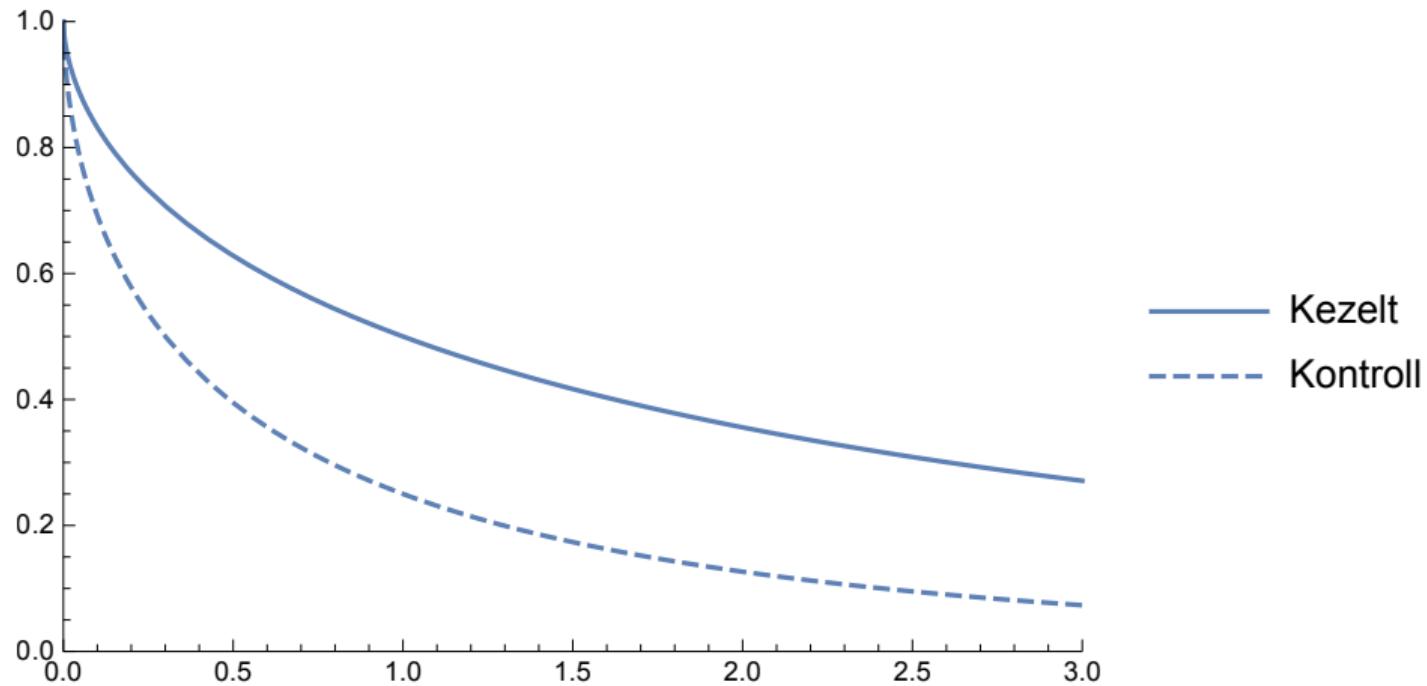
- Sokszor valamiféle „sebességgént” kezelik (első ránézésre érhető módon, mi is azt mondtuk, hogy „milyen gyorsan” fogynak a betegek), például ilyeneket mondva:
  - kétszer hamarabb vagy kétszer gyorsabban halnak meg kezelés nélkül a betegek...
  - kétszer annyian halnak meg adott időpontban...
  - kétszer annyian halnak meg adott időpontig...
  - kétszer valószínűbb, hogy meghalnak kezelés nélkül...
- Ezen megfogalmazások mindegyike hibás!
- A sebesség-analógia félrevezető: a sebesség értelmezhető egy konkrét autóra, de a „milyen gyorsan fogynak” megközelítés csak betegek egy csoportjára: az teljesen értelmezhető, hogy a betegek ezen csoportja ebben az időpontban 1%/nap gyorsasággal fogy, azaz hal meg (mert például 1 nap alatt 100-ból 1 halt meg), de ez egy konkrét betegre értelmetlen, hiszen ő nem tud 1 nap alatt 1%-ig meghalni...

Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. Antimicrob Agents Chemother. 2004 Aug;48(8):2787-92.

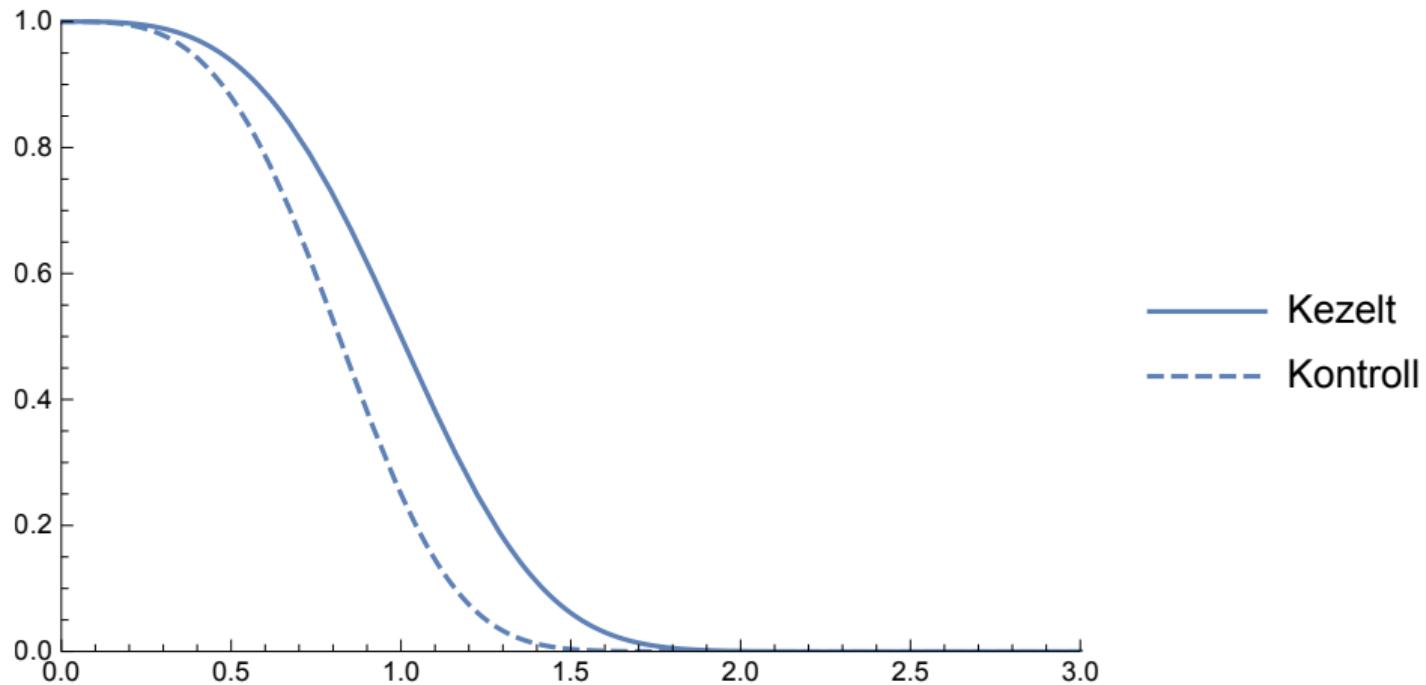
# A HR előnyei és hátrányai

- Felhasználja a teljes túlélési görbét (nem csak egy pontját, mint a fix idejű túlélési arány vagy a medián túlélési idő)
- De kérdés, hogy a proporcionalitás teljesül-e
- Figyelni kell az interpretációval, ahogy az előbb láttuk
- A HR relatív mutató, ráadásul még a túlélési görbe alakjától is függ: végeredményben a klinikailag releváns hatást egyáltalán nem biztos, hogy úgy jellemzi, ahogy azt intuitíve várunk

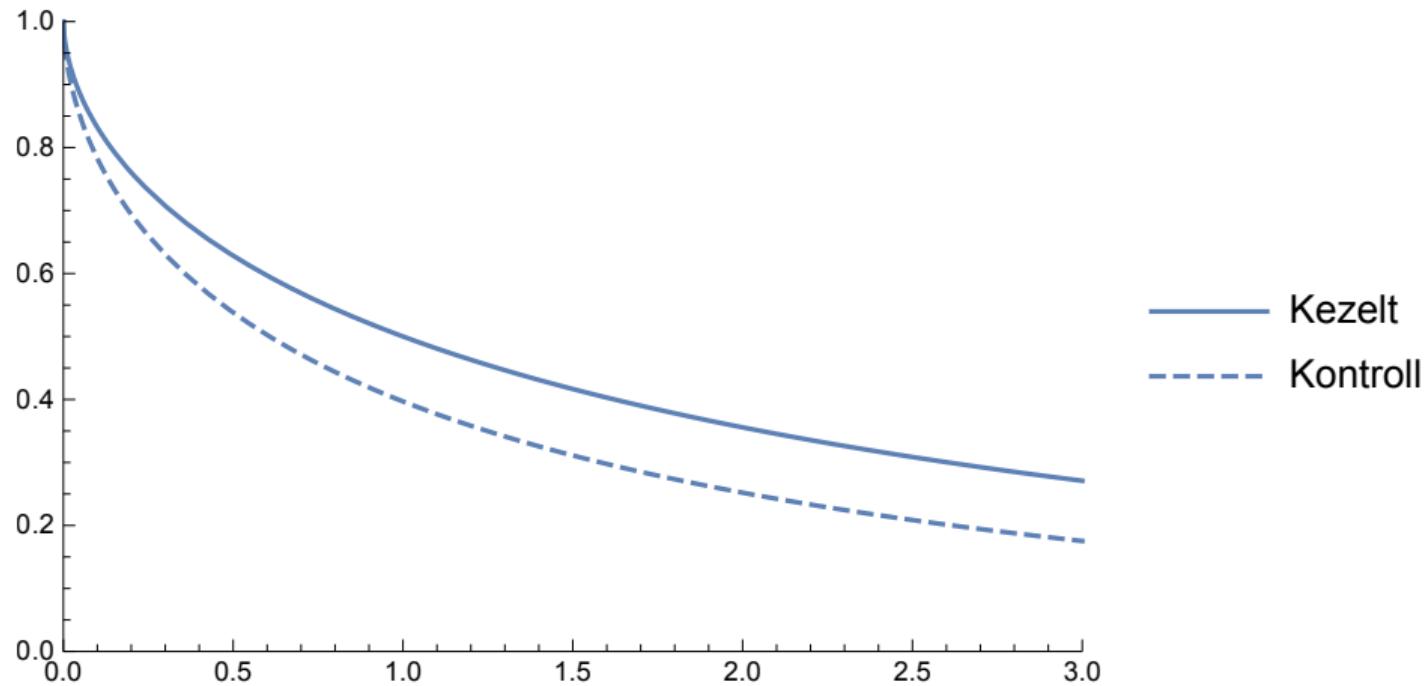
# Remek kezelés: HR=0,5



# Ugyanolyan (???) remek kezelés: HR=0,5



# Gyengécske kezelés: HR=0,75



# A HR mint a klinikai előny mutatója

- Biztos, hogy ezek így stimmelnek?!
- A HR sajnos itt még a túlélési görbe alakjától is függ
- Mindig érdemes abszolút metrikával kiegészíteni
- (A lesajnált fix idejű túlélési arány és a medián túlélési idő legalább az!)
- De ez sem problémamentes, lásd mindjárt...

# Egy általános gondolat a relatív és abszolút metrikákhoz

- Az abszolút metrika (ARR, medián túlélési különbség, 1 éves túlélési arány különbség stb.) felfogható származtatott mutatóként: a relatív metrika (RR, OR, HR stb.) és a baseline kockázat szorzata
- Bár esetről esetre meg lehetne vizsgálni, de *nagy általánosságban* az mondható, hogy
  - az abszolút metrikák fontosak a klinikai döntéshozatalhoz,
  - kutatásban viszont a relatívat érdemes lemérni, mert azok jobban átvihetők más populációkra is
- (Tehát a relatív eredmény stabilabb: lehet, hogy más populációban eltérő a baseline kockázat, de a tapasztalat az, hogy a relatív viszony *ilyenkor is* nagyon hasonló lesz – ezt többek között az is jól alátámasztja, hogy az alcsoporthanalízisekben általában elég konzisztensek szoktak lenni a relatív metrikák)
- Mi a baj a korábbi sztatistické példával?

Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses—sometimes informative, usually misleading. *BMJ*. 1999 Jun 5;318(7197):1548-51. Ferenci T. Absolute risk reduction may depend on the duration of the follow-up. *Expert Rev Clin Pharmacol*. 2017 Dec;10(12):1409-1410.

# A mutató stabilitásának kérdése: a sztatinok esete

Kísérlet neve	Kontrollcsoport rizikója	Relatív rizikó	Abszolút riziko-különbség	Utánkövetés hossza [év]
JUPITER	0,48%	0,81 (-19%)	0,09 %pont	1,9
AFCAPS/TexCAPS	0,76%	0,68 (-32%)	0,24 %pont	5,2
ASCOT-LLA	1,60%	0,90 (-10%)	0,16 %pont	3,3
WOSCOPS	2,22%	0,68 (-32%)	0,70 %pont	4,9
CARE	6,26%	0,86 (-14%)	0,87 %pont	5
HPS	9,13%	0,83 (-17%)	1,52 %pont	5
4S	9,31%	0,66 (-34%)	3,19 %pont	5,4
LIPID	9,62%	0,76 (-24%)	2,28 %pont	6,1
PROSPER	10,06%	0,86 (-14%)	1,38 %pont	3,2

# Az abszolút és relatív metrikák használata

- Azaz az általában jó stratégia: kutatásban relatívat mérni, de aztán azt a konkrét klinikai szituációban kontextusba helyezni
- A cél: „additive at the point of analysis but relevant at the point of application” (Stephen Senn)

Senn S. At odds with reality. <http://www.bmjjournals.org/rapid-response/2011/10/27/odds-reality>

# Egy illusztráció az abszolút és relatív metrikákhoz...

## ICS-LAMA-LABA vs. LAMA-LABA COPD-ben

The rate of moderate or severe exacerbations during treatment among patients assigned to triple therapy was 0.91 per year, as compared with [...] 1.21 per year among those assigned to umeclidinium–vitanterol (rate ratio with triple therapy, 0.75; 95% CI, 0.70 to 0.81; 25% difference; P<0.001).

- Szokásos sztori: a relatív („25%-os csökkenés”) jó lehet mint kísérletben kimérendő eredménye...
- ...de a klinikai döntéshozatalhoz az abszolút lesz a fontos!
- A másik fontos: a kontextusba helyezés, például a mi betegünknek a baseline kockázata is nagyobb (a vizsgálatban egészségesebb alanyok voltak!), mondjuk 1,6 exacerbáció/év
- *Feltéve, hogy a gyógyszer relatív hatása állandó, a tripla kombinációval  $1,6 \cdot 0,75 = 1,2$  exacerbációja lesz évente*

Lipson DA, Barnhart F, Brealey N, et al. Once-Daily Single-Inhaler Triple versus Dual Therapy in Patients with COPD. N Engl J Med. 2018 May 3;378(18):1671-1680.

# Egy illusztráció az abszolút és relatív metrikákhoz...

- A különbség  $(1,6-1,2=)$  0,4 exacerbáció/év
- Még kifejezőbb ha úgy fogalmazunk:  $(1/0,4=)$  2,5 évig kell egy beteget kezelnünk ezzel ahhoz, hogy egy exacerbációt megelőzzünk
- Biztos megéri 2,5 évnyi kezelés mellékhatásainak kitenni egy beteget egy exacerbáció megelőzése végett? Biztos megéri 2,5 évnyi kezelést kifizettetni egy exacerbáció megelőzése végett? (ezek persze már nem statisztikai kérdések, de a „statisztika” segít abban, hogy jól megfogalmazzuk/megértsük, hogy egyáltalán mi a kérdés!)

Lipson DA, Barnhart F, Brealey N, et al. Once-Daily Single-Inhaler Triple versus Dual Therapy in Patients with COPD. N Engl J Med. 2018 May 3;378(18):1671-1680.

## ...és egy – nagyon fontos – továbbgondolás

- A másik nagy előny ez abszolút metrikáknak, hogy megkönnyíti a különböző kimenetek egymással történő összevetését

### ICS-LAMA-LABA vs. LAMA-LABA COPD-ben

[T]he risk of clinician-diagnosed pneumonia was significantly higher with triple therapy than with umeclidinium–vitanterol, as assessed in a time-to-first event analysis (hazard ratio, 1.53; 95% CI, 1.22 to 1.92; P<0.001).

- A 25% exacerbáció rizikó csökkenés összehetetlen az 53%-os pneumoniabeli növekedéssel! (hiszen teljesen mások a baseline rizikók)
- Ha azonban jobban megnézzük a kéziratot, akkor láthatjuk, hogy az ICS-LAMA-LABA csoportban a pneumonia rizikója 95,8/1000 betegév, a LAMA-LABA csoportban 61,2/1000 betegév
- Ez lehetővé teszi, hogy abszolút metrikát számoljunk: a tripla kombináció ( $0,0958 - 0,0612 =$ ) 0,0346/évvel növeli a pneumonia rizikóját (mondjuk, hogy itt a betegünk baseline rizikója ugyanaz, mint a kutatásban)
- Azaz a szokásos további átszámítással: ( $1 / 0,0346 =$ ) 28,9 évnyi kezeléssel okozunk *egy többlet tüdőgyulladást*

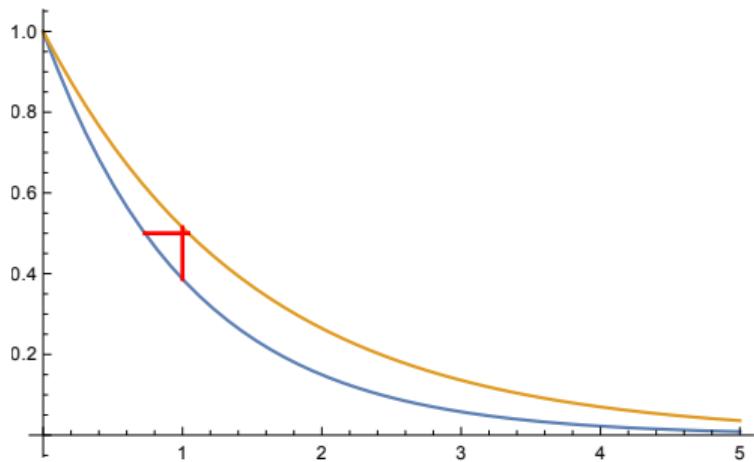
# Különböző végpontok összevethetősége

- És most már ordít, hogy miért beszélhetünk „összevethetőség megteremtéséről”: ez magyarár azt jelenti, hogy kb. 11,5 exacerbációt előzünk meg 1 tüdőgyulladás árán!
- Megint csak: az nyilván nem statisztikai kérdés, hogy ez megéri-e, „ér-e” 11,5 exacerbációt 1 pneumonia – de így legalább látjuk, mégpedig klinikailag értelmes, releváns módon, hogy egyáltalán mit kell összevetni! (a százalagos változásokból ez nem derült ki!)
- (Megjegyzés: út az exacerbáció és a pneumonia közös nevezőre hozására: életév-veszteség, minőséggel korrigált életév-veszteség...)

# Sajnos az abszolút mutatók sem problémamentesek

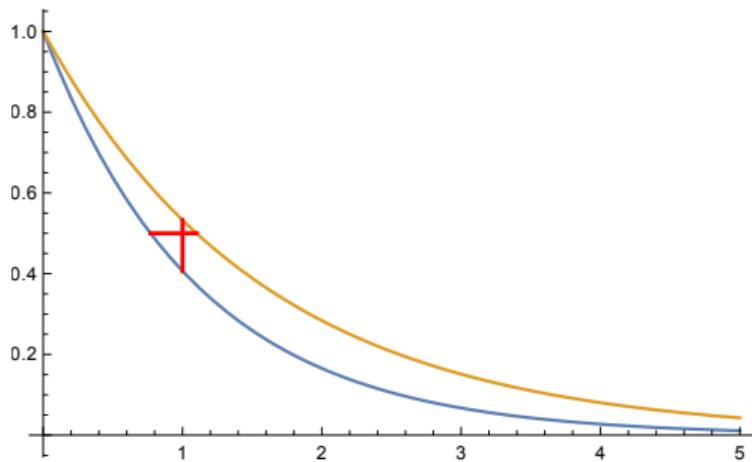
- Eltelt idő jellegű végpont esetén a két abszolút jellegű mutató (medián túlélés növekedés és fix – például 1 éves – túlélési arány javulás) sajnos *ellentmond egymásnak!*
- A következő ábrákon pontosan ugyanolyan a túlélési görbe alakja, pontosan ugyanúgy rögzítetten 0,7 a HR, az egyetlen különbség, hogy a kontrollcsoport milyen gyorsan hal (mekkora a baseline rizikó)

# Sajnos az abszolút mutatók sem problémamentesek



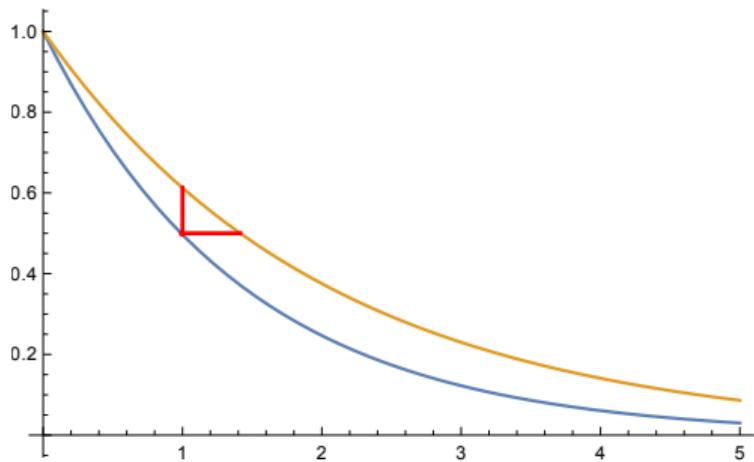
Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

# Sajnos az abszolút mutatók sem problémamentesek



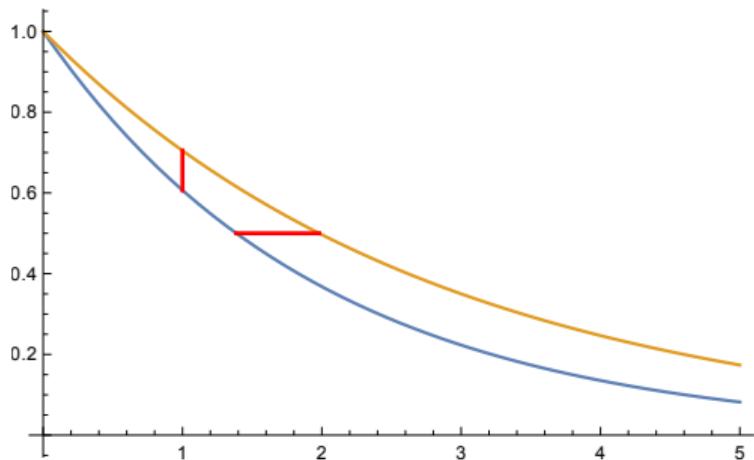
Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

# Sajnos az abszolút mutatók sem problémamentesek



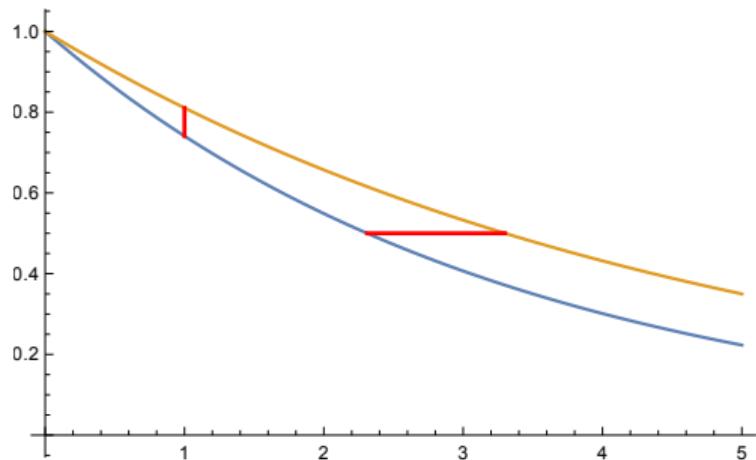
Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

# Sajnos az abszolút mutatók sem problémamentesek



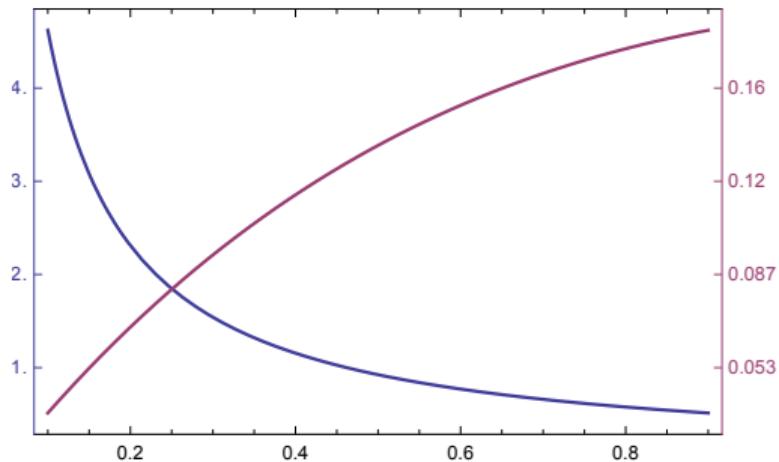
Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

# Sajnos az abszolút mutatók sem problémamentesek



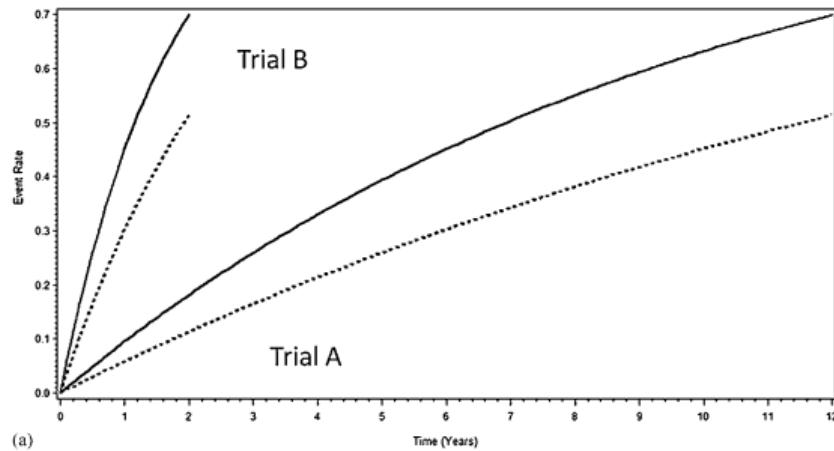
Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

# A két mutató ellentmondása



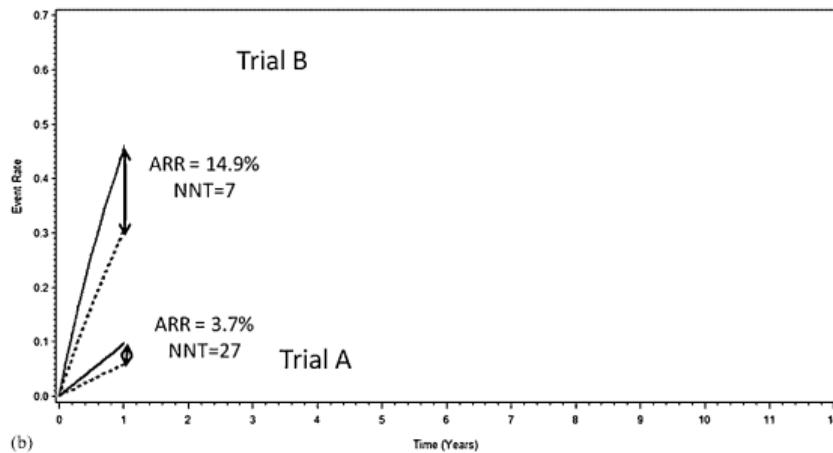
Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

# Az ellentmondás egy konkrét illusztrációja



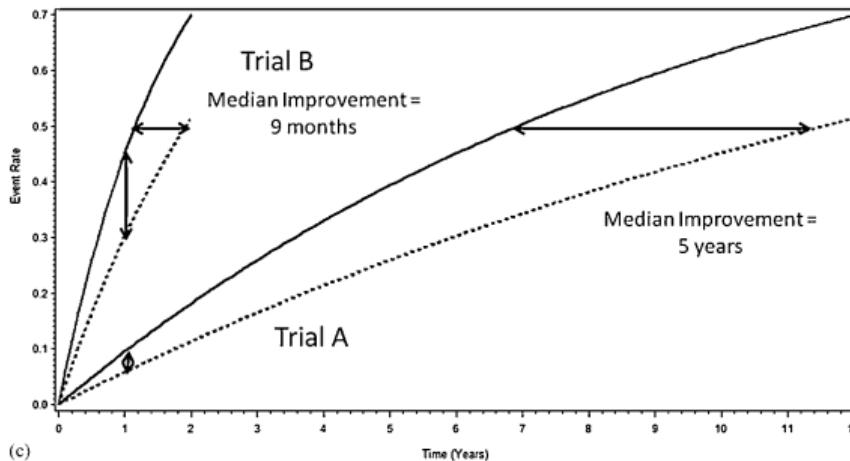
Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

# Az ellentmondás egy konkrét illusztrációja



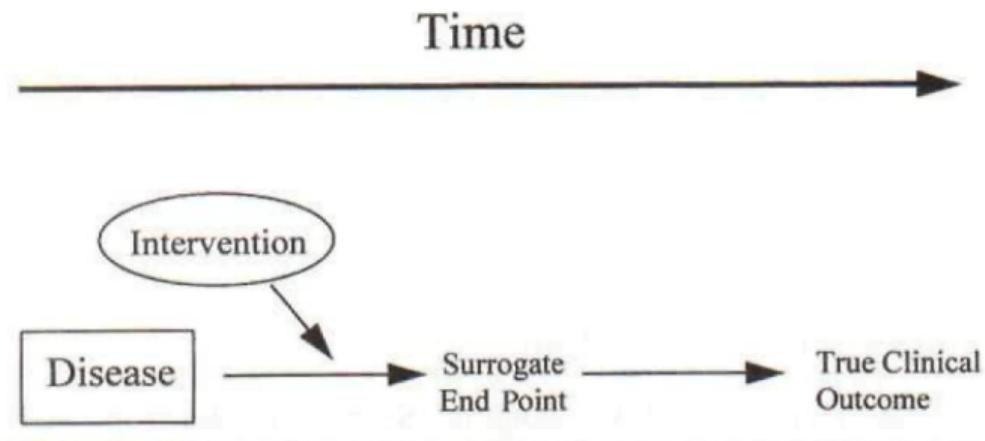
Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

# Az ellentmondás egy konkrét illusztrációja



Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat Med. 2011 Aug 30;30(19):2341-8.

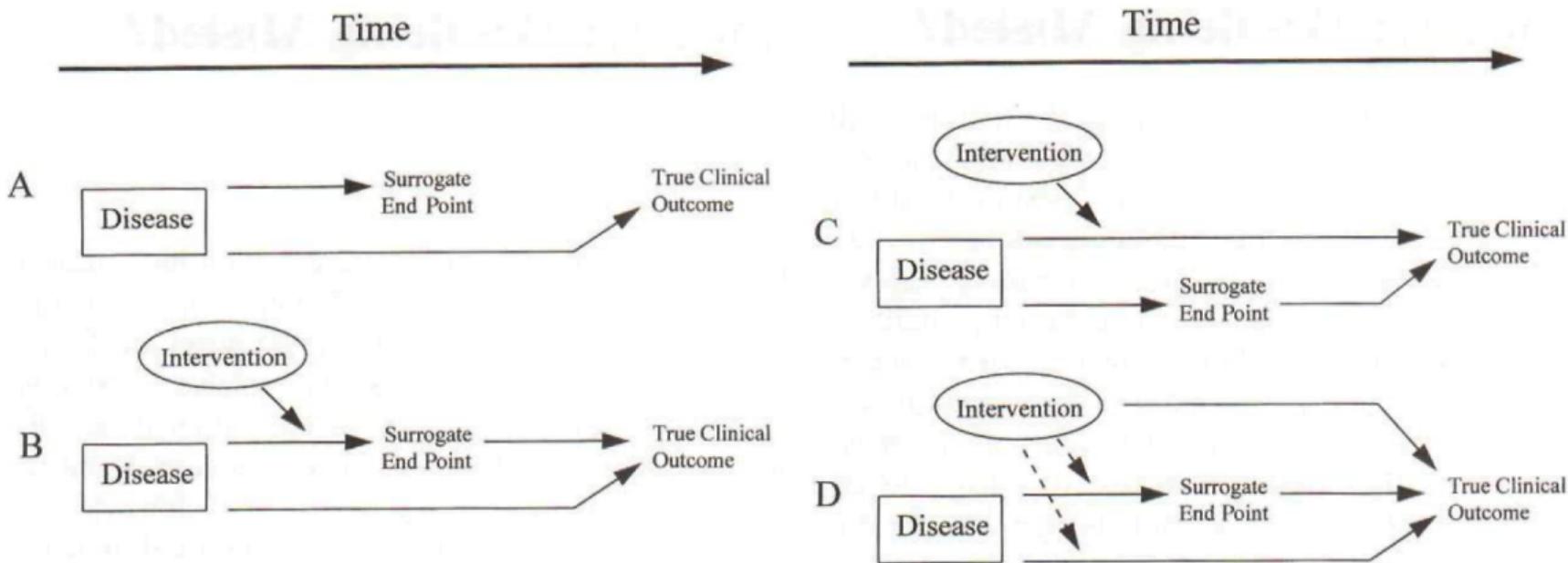
## Kemény vs. surrogate végpont: ami az optimális



Tételmondat: a jó surrogate-séghoz *nem* elég, hogy jól korrelált legyen („correlate does not a surrogate make”)!

Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996 Oct 1;125(7):605-13.

# Kemény vs. surrogate végpont: ami baj lehet



Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996 Oct 1;125(7):605-13.

# Kemény vs. surrogate végpont: pár példa

**Table 1.** Speculation on Reasons for Failures of Surrogate End Points\*

Disease and Intervention	End Points		Settings in Figure 1†			
	Surrogate	Clinical	A	B	C	D
Cardiologic disorder						
Arrhythmia						
Encainide; flecainide	Ventricular arrhythmias	Survival	+	++		
Quinidine; lidocaine	Atrial fibrillation	Survival	+	++		
Congestive heart failure						
Milrinone; flosequinan	Cardiac output; ejection fraction	Survival	+	++		
Elevated lipid levels						
Fibrates; hormones; diet; lovastatin	Cholesterol levels	Survival	+	++		
Elevated blood pressure						
Calcium channel blockers	Blood pressure	Myocardial infarction; survival	+	++		
Cancer						
Prevention						
Finasteride	Prostate biopsy	Symptoms; survival	++‡			
Advanced disease						
Fluorouracil plus leucovorin	Tumor shrinkage	Survival	+	++		
Other diseases						
HIV infection or AIDS						
Antiretroviral agents	CD4 levels; viral load	AIDS events; survival	+	+	+	
Osteoporosis						
Sodium fluoride	Bone mineral density	Bone fractures	+		+	
Chronic granulomatous disease						
Interferon-γ	Bacterial killing; superoxide production	Serious infection		++		

\* AIDS = acquired immunodeficiency syndrome; HIV = human immunodeficiency virus; + = likely or plausible; ++ = very likely.

† A = surrogate end point not in causal pathway of the disease process; B = of several causal pathways of the disease, the intervention only affects the pathway mediated through the surrogate; C = the surrogate is not in the pathway of the intervention's effect or is insensitive to its effect; D = the intervention has mechanisms of action that are independent of the disease process.

‡ In settings in which only latent disease is prevented.

Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996 Oct 1;125(7):605-13.

# Végpontok összefogása

- A legfontosabb ok: dúsítja a végpontok számát, ezáltal növeli az erőt (adott mintanagysággal kisebb különbséget is ki tudunk mutatni, illetve ugyanazon különbség kimutatásához elég kisebb mintanagyság is)
- Védelem az ellen, ha a terápia csak konvertálja a végpontot
- De: ugyanazon biológiai jelenség manifesztációit mérjék, de azért ne túlságosan ugyanazt
- Mely komponens változása adta az összesített változást...?
- Extrém széles összefogás kérdései (pl. UKPDS)

Moyé L. Introduction to Composite Endpoints. In: Moyé L. Multiple Analyses in Clinical Trials – Fundamentals for Investigators. Springer, 2003.

# Egy példa: UKPDS

	Active rx. (n = 2729)	Conv. rx. (n = 1138)	P -value	Rel. risk	95% CI*
Composite endpoints	963	238	0.029	0.88	[0.79 – 0.99]
Fatal endpoints	285	129	0.340	0.90	[0.73 – 1.11]
Fatal MI	207	90	0.630	0.94	[0.68 – 1.30]
Stroke deaths	43	15	0.600	1.17	[0.54 – 2.54]
Renal deaths	8	2	0.530	1.83	[0.21 – 12.49]
Glucose related**	1	1	0.523	0.420***	[0.03 – 6.66]
Sudden death	24	18	0.047	0.54	[0.24 – 1.21]
Death from PVD****	2	3	0.120	0.26	[0.03 – 2.77]
Nonfatal endpoints					
Nonfatal MI	197	101	0.067	0.79	[0.58 – 1.09]
Angina pectoris	177	72	0.910	1.02	[0.71 – 1.46]
Major stroke	114	44	0.720	1.07	[0.68 – 1.69]
Amputation	27	18	0.990	0.81	[0.28 – 1.33]
Blindness	78	38	0.390	0.84	[0.51 – 1.40]
Renal failure	16	9	0.450	0.76	[0.53 – 1.08]
Photocoagulation*****	207	117	0.003	0.71	[0.53 – 0.96]

UKPDS Study Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33).

UK Prospective Diabetes Study (UKPDS) Group. Lancet. 1998 Sep 12;352(9131):837-53.

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
  - Megfigyeléses vizsgálatok
  - Kísérletes vizsgálatok
  - Metaanalízisek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
  - Megfigyeléses vizsgálatok
  - Kísérletes vizsgálatok
  - Metaanalízisek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# A megfigyeléses vizsgálatok alaptípusai

- Egyedi adatok alapján
  - Kohorsz: Az expozíció szerinti csoportok *később* eltérnek-e a végpontban
    - A távvezeték közelében élők körében *később több lesz* a rákos mint az attól távol élők körében?
  - Eset-kontroll: A végpont szerinti csoportok a *múltban* eltértek-e az expozícióban
    - A rákosok körében a *múltban több volt* a távvezetők közelében élő mint az attól távol élő?
  - Keresztmetszeti vizsgálat: Az expozíció szerinti csoportok *most* eltérnek-e a végpontban
    - (Itt nehezen értelmezhető, inkább prevalencia-becslésnél jön szóba)
- Aggregált adatok alapján: ecological módszerek
  - A több távvezetékkel rendelkező országban több a rákos?
- Kontroll nélkül: case report, case series
  - „Beszámolunk Andrásról, aki távvezeték közelében élt és rákos lett...”, „Beszámolunk Andrásról, Béláról és Cecíliáról, akik távvezeték közelében éltek és most rákosak lettek...”

# Kohorsz vizsgálat

- Homogenitást a csoport hasonlósága adja (ugyanazon születési év, ugyanazon lakhely, ugyanazon iskola stb.)
- A kohorsz időben előremenő
- Lehet utólag összerakni (prospektív vs. retrospektív kohorsz), de az alaplogika akkor is időben előremenő
- Általánosságban véve jobb homogenitás érhető el
- Rizikó közvetlenül becsülhető
- Többféle végpont is vizsgálható
- Ritka végpont? (ebből adódóan szükséges mintanagyság?)
- Szükséges idő?

# Eset-kontroll vizsgálat

- Megoldja a ritka végpont és a kivárás ideje problémákat is, viszont...
- Általánosságban véve korlátosztabb homogenitás
- A homogenitást az esetek és a kontrollok hasonlósága adja (matching)...
- ...néhány – általunk fontosnak tartott – szempont ( $\rightarrow$  confounderek) szerint
- Alaplogika időben visszafelé haladó, és szükségképp retrospektív is
- A retrospektivitás miatt kifejezetten sújtja a felidézési torzítás (lásd mindjárt)
- Rizikó nem becsülhető közvetlenül (de: OR számolható, ritka betegségnél jól közelíti az RR-t)
- Egyfélé végpont vizsgálható

## Felidézési torzítás

- Eset-kontrollnál fel kell idézni az expozíciót
- Egyáltalán nem biztos, hogy ezt *ugyanolyan hatékonysággal* teszik meg az esetek és a kontrollok, *pláne* ha az esetek kifejezetten gyanakodnak is arra, hogy az expozíció a ludas a betegségükben
- (A probléma tehát nem is a tökéletlen memória, hanem a *különbség* a két csoport között)
- Lindefors-Harris (1991): ABC-hipotézis, az esetek 22-szer inkább tévednek a nem is megtörtént abortusz irányába
- Rockenbauer (2001): szülés után mennyire emlékeznek arra a nők, hogy szedtek-e adott gyógyszert, 2-szeres különbség az eredményekben aszerint, hogy mi alapján döntjük el, hogy használtak-e gyógyszert

Lindefors-Harris BM, Eklund G, Adami HO, Meirik O. Response bias in a case-control study: analysis utilizing comparative data concerning legal abortions from two independent Swedish studies.

Am J Epidemiol. 1991 Nov 1;134(9):1003-8. Rockenbauer M, Olsen J, Czeizel AE, et al. Recall bias in a case-control surveillance system on the use of medicine during pregnancy. Epidemiology.

2001 Jul;12(4):461-6.

# Információs torzítás

- Az expozíció és/vagy a végpont helytelen rögzítése (mérési hiba) – egyéb okokból is lehet
- Az eseteknél intenzívebben felveszik a kórelőzményt (például kórházi vizsgálat keretében, a kontrollokat csak felhívják telefonon)
- Vagy: ha a kérdező tudja, hogy akit kérdez az eset, akkor jobban kikérdezi
- Vagy: aki maga is érintett, az jobban tudja, hogy a családban előfordult-e korábban hasonló betegség
- stb.

# Ami általános probléma: szelekciós torzítás

- minden egyedi adatokon nyugvó megfigyeléses vizsgálatot sújt
- Vajon a minta milyen része a sokaságnak...?
- Kísérletnél is ugyanúgy kérdés, ott fogjuk tárgyalni

# Ecological módszerek

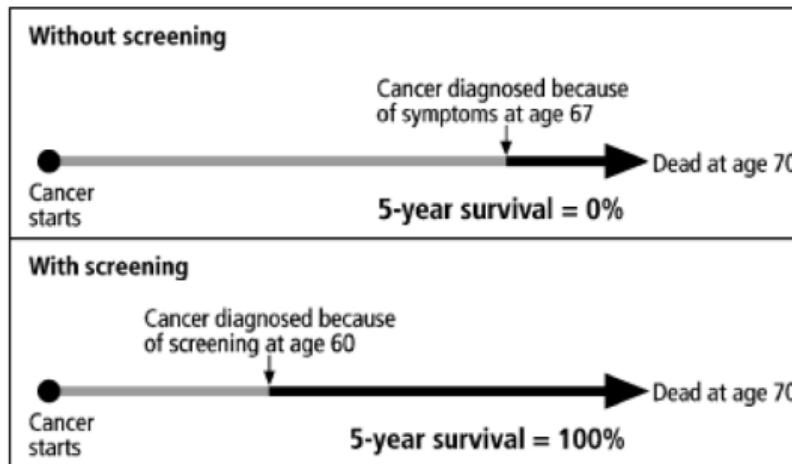
- Az első probléma: minden amit a confounding-ról mondtunk (de a legdurvábban persze)
- Egyik kérdés az aggregáció szintje
- Másik kérdés: longitudinális vagy keresztmetszeti
  - A döntés azon múlik, hogy mi homogénebb: *ugyanaz* az ország (de különböző években), vagy *ugyanaz* az év (de különböző országok); nyilván az adott kérdéstől függ
  - Nagyon nagy általánosságban véve a longitudinális szokott problémásabb lenni (sok idősornak van – növekvő vagy csökkenő – hosszútávú trendje, ezek épp e trend miatt mind korreláltak lesznek egymással!)
- Ráadásul itt még az ecological fallacy is bejön (lásd mindjárt)

# Ecological fallacy

- A második probléma: a korreláció még nem kauzalitás, ezt jól tudjuk, de az ecological adatok még a korrelációt sem feltétlenül bizonyítják!
- Ugyanis érdekes módon az egyéni szinten meglevő korreláció és az aggregátumokon tapasztalt korreláció nem feltétlenül esik egybe
- Matematikailag sem igaz + kontextuális hatások
- Tehát ha korreláltak lennének az is nagyon gyenge bizonyíték lenne, de ráadásul ilyen adatokból még az sem biztos, hogy egyáltalán korreláltak...

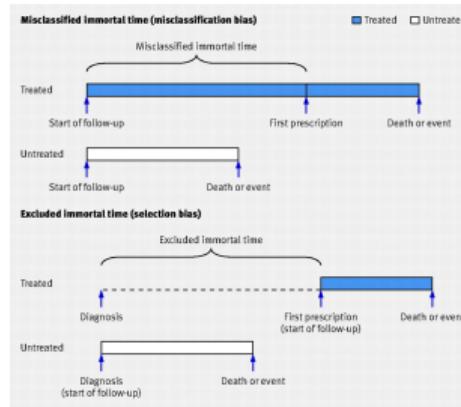
Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. Annu Rev Public Health. 1995;16:61-81.

# Lead time bias



Gigerenzer G, Gaissmaier W, Kurz-Milcke E, et al. Helping Doctors and Patients Make Sense of Health Statistics. *Psychol Sci Public Interest*. 2007 Nov;8(2):53-96.

# Immortal time bias



Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. BMJ. 2010 Mar 12;340:b5087. Suissa S. Immortal time bias in pharmaco-epidemiology. Am J Epidemiol. 2008 Feb 15;167(4):492-9. Kiri VA, Mackenzie G. Re: "Immortal time bias in pharmacoepidemiology". Am J Epidemiol. 2009 Sep 1;170(5):667-8; author reply 668-9. Ho AM, Dion PW, Ng CS, Karmakar MK. Understanding immortal time bias in observational cohort studies. Anaesthesia. 2013 Feb;68(2):126-30.

## Case report és case series

- Ha a kontrollnélküliség miatt lényegében nulla a bizonyítóerejük, akkor miért alkalmazzuk egyáltalán?
- Azért, mert a gyanú felvetésére, tippadásra, hogy mit érdemes egyáltalán – korrekt módszerekkel – megvizsgálni, alkalmas
- Nem hipotézist vizsgáló, hanem hipotézisgeneráló módszerek!
- Sajnos néha nem könnyű megítélni, hogy mi minősül „gyanús”-nak
  - Például gyanús-e, ha 74 csecsemő hal meg egy adott védőoltás beadását követő egy-két héten belül? Eléggé annak hangzik!
  - A világ élvonalába tartozó (de ilyen oltást nem is használó) országok csecsemőhalandósága 2-4, újszülöttalandóságuk 0,5-3 per 1000 körüli
  - Konstans rátójú halálozással számolva 1 hónapos és 1 éves kor között kb.  $3 \cdot 10^{-6}$  a napi halálozási valószínűség
  - Az oltásból kb. 70 milliót adtak be, az oltás és a halálozás *függetlensége* esetén várt egybeesések száma egy nap alatt  $70 \cdot 10^6 \cdot 3 \cdot 10^{-6}$ , azaz kb. 210!

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
  - Megfigyeléses vizsgálatok
  - Kísérletes vizsgálatok
  - Metaanalízisek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# Emlékeztetőül

- Ez kell ahhoz, hogy randomizálhassunk...
- ... ami az egyetlen módszer ahhoz, hogy *biztosan* megszabaduljunk a confoundingtól
- Ez adja a fő erejét
- De ez nem azt jelenti, hogy lehetetlen rosszul csinálni!

# Maszkolás

- Mivel a kezelésben részesülés ténye is képes hatást kiváltani (*placebo-hatás*), így ebben a tudatban sem lehet eltérés
- Megoldás: nem mondjuk meg, hogy ki mit kap (nem is az a lényeg, hogy a kontrollcsoport se tudja, hogy ő nem kezelt, hanem, hogy ne legyen eltérés, ebből a szempontból sem, a két csoport között – hiszen innentől a kezelt csoport sem tudhatja, hogy ő a kezelt, ennek tudatában is teljesen azonosak lesznek)
- Minimum a betegnek (egyszeresen vak vizsgálat), de jobb esetben az orvosnak sem (dupla vak vizsgálat), így még tudattalan módon sem tudja a beteget befolyásolni
- (Mi van akkor, ha a két gyógyszer formája is eltérő? Ez is megoldható, double dummy elrendezés)
- A hatás mérete függhet a várakozásuktól (és a végponttól is)
- A klinikai kísérletek további előnye, hogy maszkolhatóak

# A kontrollálás jellege

- Mi az optimális?
- Placebo? „best available treatment”?
- Csakhogy ez utóbbinál meg nagyon számít a komparátor megválasztása!
  - Nem megfelelő komparátor
  - Megfelelő komparátor nem megfelelő dózisban

Mann H, Djulbegovic B. Comparator bias: why comparisons must address genuine uncertainties. J R Soc Med. 2013 Jan;106(1):30-3. Castro M. Placebo versus best-available-therapy control group in clinical trials for pharmacologic therapies: which is better? Proc Am Thorac Soc. 2007 Oct 1;4(7):570-3. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Ann Intern Med. 2000 Sep 19;133(6):455-63. Ellenberg SS, Temple R. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 2: practical issues and specific cases. Ann Intern Med. 2000 Sep 19;133(6):464-70. Simon R. Are placebo-controlled clinical trials ethical or needed when alternative treatment exists? Ann Intern Med. 2000 Sep 19;133(6):474-5. Wolinsky H. The battle of Helsinki: two troublesome paragraphs in the Declaration of Helsinki are causing a furore over medical research ethics. EMBO Rep. 2006 Jul;7(7):670-2.

# Szelekciós torzítás: kizárási és bevonási kritériumok

**Table 3** Percentage of individuals who potentially meet eligibility criteria for included asthma randomised controlled trials

RCT reference	Current asthma (%), n = 179	Current asthma on treatment (%), n = 127
13	5	7
14	7	10
15	6	9
16	6	8
17	0	0
18	4	6
19	2	3
20	1	1
21	7	9
22	8	11
23	7	9
24	36	43
25	2	2
26	1	0
27	1	2
28	2	3
29	3	5

RCT, randomised controlled trial.

Travers J, Marsh S, Williams M, Weatherall M, Caldwell B, Shirtcliffe P, Aldington S, Beasley R. External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? Thorax. 2007 Mar;62(3):219-23.

# De!

- A kérdés nem annyira kritikus, ha a relatív metrikákat használunk, mert azok általában jól vonatkoztathatóak más populációkra is (lásd a végpontoknál mondottakat)
- Ez persze csak az általánosság, a helyzetet adott esetben külön-külön minden esetben meg lehet vizsgálni

## Kísérletek korai leállítása

- Egy rendesen vakosított vizsgálatban csak a legvégén derül(het)nek ki az eredmények
- Nem feltétlenül jó: sem akkor, ha nagyon hat a gyógyszer, de pláne akkor, ha véletlenül káros lenne
- Nézzünk bele menet közben (egy lehetőleg független bizottság), hogy „hogyan alakul” az ábra

# Kísérletek korai leállítása

- Különösen kérdéses a nagyon jó hatás miatti leállítás: megnő a fals találat esélye azáltal, hogy többször belekukucskálok, több esélyt adok arra, hogy *valamikor* hatásos legyen (mint a szignifikanciaadászat, lásd ott)
- Ezt figyelembe kell venni a végelszámoláskor *még akkor is* (igen, agyzsibbasztó, de tényleg így van!), ha végülis nem állítottam meg semmikor sem

Zannad F, Gattis Stough W, McMurray JJ, et al. When to stop a clinical trial early for benefit: lessons learned and future approaches. *Circ Heart Fail.* 2012 Mar 1;5(2):294-302. Cuzick J, Howell A, Forbes J. Early stopping of clinical trials. *Breast Cancer Res.* 2005;7(5):181-3. Guyatt GH, Briel M, Glasziou P, et al. Problems of stopping trials early. *BMJ.* 2012 Jun 15;344:e3863. Cannistra SA. The ethics of early stopping rules: who is protecting whom? *J Clin Oncol.* 2004 May 1;22(9):1542-5. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet.* 2005 May 7-13;365(9471):1657-61. Chow SC. Adaptive clinical trial design. *Annu Rev Med.* 2014;65:405-15.

## Kiértékelés elve: ITT vs PP

- Protokollsértések szinte minden klinikai kísérletben elkerülhetetlenek: az alany nem kapja meg a karjának megfelelő készítményt, a másikat kapja meg, elfogadhatatlan mellékhatás miatt muszáj abbahagyni a kezelést, kizárasra kerül (visszavonja a beleegyezést, kizáró készítményt kap, teherbe esik stb.), lost to follow-up stb. stb.
- (A „sértés” szót általános értelemben használom, ez lehet valamelyen előre tervezetten megengedett esemény is, például progresszió utáni átkeresztezés a kezelt csoportra egy onkológiai vizsgálatban)
- A megsértés lehet all-or-nothing (pl. egy műtéti eljárás vizsgálatánál), vagy időbeli, dinamikus jelenség (például egy hosszabb időn keresztül adott gyógyszer esetén); ez utóbbi lesz most inkább a fókuszunk
- Mit kezdjünk ezzel?
- Zárjuk ki a vizsgálatból a nem protokoll szerint kezelteket?
- Első ránézésre logikus, sőt, csak ez a logikus, hiszen *nyilván* egy protokoll értékeléséhez nem használhatjuk a nem is a protokoll szerint kezelt betegeket

Dodd S, White IR, Williamson P. Nonadherence to treatment protocol in published randomised controlled trials: a review. Trials. 2012;13(1):84.

# Kiértékelés elve: ITT vs PP

Csakhogy:

- Az még a kisebbik probléma, hogy csökken a mintanagyság, csökken az erő
- Az már nagyobb gond, hogy ilyenek a valóságban is előfordulhatnak, tehát miért ne kellene ezt is belemérni
- De a fő probléma, ha a kizáras nem véletlenszerű: ez elrontja a randomizációt!
- Márpedig ez nagyon is várható (súlyos mellékhatás is gyakoribb a rosszabb állapotú betegeknél, a nem-placebot kapóknál, értelemszerűen; továbbá ők valószínűbb, hogy reményüket is vesztik és kiszállnak a kutatásból; az adherensek általában jobb outcome-ot érnak el stb.)
- Összességében ez szinte biztos, hogy felnagyítja a gyógyszer hatását

# Kiértékelés elve: ITT vs PP

- Megoldás: intention-to-treat (ITT) vs. per-protocol (PP) elv
- ITT: mindenkit abban a csoportban számolunk el, ahová randomizáltuk, *függetlenül* attól, hogy mi történt vele
- Megtartja a mintanagyságot, realisztikusabb, fenntartja a randomizálásból fakadó kiegyensúlyozottságot
- Konzervatív kiértékelés (értsd: kevésbé valószínű, hogy felülbecsli a gyógyszer hatásosságát)
- Ma már lényegében de facto kötelező az ITT használata (de legalább mindenki közlése)
- Persze így hatást is nehezebb lesz kimutatni (gondoljunk bele, elvileg még azokat is elszámoljuk a kezelt csoportban, akik egyetlen kezelést sem kaptak, hígítja a hatást – mITT)
- (A nem-inferioritás külön történet, ott egyáltalán nem biztos, hogy az ITT a konzervatív elv, sőt)

Priya Ranganathan, C. S. Pramesh, Rakesh Aggarwal. Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis. *Perspect Clin Res.* 2016 Jul-Sep; 7(3): 144–146.

Victor M. Montori, Gordon H. Guyatt. Intention-to-treat principle. *CMAJ.* 2001 Nov 13; 165(10): 1339–1341. Sandeep K. Gupta. Intention-to-treat concept: A review. *Perspect Clin Res.* 2011

Jul-Sep; 2(3): 109–112. Giovanni Tripepi, Nicholas C Chesnaye, Friedo W Dekker, Carmine Zoccali, Kitty J Jager. Intention to treat and per protocol analysis in clinical trials. *Nephrology (Carlton).* 2020 Jul;25(7):513-517. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ.* 1999 Sep 11;319(7211):670-4.

Montedori A, Bonacini MI, Casazza G, et al. Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. *Trials.* 2011 Feb 28;12:58.

# Tartalom

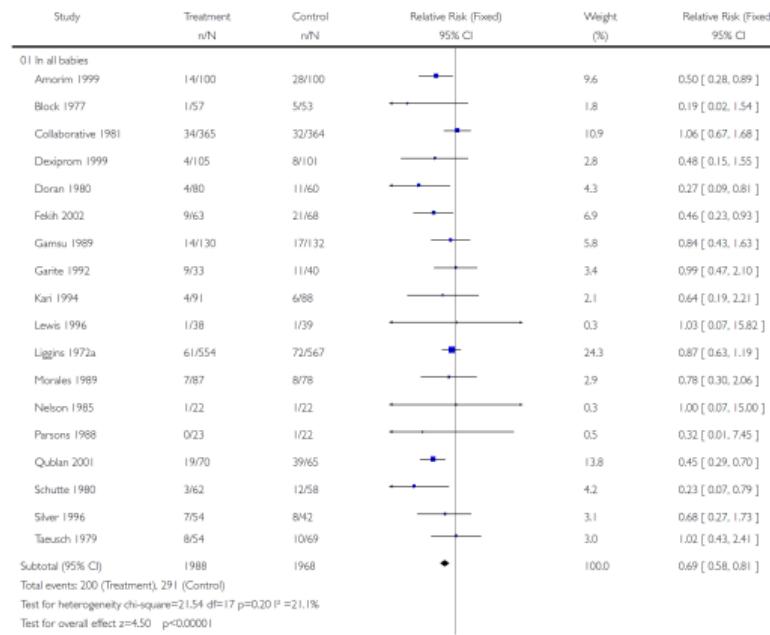
- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
  - Megfigyeléses vizsgálatok
  - Kísérletes vizsgálatok
  - Metaanalízisek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# A metaanalízis fogalma

- Több, ugyanarra a kérdésre vonatkozó vizsgálat eredményeinek – bizonyos módszer szerinti – aggregálása
- (Itt természetesen kvantitatív eredményű, empirikus vizsgálatokról beszélünk)
- Akkor alkalmazzuk jellemzően, ha a páciens-szintű adatok nem elérhetőek (különben klasszikus módszerekkel is aggregálhatóak a eredmények)
- Tehát többször vizsgáljuk ugyanazt a kérdést: hát persze, a kutatások replikációja a tudományos munka egyik alappillére
- De hogyan kell ezt megtenni? Mi erre a korrekt (statisztikai) módszertan?

# Metaanalízis végeredményének közlése

Tipikus grafikus megoldás a forest plot:



# A metaanalízis története

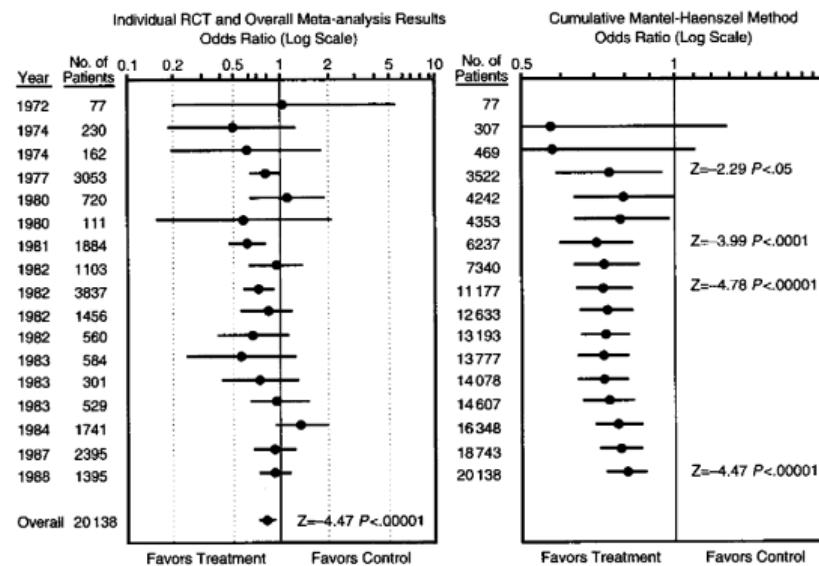
- Előzmények: 17., 18. században (George Biddell Airy)
- Karl Pearson és a tífusz elleni oltás (1904)
- Modern kezdetek: agrometriai eredmények kombinálása (már az 1930-as évektől), Ronald Fisher és Cochran, Yates
- 20. század közepe: elsősorban társadalomtudományi alkalmazások (szociológia, pszichológia)
- A kifejezést Gene Glass vezette be, 1976-ban („analysis of analysis”)
- Egyik első emlékezetes orvosi alkalmazás: aspirin és szívinfarktus (1974-1980)
- Elképesztő fejlődés azóta, alapvető eszközzé vált az orvostudományban (együtt a vizsgálatok számának robbanásszerű növekedésével)

# Miért végzünk metaanalízist?

- Egyszerűen ugyanazért, amiért a nagyobb mintanagyság jobb (szűkebb konfidenciaintervallumok, erősebb tesztek)
- ... de ez „olcsóbb” (nem csak forintban, kivitelezési időben is)
- Konzisztencia megítélése vizsgálatok között
  - Sőt, adott esetben megítélhető, hogy a vizsgálatra jellemző valamelyen tényező számszerűen hogyan függ össze a vizsgálatban talált eredménnyel (metaregresszió)
- Publikációs torzítás felfedése
- Persze a metaanalízis sem hibátlan: hogy milyen tanulmányokat vonunk be, az maga is lehet torzítás forrása (minél több paramétert kell hozzá megszabni, annál inkább – „összelőhetjük” úgy, hogy pont az jöjjön ki, ami a prekonцепciónk)
- „GIGO-elv”

# Nagyobb mintanagyság

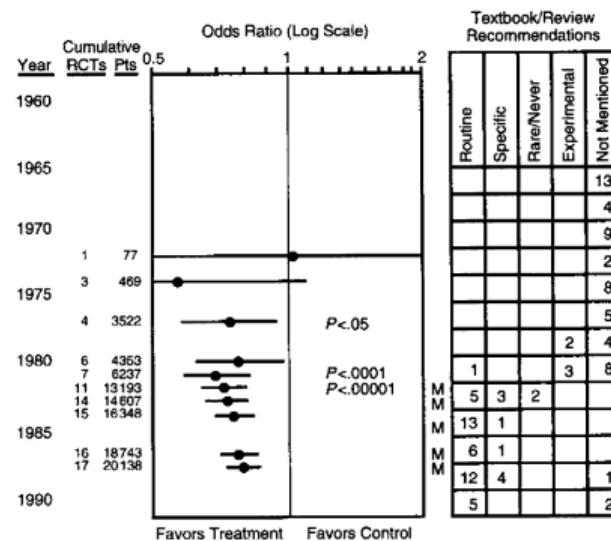
Béta-blokkolók adása szívinfarktus másodlagos prevenciójában:



Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA. 1992 Jul 8;268(2):240-8.

# Nagyobb mintanagyság

Béta-blokkolók adása szívinfarktus másodlagos prevenciójában:



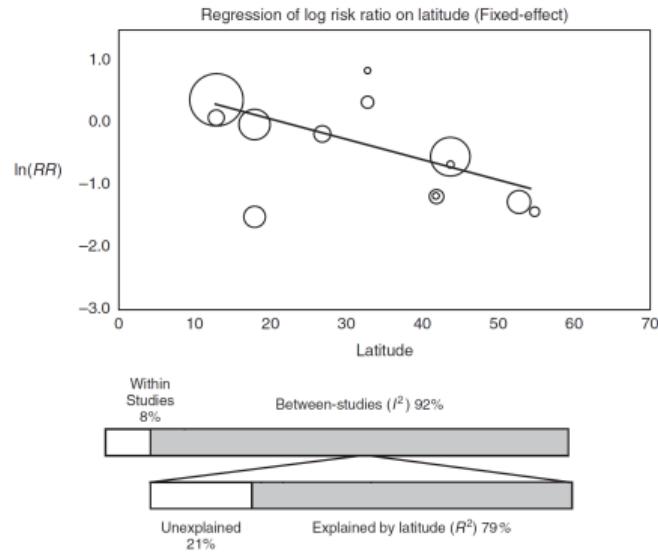
Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA. 1992 Jul 8;268(2):240-8.

# A Cochrane-logó és története



# Konzisztencia megítélése és metaregresszió

BCG hatásossága a tüdő-tuberkulózis ellen:



Borenstein M, Hedges LV, Higgins J, Rothstein HR. Introduction to Meta-Analysis. Wiley. 2009.

# Publikációs torzítás megítélése

Lásd később, a rendszerszintű hibáknál

# Mikor végezhető metaanalízis?

- Ugyanazt a kérdést vizsgálja több kutatás
- Mi az, hogy „ugyanaz”...?
- Két kutatás soha nem tökéletesen ugyanolyan, de legyen
  - hasonló, vagy legalábbis összevethető a minta (definíció, bevonási, kizárási kritériumok!), a kutatás alapkoncepciója, tervezése
  - ugyanaz az expozíció
  - ugyanaz, vagy legalábbis közös nevezőre hozható a hatásnagyság-mutató

# Metaanalízis lépései

- Nulladik lépés: az előbbiben szereplő kérdéseket rögzíteni, definiálni kell (több tucat vagy annál is több oldal lehet egy ilyen protokoll!)
- Elsőként össze kell gyűjteni az aggregált tanulmányokat
- Lehetőleg nem ad hoc módon: szisztematikus review
- Egységes formátumra kell hozni őket, megteremtve az összehasonlíthatóságot (pl. azonos hatásnagyság-mutató kiszámítása)
- El kell végezni a metaanalízist...
- ...majd a szükséges diagnosztikát: részint magára a metaanalízisre vonatkozóan (pl. különböző analitikus módszerek hatása a végeredményre), részint a nyers adatokra vonatkozóan (pl. publikációs torzítás megítélése)

# Hatásnagyság mutató fogalma

- A tanulmányok összehasonlíthatóságának alapfeltétele, hogy ugyanazon a módon mérjük a végpontot (kompatibilis hatások)
- A mutatók persze függnek a végpont jellegétől:
  - Bináris (vagy arány): kockázat-különbség, kockázat-arány (relatív rizikó), log relatív rizikó, esélyhányados, log esélyhányados, ...
  - Folytonos: átlag-különbség, standardizált átlag-különbség, relatív átlag-különbség, ...
  - Korreláció: korrelációs együttható, transzformáltjai, ...

# A szisztematikus review

- Egy ilyen kutatás nem a metaanalízisnél kezdődik...
- ... hanem a metaanalizált tanulmányok összegyűjtésénél!
- Szisztematikus vs. nem-szisztematikus
- Számos kérdést vet fel: betegség definíciója, kontrollcsoport, betegjellemzők, hosszúság, stb.

# Módszertani alaplogika

- A továbbiakban feltételezzük, hogy az előkészületek megtörténtek:
  - összegyűjtöttük a megfelelő tanulmányokat (lehetőleg szisztematikusan)
  - összevethető formátumra hoztuk őket
- Így tehát van egy adatbázisunk: minden sor egy tanulmány, azaz minden sorban van egy  $Y_i$  hatásmutató (ugyanaz!), és esetleg még egyéb információk a kutatásról
- A kérdés már csak az (innen jön a matematika): hogyan aggregáljuk őket?

# Tanulmányok eredményeinek aggregálása

- Az alapötlet: az átlagos hatást több kutatásból úgy kapjuk meg, hogy kiátlagoljuk őket
- De nem egyszerű átlaggal (ha egy 10 fős tanulmány szerint -5% a hatás, egy 10 ezer fős szerint pedig +5%, akkor aligha mondható, hogy átlagban nincs hatás)...
- ... hanem súlyozott átlaggal!
- A fontosabb, lényegesebb tanulmányok nagyobb súlyt kapnak az átlagolásban!
- Na de mi az, hogy „fontosabb”?
- Legáltalánosabb módszer: inverse variance weighting

# Az inverse variance weighting lényege

- A súly a varianciával fordítottan arányos:  $W_i = \frac{1}{V_i}$
- Itt a variancia természetesen a mintavételi variancia...
- ... azaz a súlyozás a mintavételi bizonytalansággal fordítottan arányos: minél kisebb a mintavételi bizonytalanság egy kutatásban, annál nagyobb súlyt kap az átlagolás során!

# Fix és random hatás feltevése

- Alapfeltételezés: az  $Y_i$  hatásmutatók eloszlása normális
- Fix hatás:  $Y_i = \theta + \varepsilon_i$ , ahol  $\varepsilon_i \sim \mathcal{N}(0, V_i)$
- Random hatás:  $Y_i = \theta + \theta_i + \varepsilon_i$ , ahol  $\theta_i \sim \mathcal{N}(0, \tau^2)$  és  $\varepsilon_i \sim \mathcal{N}(0, V_i)$

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételei hiba, következtető statisztika
  - A következtető statisztikáról részletesebben
  - A legfontosabb félreértések a *p*-érték kapcsán
  - Klinikai és matematikai szignifikancia
  - Szignifikanciaadászat és a többszörös összehasonlítások helyzete
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételi hiba, következtető statisztika
  - A következtető statisztikáról részletesebben
    - A legfontosabb félreértések a *p*-érték kapcsán
    - Klinikai és matematikai szignifikancia
    - Szignifikanciaadászat és a többszörös összehasonlítások helyzete
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# A sokaság és a minta fogalmáról

- „Mennyi a jelenlevők átlagos testtömege?” → véges sokaság ( $N = 30$ ), név szerint felsorolhatóak az elemei
- De: „Egy új vérnyomáscsökkentő gyógyszer-jelölt valóban csökkenti a vérnyomást?”
  - Mi itt a sokaság?
  - Ennek nem lehet név szerint felsorolni az elemeit, ez egy absztrakt konstruktum
  - Szokás fiktív sokaságnak hívni (vagy, szintén használt találó névvel, végtelen sokaságnak)
- Általában nem tudjuk az egész sokaságot megfigyelni → *mintavételes helyzet*
- Amit meg tudunk figyelni: *minta*

# A problémafelvetés

- Nagyon sok esetben tehát technikai okokból, vagy elvileg is lehetetlen a teljes sokaság megfigyelése
- Csak egy részét, a mintát ismerjük
- És itt jön a kulcsprobléma: *mi mégis a sokaságról akarunk nyilatkozni!*
- Lehet egyáltalán? Hogyan?
- Biztosan már nem tudunk mondani... de valószínűségi állítást igen!

# Mintavételi ingadozás

- Ha csak a sokaság egy részét (a mintát) ismerjük, akkor minden belőle számolt jellemző két dologtól fog függeni:
  - ① a jellemző sokaságbeli értékétől (a gyógyszer valódi hatásától)
  - ② attól, hogy konkrétan hogyan választottuk ki a mintát (kik kerültek be a gyógyszerkísérletbe)
- Mi értelemszerűen az elsőre vagyunk kíváncsiak
- ...csakhogy a – kikerülhetetlen – második („pont milyen mintát vettünk”) azt fogja okozni, hogy minden eredményünk mintáról-mintára változni fog
- A szerencse: ez az ún. mintavételi ingadozás követ valószínűségszámítási törvényeket, így valószínűségi állításokat meg tudunk fogalmazni!
- Hibázhatunk, de ennek természetéről tudunk nyilatkozni

# Mintavételi hiba

- Figyelem, ennél a hibázásnál nem arról van szó, hogy „rosszul” veszünk mintát: például a legtökéletesebben véletlenszerű mintavételnél is előfordulhat, hogy egy 100 fős sokaságból úgy becsüljük az átlagos testtömeget, hogy pont a 20 legkönnyebbet választjuk ki
- De: ennek a valószínűsége extrém kicsi! (egész pontosan  $1/\binom{20}{100} \approx 2 \cdot 10^{-19}\%$ )
- Az érzékeltetés kedvéért: ha ekkora valószínűsséggel dobunk fejet egy cinkelt pénzérmével, és másodpercenként 1-szer fel tudjuk dobni, akkor az első fej várhatóan 17 ezer milliárd év múlva fog kijönni...
- (A trükkös rész: egy konkrét mintánál soha nem tudhatjuk, hogy az véletlenül nem pont az az 1-e, ami átlag 17 ezer milliárd évente egyszer fordul elő!)
- Az egész csak ilyen képzeletbeli ismételt mintavételi értelemben nyer értelmet; ezt a megközelítést szokás frekvencionista statisztikai iskolának nevezni
- Statisztikai apparátus: becsléselmélet („konfidenciaintervallumok”) és hipotézisvizsgálat („ $p$ -értékek”)

## Becsléselmélet

- Becslőfüggvény (pl. a sokasági átlagot „tippeljük” a minta átlagával)
- Persze tévedhetünk, de reméljük, hogy vannak jó tulajdonságai a becslőfüggvényünkben (torzítatlanság, hatásosság)
- A fentiekkel egyetlen számot, „a” legjobb becslést adjuk vissza eredményként
- Nem adunk számot arról, hogy ebben mekkora a bizonytalanság...
- ...pedig erről is tudunk nyilatkozni! („Kalkulálható bizonytalanság”)
- Tipikus szemléltetés: *konfidenciaintervallum* (CI): mi az a tartomány, amire igaz, hogy ha sokszor megismételnék a mintavételt, és minden egyik mintából megszerkesztenénk a CI-t, akkor ezen CI-k várhatóan 95%-a tartalmazná az igazi (sokasági) értéket (95% megbízhatóság mellett CI)

## A konfidenciaintervallumról bővebben

- „Fordított logika” (ezt később is látni fogjuk): nem azt mondjuk, hogy ha ez a minta, akkor hol lehet a valódi érték, hanem, hogy ha ez a valódi érték, akkor hihető-e még, hogy ez jöjjön ki mintaként
- Úgy is szokták mondani, hogy a CI megadja, hogy – ilyen értelemben – mi kompatibilis a mintával
- (Finom különbségnek tűnik, pedig hatalmas jelentősége van, lásd később, a hipotézisvizsgálatnál)
- A „hihetőségre” természetesen határt kell húzni – ez lesz az a bizonyos megbízhatósági szint
- minden CI szerkesztő eljárásnak vannak előfeltevései (a több ilyet használók általánosságban szűkebb intervallumot adnak, de ha nem fennálló előfeltevésre építünk, akkor invalid lesz a CI, tehát nem 95% lesz a lefedése)
- Nagyobb megbízhatóság ↔ semmitmondóbb intervallum

# A hipotézisvizsgálat alapfogalmai

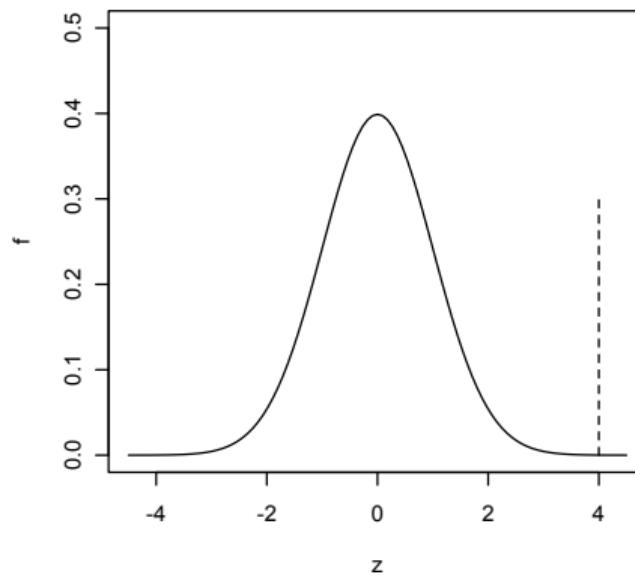
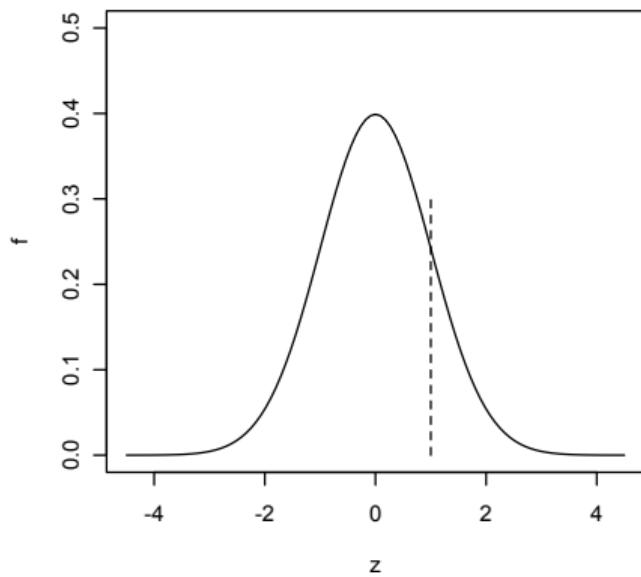
- Feladat: sokaságra vonatkozó állítás eldöntése minta alapján
- Lényegében az intervallumbecslés ikertestvére, de hatalmas gyakorlati jelentősége miatt külön eszköztára van
- Alapeszköze a *statisztikai próba* (vagy teszt), mely a mintaelemek alapján kiszámol egy ún. *tesztstatisztikát* (próbafüggényt)
- Vizsgált állításaink: nullhipotézis – ellenhipotézis
- Egy tipikus példa:

$$H_0 : \text{HR} = 1$$

$$H_1 : \text{HR} \neq 1$$

## Döntés a hipotézisvizsgálatban

- Hihető-e, hogy az empirikus (adott, konkrét mintából kapott) érték ebből az eloszlásból származik?
- Biztos döntés nincs! De: mennyire hihetőek ezek?



# Döntés a hipotézisvizsgálatban

- Megint csak, valahol „határt kell húznunk” → szó szerint is!
- Azt mondjuk, hogy a nagyon kis valószínűségű területekre esést már nem hisszük el
- Pedig az *nem* lehetetlen, sőt: az is tudható, hogy az oda esés (azaz a fenti logikával történő hibázás) valószínűsége épp ez a nagyon kis valószínűség
- Tipikus, hogy a felső és alsó szélén is 2,5-2,5 % valószínűségű területet jelölünk ki ( $\alpha = 5\%$ , ez a szignifikanciaszint), határai: a  $c_a$  alsó és a  $c_f$  felső kritikus értékek (példában:  $\pm 1,96$ )

# *p*-érték

- Vagy: Mennyi lenne az a szignifikanciaszint, ami mellett a mintából kapott (empirikus) tesztstatisztika-érték épp az elfogadás és az elutasítás határára kerülne?
- (Ez nem más, mint az empirikus értéktől extrémebb helyeken vett integrálja a mintavételei eloszlásnak)
- A neve: *p*-érték
- „Az olvasó is tud döntení”: ha a választott szignifikanciaszint nagyobb, mint a *p*-érték, akkor elutasítunk, különben elfogadunk
- Ez természetesen ugyanaz, mint a korábban látott definíció: ha fennáll a nullhipotézis (pl. nem hat a gyógyszer), mekkora valószínűsséggel kapunk olyat – vagy még extrémebbet – mint amit ténylegesen kaptunk is
- Számos félreérts, lásd mindjárt és filozofiai kérdőjel, lásd később...

# Döntés a hipotézisvizsgálatban

- „Fordított logika” (megint csak) a döntéshez: *ha* fennáll a nullhipotézis – nem hat *valójában* a gyógyszer – *akkor* mennyire hihető, hogy a mintában *mégis* hat
- A mintavételei hiba miatt nem nyilvánvaló kérdés: ha a kezelt csoportban 10% halt meg, a kontrollcsoportban pedig 20, attól még lehet, hogy nem hat a gyógyszer – mert a valóságban minden két csoportban 15% hal meg, de mi pont ilyen mintát vettünk (pusztán a véletlen ingadozás miatt!)
- Ennek a valószínűségét méri a *p*-érték (precízen: hogy ilyen vagy ennél is extrémebb mintát kapunk, feltéve, hogy fennáll a nullhipotézis)
- Logika: ha ez nagyon kicsi, akkor azt mondjuk, hogy hat a gyógyszer (mert a másik magyarázat, hogy egy nagyon valószínűtlen esemény történt)
- Vigyázat, hipotézisvizsgálatban minden az elutasítás az erős döntés (nem véletlen, hogy úgy fogalmazunk, hogy „nem tudtuk elvetni” a nullhipotézist, nem úgy, hogy „elfogadtuk”)
- A hiány bizonyítéka nem a bizonyíték hiánya!

# Próba hibái

- Elvetjük  $H_0$ -t, pedig fennáll (elsőfajú hiba,  $\alpha$ ): pontosan szabályozható valószínűségű
- Elfogadjuk  $H_0$ -t, pedig el lehetne vetni (másodfajú hiba,  $\beta$ ): általánosságban nem ismert, függ a valóságtól
- $1 - \beta$ : próba ereje („mennyire ismeri fel az eltérést, ha tényleg van”)
- Mi két dologgal tudjuk befolyásolni a próba erejét, mindkettőhöz egy-egy tételemondat:
  - ① Választott próba: „mindig annyi előfeltevésre építő próbát használunk, amennyit tudunk, se többet se kevesebbet” (több előfeltevésre építő próbák erősebbek ugyan, de ha szükséges előfeltevés nem teljesül, a próba nem lesz valid)
  - ② Mintanagyság: „kis hatáshoz nagy minta kell, nagy hatáshoz elég a kisebb minta is” → ha feltételezzük egy eltérést, akkor kiszámítható, hogy annak adott valószínűségű – erejű – kimutatásához hány beteget kell bevонni!

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételei hiba, következtető statisztika
  - A következtető statisztikáról részletesebben
  - **A legfontosabb félreértések a *p*-érték kapcsán**
  - Klinikai és matematikai szignifikancia
  - Szignifikanciaadászat és a többszörös összehasonlítások helyzete
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# A legfontosabb félreértés a $p$ -érték kapcsán

- Az első és legfontosabb félreértés: a  $p$ -érték csak *kizárolag* a mintavételi hibát jellemzi, a világon semmit, de semmit nem mond a nem-mintavételi hibákról
- ...miközben lehet, hogy egy kutatásban a nem-mintavételi hibák dominálnak!
- (Ebben az esetben a  $p$ -értékek közlése akár kifejezetten félrevezető is lehet!)
- Közel nem igaz tehát, hogy a  $p$ -érték (és társai) valamiféle univerzális metrikái lennének egy kutatás potenciális hibázásának

## Második legfontosabb félreértés: egy kis motiváció

- Terroristák a városban...
- Mellesleg mitől medikális ez a példa?
- ...ritka betegségek szűrése!
- Na és mi köze a *p*-értékekhez?
- Íme...

## A második legfontosabb félreértés: $p$ -érték = hibavalószínűsége

- Az egyszerűség kedvéért a kísérletünknek csak két kimenete lehet: a gyógyszer hatásos és hatástalan
- Kétféle hiba van, legyen a hatástalan hatásosnak minősítése 5, a hatásos hatástanannak minősítése 20% valószínűségű
- Az utóbbi azt jelenti, hogy 80% az erő...
- ...az előbbi épp a választott szignifikanciaszint
- A hatástananság a priori valószínűsége 90%

# Mi az a prior valószínűség?

- „Előzetes” valószínűség: *még mielőtt* egyáltalán megkezdenénk a vizsgálatot, mennyi a valószínűsége annak, hogy hatástalan a gyógyszer
- Hogy micsoda? Hát ennek mi értelme? Pont azért csináljuk a kutatást, mert nem tudjuk, hogy hat-e...!
- Korábbi vizsgálatok, kutatáson kívül egyéb okok, például más területek eredményei (biológiai megfontolások), stb.
- Ha valaki szeretné valami konkréthoz kötni, akkor gondoljon arra, hogy egy számos már sikeres gyógyszerrel rendelkező gyógyszercsalád egy minimálisan módosított új tagjánál ez a valószínűség nagy, annak viszont, hogy a hiperpulzatív mágneses térrrel kvantumtranszformált rezgőkristály hat, ez a prior valószínűség csekély

## Egy kis fejszámítás

- Hogy matematikai részletek nélkül megindokoljam ezt, végezzünk egy gondolatkísérletet
- 1000-szer szimuláljuk ezt a világot (avagy 1000 párhuzamos univerzumot tekintünk)
- $1000 \cdot 0,1 = 100$ -szor hatásos lesz a gyógyszer...
  - ... ebből  $100 \cdot 0,2 = 20$ -szor hatástalannak minősítjük, a maradék 80 esetben hatásosnak
- $1000 \cdot 0,9 = 900$ -szor hatástanlan lesz a gyógyszer...
  - ... ebből  $900 \cdot 0,05 = 45$ -szor hatásosnak minősítjük, a maradék 855 esetben hatástalannak
- Összességében  $80 + 45 = 125$  esetben lesz hatásosnak minősítve a gyógyszerünk
- A hibarány, tehát, hogy a hatásosnak minősítések közül mekkora arányban hatástanlan valójában a gyógyszer  $45/125 = 36\% - \text{marhára nem } 5\%$ !

Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. R Soc Open Sci. 2014 Nov 19;1(3):140216.

# A hatás

Power of study (proportion (%) of time we reject null hypothesis if it is false)	Percentage of "significant" results that are false positives		
	P=0.05	P=0.01	P=0.001
<b>80% of ideas correct (null hypothesis false)</b>			
20	5.9	1.2	0.10
50	2.4	0.5	0.05
80	1.5	0.3	0.03
<b>50% of ideas correct (null hypothesis false)</b>			
20	20.0	4.8	0.50
50	9.1	2.0	0.20
80	5.9	1.2	0.10
<b>10% of ideas correct (null hypothesis false)</b>			
20	69.2	31.0	4.30
50	47.4*	15.3	1.80
80	36.0	10.1	1.10
<b>1% of ideas correct (null hypothesis false)</b>			
20	96.1	83.2	33.10
50	90.8	66.4	16.50
80	86.1	55.3	11.00

\*Corresponds to assumptions in table 2.

Azaz: elborult ötleteknél még egy kimondottan alacsony *p*-érték is jelentheti azt, hogy valószínűleg *nem* vethető el a nullhipotézis (az ilyeneknél nagyon erős bizonyíték kell)!

Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? BMJ. 2001 Jan 27;322(7280):226-31.

# A probléma oka

- A kutya tehát ott van elásva, hogy mi a prior valószínűsége
- A *p*-érték „előnye”, hogy ez nem kell hozzá
- De emiatt nem is azt méri, amit sokan gondolnának!
- Azt is lehet, de ahhoz kell a prior valószínűség: Bayes-tétel

$$\mathbb{P}(H_0|\text{Minta}) = \frac{\mathbb{P}(\text{Minta}|H_0) \cdot \mathbb{P}(H_0)}{\mathbb{P}(\text{Minta})}$$

- Út a Bayes-faktorokhoz

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. Ann Intern Med. 1999 Jun 15;130(12):1005-13.

## Cromwell-elv

- De: még a legelborultabb ötletekhez se rendeljünk 0 prior valószínűséget, rendeljünk nagyon picit, de ne 0-t
- „Ahhoz se 0 prior valószínűséget rendeljünk, hogy sajtóból van a Hold, különben egy hadseregnyi, sajtal visszatérő űrhajós sem fog meggyőzni minket” (Dennis Lindley)
- Tehát ahhoz, hogy hat a hiperpulzatív mágneses térrrel kvantumtranszformált rezgőkristály se 0 prior valószínűséget rendeljünk! Rendeljünk  $10^{-10}$ -t,  $10^{-100}$ -t,  $10^{-1000}$ -t, de *ne* nullát; ez a Cromwell-elv
- Ez tökéletesen racionális: nagyon erős bizonyíték kell, hogy elhiggyük, hogy hat (összhangban azzal, hogy az alapkutatási eredmények ennek ordítóan ellentmondanak), de közben nem is lehetetlen, tehát ha sorban meggyógyít mindenkit, akkor be fogjuk látni, hogy tényleg hat

# Frekventista és bayes-i statisztika

- Az előzőekből látható, hogy miért gondolják a legtöbben, hogy jobb lenne bayes-i keretet használni az orvosi vizsgálatok kiértékelésére
- Mégis, szinte kizárolagos a frekvencionista elv használata:
  - A bayes-i statisztika számításigényes (ma már nem lényeges ellenérv, van számítási kapacitás könnyen elérhetően)
  - Apróbb filozófiai és statisztikai ellenérvek (pl. a valószínűség bayes-i interpretációja vagy, hogy minden hipotézist azonosítani kell)
  - De a legeslegfontosabb: az orvosok gyomra nem veszi be a prior valószínűség használatát (mondván, hogy az „szubjektív” – noha a frekvencionista keretnek nem kevesebb feltevése van, csak azok nem ilyen látványosak)
  - És végül a tehetségtelenesség: a frekvencionista iskola rendkívül jól kidolgozott elméletileg és gyakorlatilag is, kitűnő számítógépes támogatással rendelkezik, meg egyáltalán, 100 éven keresztül megszámlálhatatlan sok vizsgálatban ezt alkalmazták, ezt tanulta szinte mindenki

# A *p*-értékek minősítése, a nagyságának a jelentősége

P-VALUE	INTERPRETATION
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	
0.06	ON THE EDGE OF SIGNIFICANCE
0.07	
0.08	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P<0.10$ LEVEL
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
$\geq 0.1$	

Aki azt gondolná, hogy ez egy vicc:

<https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/>

## A harmadik legfontosabb félreértés: a bizonyíték hiánya nem a hiány bizonyítéka

- A kétféle döntés nem szimmetrikus: a hatástalanságot erősen el tudjuk vetni, de olyan nincs, hogy erősen elfogadjuk
- Ha 100 kezeltből 0 hal meg, 100 kontrollból 100, akkor nagyon erősen (megj.: nem biztosan, ugye!) el tudjuk vetni a feltevést, hogy a gyógyszer hatástalan
- De ha 100 kezeltből és 100 kontrollból egyaránt 50 hal meg, még akkor sem mondhatjuk, hogy nagyon erősen elfogadjuk, hogy hatástalan
- Miért? Mert erő is van a világon! Simán lehet hatása, de  $2 \times 100$  betegből kimutathatatlan (gondoljuk arra, ha csak  $2 \times 10$  betegen találtuk volna ugyanezt!)
- A „nem tudtuk bizonyítani, hogy van hatás” tehát *nagyon nem ugyanaz*, mint az, hogy „bizonyítottuk, hogy nincs hatás”!

Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ. 1995 Aug 19;311(7003):485. Alderson P. Absence of evidence is not evidence of absence. BMJ. 2004 Feb 28;328(7438):476-7.

# Egy példa: dexmedetomidine jót tesz lélegeztetett szeptikus betegeknél?

## Key Points – Meaning

Treatment with dexmedetomidine in patients with sepsis did not improve either ventilator-free days or 28-day mortality.

## Számszerűen

Mortality at 28 days was not significantly different in the dexmedetomidine group vs the control group (19 patients [22.8%] vs 28 patients [30.8%]; hazard ratio, 0.69; 95% CI, 0.38-1.22; P = .20).

## Jobbak eset, mert ők legalább tudják

Among patients requiring mechanical ventilation, the use of dexmedetomidine compared with no dexmedetomidine did not result in statistically significant improvement in mortality or ventilator-free days. However, the study may have been underpowered for mortality, and additional research may be needed to evaluate this further.

Kawazoe Y, Miyamoto K, Morimoto T. Effect of Dexmedetomidine on Mortality and Ventilator-Free Days in Patients Requiring Mechanical Ventilation With Sepsis: A Randomized Clinical Trial.

JAMA. 2017 Apr 4;317(13):1321-1328.

## Az „underpowered” kutatások problémái

- Első ránézésre a szponzor baja, ha a kutatás ereje kicsi (hat a gyógyszer, mégis kidobják az ablakon a pénzt, hogy ezt kimutassák)
- Tehát az orvosi tudásunk csak olyan irányban lehet tőle hibás, hogy valamiről azt hisszük, hogy nem hat, pedig igen
- Ez nem így van:
  - Fontos lenne tudni, hogy egyáltalán volt-e esély kimutatni valamit – ugyanis nem csak hatásosságot vizsgálunk, hanem biztonságosságot is!
  - Még a hatásosságnál sem teljesen így van: a kis erejű vizsgálatok, ha találnak is hatást, az várhatóan nagyobb lesz (az ilyenekben nagyobb ingadozás, a hatásosság kimutatásához nagyobb hatás kell: a fals hatás kimutatásának a valószínűsége így ugyanannyi lesz, csak épp ha megtörténik, akkor a kimutatott hatás nagyobb lesz)
  - Azaz felülbecsüljük a hatást!

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013

May;14(5):365-76.

# A *p*-érték aszimmetriája

- Ha nagyon kicsi a *p*-érték, az nagyon erősen amellett szól, hogy van hatás; minél kisebb, annál inkább
- De ez fordítva nem igaz: a nagy *p*-érték (ad abszurdum a  $p = 1$ ) sem mond semmit arról, hogy *mennyire* erősen gondoljuk azt, hogy nincs hatás!
- Gondoljunk bele: 347/5033 (6,89%) vs. 351/5033 (6,98%),  $p=0,87$  és 5/19 (26,32%) vs. 4/14 (28,57%),  $p=0,89$  – ez a kettő biztos ugyanaz?!
- A dolog tehát aszimmetrikus, a hatástalanságban való bizonyosságot nem jellemzi a *p*-érték
- Bayes-i keretben jól kezelhető
- Illetve megint csak: ha nem is térünk át a bayes-i keretre, legalább preferáljuk a konfidenciaintervallumokat a *p*-értékekkel szemben

Hoekstra R, Monden R, van Ravenzwaaij D, Wagenmakers EJ. Bayesian reanalysis of null results reported in medicine: Strong yet variable evidence for the absence of treatment effects. PLoS One.

2018 Apr 25;13(4):e0195474.

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételei hiba, következtető statisztika
  - A következtető statisztikáról részletesebben
  - A legfontosabb félreértések a  $p$ -érték kapcsán
  - **Klinikai és matematikai szignifikancia**
  - Szignifikanciaavadászat és a többszörös összehasonlítások helyzete
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# Klinikai vs. matematikai szignifikancia

- A „szignifikáns hatás” tehát azt jelenti: túlmutat a mintavételi ingadozáson
- De sajnos a szót nem csak ilyen értelemben használjuk!
  - „A gyógyszer szignifikánsan növeli a túlélési időt ( $p = 0,02$ )”
  - „Peti szignifikánsan magasabb mint Béla”
- Klinikai és matematikai szignifikancia fogalmának keveredése
- Klinikailag jelentéktelen hatás is lehet – statisztikai értelemben – szignifikáns (és fordítva)

# Egy példa: erlotinib a gemcitabine mellett

## Az absztrakt konklúziója

To our knowledge, this randomized phase III trial is the first to demonstrate statistically significantly improved survival in advanced pancreatic cancer by adding any agent to gemcitabine. The recommended dose of erlotinib with gemcitabine for this indication is 100 mg/d.

És tényleg...

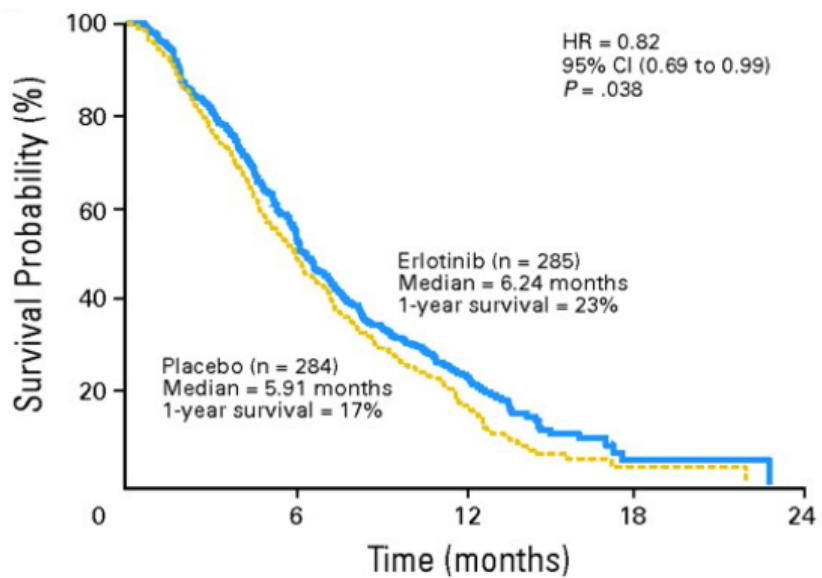
P = .038, adjusted for stratification factors

Nézzük csak meg közelebbről!

median 6.24 months v 5.91 months

Moore MJ, Goldstein D, Hamm J, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol.* 2007 May 20;25(15):1960-6.

## Egy példa: erlotinib a gemcitabine mellett



Moore MJ, Goldstein D, Hamm J, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol.* 2007 May;25(15):1960-6.

# Egy gondolat az információk terjedéséről

Miksad et al. JCO, 2007

Moore and colleagues [2] report a statistically significant overall survival benefit of 0.33 months for erlotinib plus gemcitabine compared with gemcitabine alone for first-line treatment of advanced pancreatic cancer (National Institute of Canada PA.3 trial). However, the small size of the survival gain and the increased risk of toxicities (diarrhea, interstitial lung disease, and treatment-related death) suggest that the clinical value of erlotinib may be limited.

Vincent et al. Lancet, 2011

In phase III clinical trials, the combination of gemcitabine and the epidermal growth factor receptor (EGFR) inhibitor, erlotinib, was modestly superior to gemcitabine alone. [134]

Schultz et al. PLoS ONE, 2016

PA.3 was a randomized phase III trial conducted by the NCIC Clinical Trials Group that demonstrated improved survival in patients treated with erlotinib plus gemcitabine compared to gemcitabine alone, in which, EGFR expression was not predictive of a response to erlotinib [6].

De!

De vigyázat: a nagyon kis hatások is fontosak lehetnek, ha nagyon sok embert teszünk ki nekik!

# *p*-érték és konfidenciaintervallum

- Egy másik fontos tanulsága ugyanennek: ne varázsoljanak el (teljesen) a *p*-értékek, ha cikket olvasunk!
- ...a hatásnagyság is fontos
- Ha már mindenképp a frekvencionista keretben maradunk, legalább annyit tegyük meg, hogy a *p*-értékek helyett a konfidenciaintervallumokat preferáljuk!
- A *p*-érték ugyanis összemossa ezt a két dolgot: a  $p = 0,02$  jelentheti azt is, hogy egy kis hatást kimértünk pontosan, meg azt is, hogy egy nagy hatást pontatlanul

# Tartalom

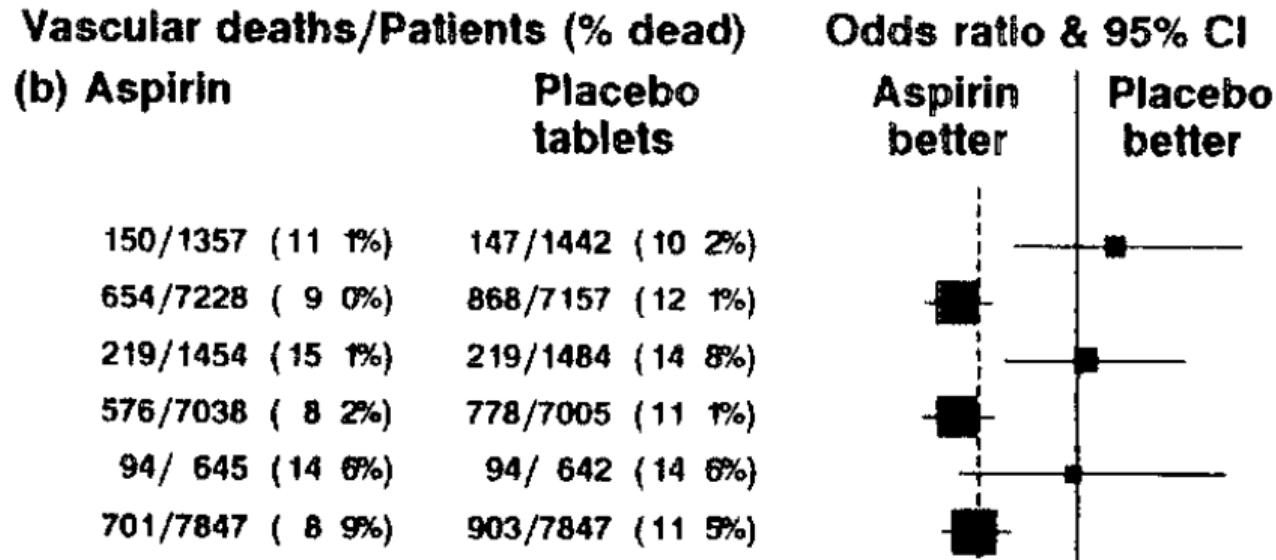
- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételei hiba, következtető statisztika
  - A következtető statisztikáról részletesebben
  - A legfontosabb félreértések a  $p$ -érték kapcsán
  - Klinikai és matematikai szignifikancia
  - Szignifikanciaavadászat és a többszörös összehasonlítások helyzete
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# A szignifikanciaavadászatról

- Zűrök lehetnek akkor, ha több hipotézisvizsgálatot végzünk és az eredményét vaglagosan értjük (azt mondom, hogy van hatás, ha bárhol találtam hatást): többszörös összehasonlítások helyzete
- $\alpha$ -infláció ( $\alpha_C = 5\%$ -nál  $\alpha_{E,2} = 9,75\%$ ,  $\alpha_{E,5} = 23\%$ ,  $\alpha_{E,50} = 92\%$ )
- Avagy hogyan befolyásolja az ember vérét a neve
- Az igazi gond, hogy ez nem csak rosszhiszeműség esetén kerül elő, tipikus példa: alcsoport-analízis

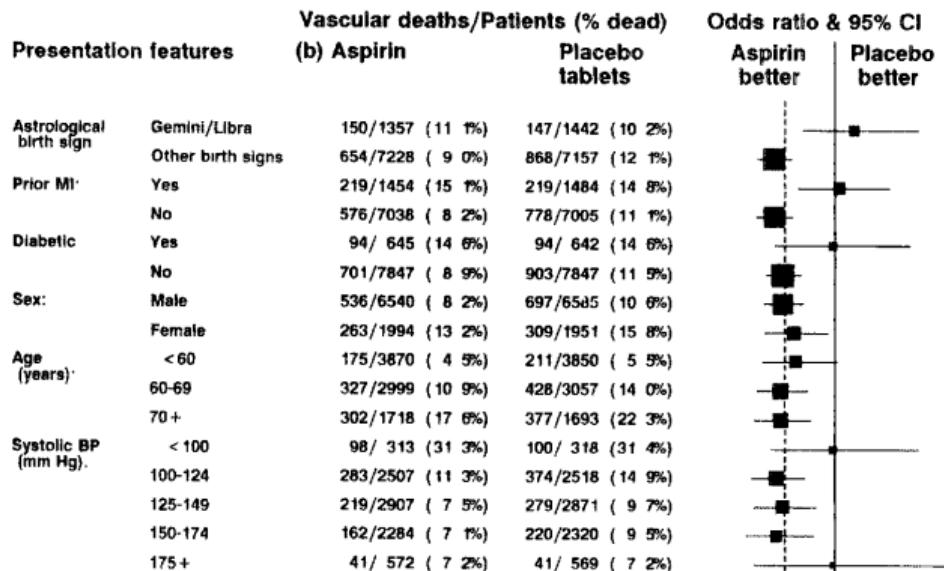
Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. Lancet. 2005 Apr 30-May 6;365(9470):1591-5.

# Ki az, aki azt hitte, hogy viccelelek a névvel?



Sleight P. Debate: Subgroup analyses in clinical trials: fun to look at - but don't believe them! Curr Control Trials Cardiovasc Med. 2000;1(1):25-27. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. Lancet. 2005 May 7-13;365(9471):1657-61.

# Ki az, aki azt hitte, hogy viccelelek a névvel?



Sleight P. Debate: Subgroup analyses in clinical trials: fun to look at - but don't believe them! Curr Control Trials Cardiovasc Med. 2000;1(1):25-27. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. Lancet. 2005 May 7-13;365(9471):1657-61.

# Multiplicitás: tipikus példák gyógyszervizsgálatokban

- Több kezelés
- Több végpont
- Több populáció (betegcsoport)
- Több időpontbeli vizsgálat (ez elvezet az interim analízisek kérdéséhez)

# Multiplicitás: mit tehetünk?

- Nem kell korrigálni, ha
  - Nem alapozunk rá állítást
  - Ha nem vagylagos a válaszunk ( minden végpontnak teljesülnie kell – ekkor persze az erő fog leromlani)
  - Megfelelő hierarchikus teszteléssel
- Enyhítés: megfelelő összevonás
- Statisztikai korrekciók

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételei hiba, következtető statisztika
- 4 A statisztikai modellek alapjai**
  - A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

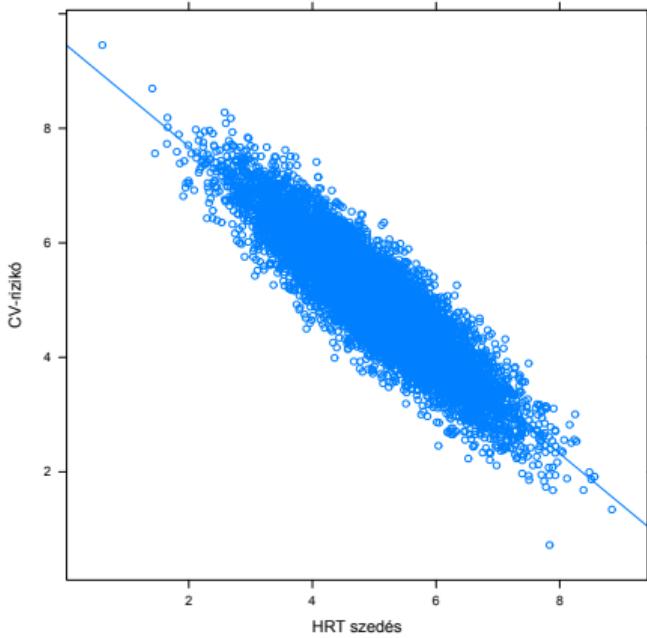
# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
  - A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések

# A statisztikai modellek alkalmazása

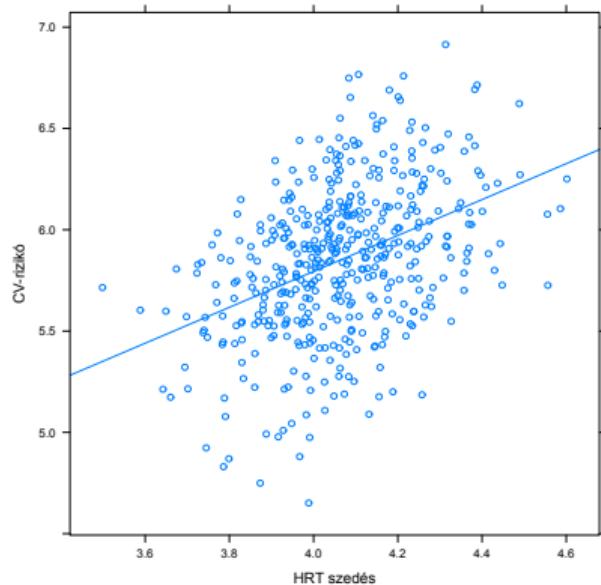
- A statisztika modellek rengeteg módon vezethetők be
- Én most úgy fogom tekinteni, mint egy eszközt a confounding kezelésére
- Nézzünk meg először egy – szimulált – példát!

# A confounding alaphelyzete



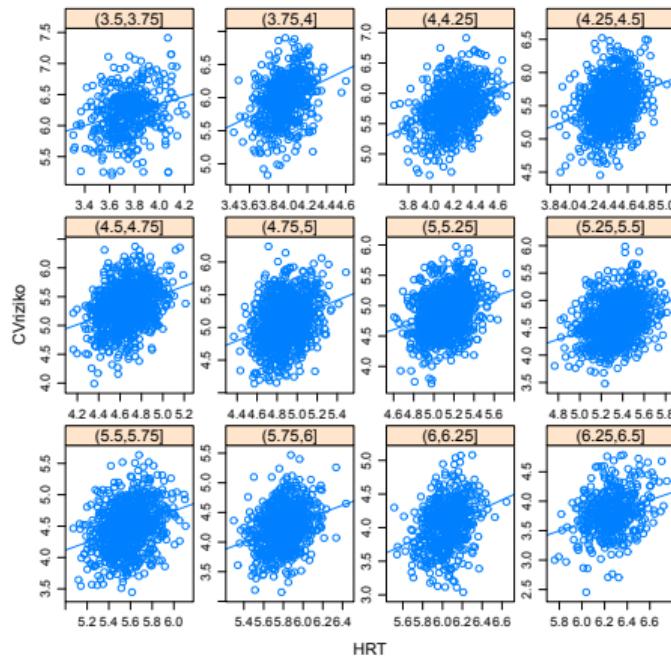
## Első kezelési lehetőség: rétegzés

A confounder szerint bontsuk meg – rétegezzük – a vizsgálatot; például 4 körüli (3,9 és 4,1) közötti SES-nél az összefüggés:



# Első kezelési lehetőség: rétegzés

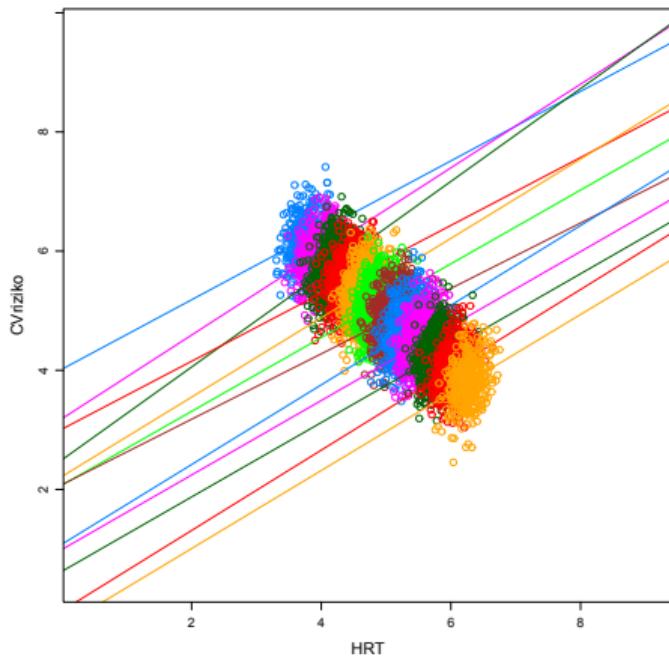
Az összes ilyen együtt:



# A confounding illusztrációja

Ez mellesleg a confounding jelenségét is jól illusztrálja! Még térlátás-igényesebb megoldás: ugyanez 3D-ben...

# A confounding illusztrációja és a rétegzés



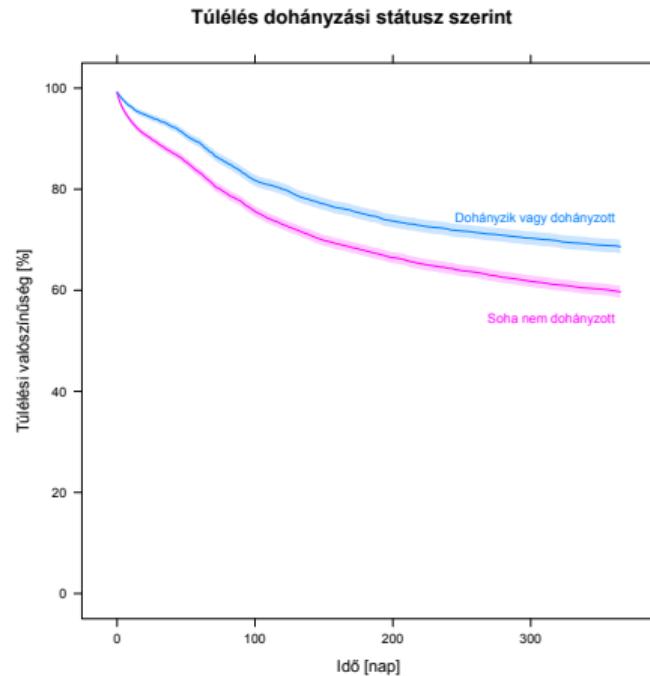
# Rétegzés a gyakorlatban

Igazából a Simpson-paradoxon is ez:

	Nyílt feltárás	Perkután eljárás
Kőátmérő < 2cm	93% (81/87)	87% (234/270)
Kőátmérő $\geq$ 2cm	73% (192/263)	69% (55/80)
Összességében	78% (273/350)	83% (289/350)

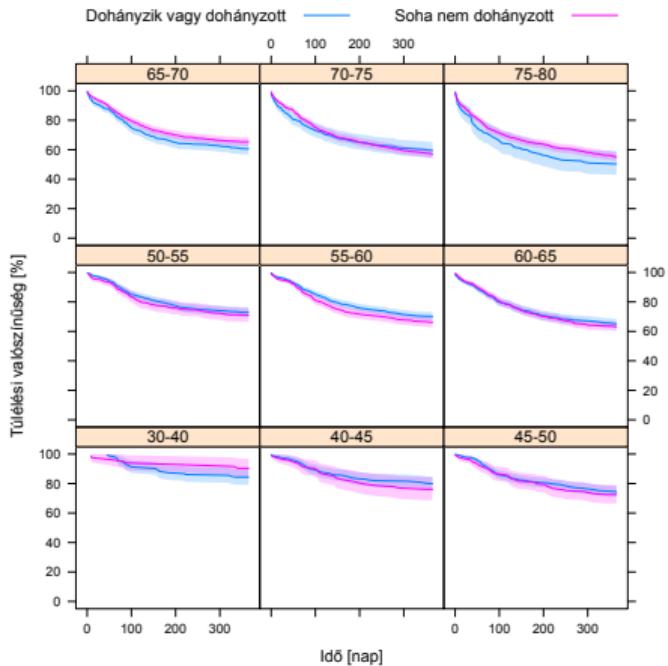
# Rétegzés a gyakorlatban

Végrehajtható a szívinfarktusos esetben is:



# Rétegzés a gyakorlatban

Végrehajtható a szívinfarktusos esetben is:

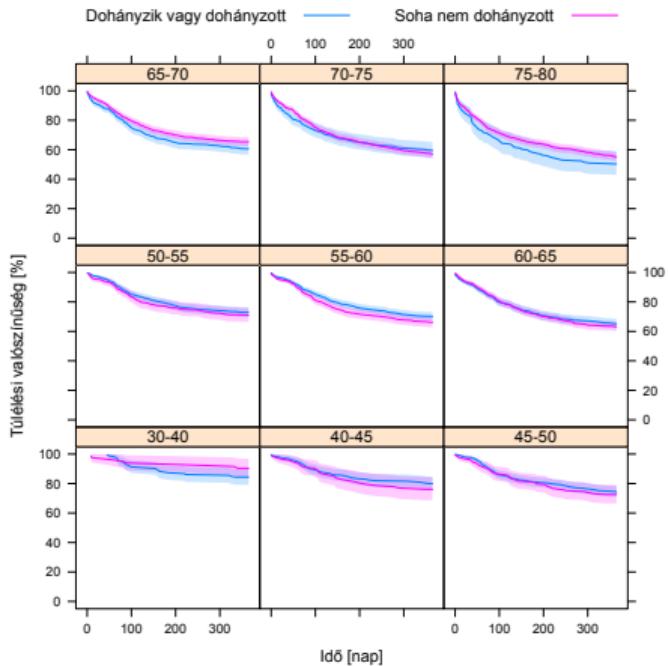


# Mi a baj a rétegzéssel?

- Noha ez sem tökéletes (nem lehet „adott” SES-nél vagy adott életkornál vizsgálni, össze kell vonni egy tartományt, egybemosva az ottani viszonyokat), egy confoundernél még tulajdonképpen különösebb gond nélkül alkalmazható
- De mi van akkor, ha több potenciális confounder van...?
- A rétegek („cellák”) száma *kombinatorikusan* nő, egy idő után kezelhetetlen lesz (gondoljunk a császármetszés vs. T1DM példájára)!
- (Nem is egyszerűen olyan értelemben, hogy sok lesz, hanem, hogy a mintaméréthez *képest* lesz sok!)
- Valamint: nehezen értelmezhető az eredmény (hogyan mondunk valamilyen kezelhető, kompakt információt abból, hogy a 12 rétegben mi az összefüggés?)

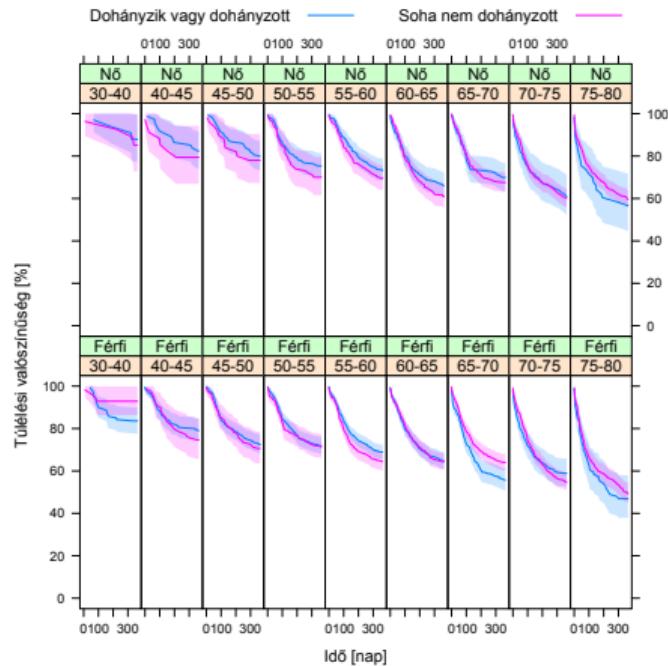
# Egy gyakorlati példa mindenre

Végrehajtható a szívinfarktusos esetben is:



# Egy gyakorlati példa mindenre

De a rétegek száma egy idő után kezd nagyon elszaladni...



## Egy megoldási lehetőség

- A rétegzés előnye, hogy *semmilyen* feltevéssel nem kell élnünk a változók kapcsolatáról
- De ha ez már nem működik, akkor egyszerűsíteni kell a helyzetet: éljünk valamilyen feltételezéssel arról, hogy hogyan függnek össze ezek a változók!
- (Lényegében: redukáljuk a paramétereink számát)
- Például: lineárisan függ a CV-rizikó a SES-től, és a HRT-szedéstől is:

$$\text{CVriziko} = \beta_0 + \beta_{\text{SES}} \text{SES} + \beta_{\text{HRT}} \text{HRT} + u$$

- Itt  $\beta_{\text{HRT}}$  a HRT-szedés szocioökonómiai státusztól *tisztított* hatása lesz (gondoljuk végig a ceteris paribus értelmezést)
- Valóban, az előbbi szimulált példán  $\widehat{\beta}_{\text{HRT}} = 0,89$ , míg ha a SES-t nem raktuk volna bele, akkor  $-0,90$

# A regressziós modell

- Ezt hívjuk regressziós modellnek
- Szétválogatja az egyes tényezők hatásait
- Több változós is teljesen hasonlóan kezelhető
- De a dolog nincs ingyen: beépítettük a feltevéseket, amiknek teljesülnie kell, hogy hiteles képet kapunk
- Például linearitás, nincs interakció

McNamee R. Regression modelling and other methods to control confounding. Occup Environ Med. 2005 Jul;62(7):500-6, 472.

# Folytonos eredményváltozó: lineáris regresszió (esettanulmány: CAP és HS)

- Egyváltozós vizsgálatok
- Confounding
- Többváltozós regresszió
- $\beta$ -k és értelmezésük, confounding elleni védekezés
- Nemlinearitás (spline-nal) és tesztelése
- Vizualizáció (forest plot és teljes tartományos)
- Modellszelekció és csapdái (nem prespecifikált modellek, automatikus modellszelekció)
- Túllilleszkedés
- Modell (belő) validálása és kalibrálása: split-sample (hold-out sample), keresztvalidáció, bootstrap
- Regularizált (penalizált) regresszió

# Kategoriális eredményváltozó: logisztikus regresszió (esettanulmány: antiarritmiás szerek paradox hatása)

Mint az előbbiek +

- Potenciális confounderekhez szükséges paraméterek számának csökkentése (blinded to the outcome!)
- OR-k és értelmezésük
- Cut-off, szenzitivitás, specificitás
- ROC-görbe, AUC
- Bináris logisztikus regresszió kiterjesztése több kategóriára (multinomiális és ordinális logisztikus regresszió)

# Time-to-event eredményváltozó: Cox-regresszió (esettanulmány: culprit ér szerepe AMI utáni túlélésben)

Mint az előbbiek +

- HR-k és értelmezésük
- Proporcionalitási feltevés

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
  - Rendszerszintű problémák
- 6 Filozófiai kérdések

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
  - Rendszerszintű problémák
- 6 Filozófiai kérdések

# Egyszerűen elszámolják, egyéb fundamentális hiba

Tesztstatisztikából lehet  $p$ -értéket számolva (az egyéb paraméterek ismeretében) és fordítva – vajon ugyanazt kapjuk?

## Egy 2001-es vizsgálat eredményei (Nature, BMJ!)

11.6% (21 of 181) and 11.1% (7 of 63) of the statistical results published in Nature and BMJ respectively during 2001 were incongruent, probably mostly due to rounding, transcription, or type-setting errors. At least one such error appeared in 38% and 25% of the papers of Nature and BMJ, respectively. In 12% of the cases, the significance level might change one or more orders of magnitude.

García-Berthou E, Alcaraz C. Incongruence between test statistics and P values in medical papers. BMC Med Res Methodol. 2004 May 28;4:13.

# Egyszerűen elszámolják, egyéb fundamentális hiba

- Ilyen mindenig is volt, és mindenig is lesz...
- ...csakúgy, mint egyéb drasztikus hibák (minta összekeveredése, elcserélése, hibás kódolás stb.), minden rosszhiszeműség nélkül is
- Ez egy – nagyon pici, de nem nulla – kockázat, ha bármilyen cikkre építünk
- Ha kiderül, visszavonják (hasonlóan a rosszhiszemű csalásokat is)
- Ez önmagában még nem „rendszerszintű” hiba, akkor válik azzá, ha erről a visszavonásról nem értesülünk, illetve ha az előtt keletkezett citációt felhasználunk
- Egy jó példa: <http://retractionwatch.com/>

# Interpretáció kérdése

Comparison of two systematic reviews of albumin for fluid resuscitation

	Wilkes, 2001w1	Alderson, 2002w2
Funding body	Plasma Proteins Therapeutic Association	NHS
Trials included	42 short term trials reporting mortality	31 short term trials reporting mortality
Overall difference in mortality between albumin and crystalloid (relative risk, 95% CI)	1.11 (0.95 to 1.28)	1.52 (1.17 to 1.99)
Difference in burns patients	1.76 (0.97 to 3.17)	2.4 (1.11 to 5.19)
Interests of funding body	Promotes access to and reimbursement for the use of albumin	Pays for use of albumin
Authors' conclusions	Results "should serve to allay concerns regarding the safety of albumin"	Recommend banning use of albumin outside rigorously conducted randomised controlled trials

Egy radikális javaslat: el se olvassuk a diszkussziót?

Montori VM, Jaeschke R, Schünemann HJ, Bhandari M, Brozek JL, Devoreaux PJ, Guyatt GH. Users' guide to detecting misleading claims in clinical research reports. BMJ. 2004 Nov 6;329(7474):1093-6.

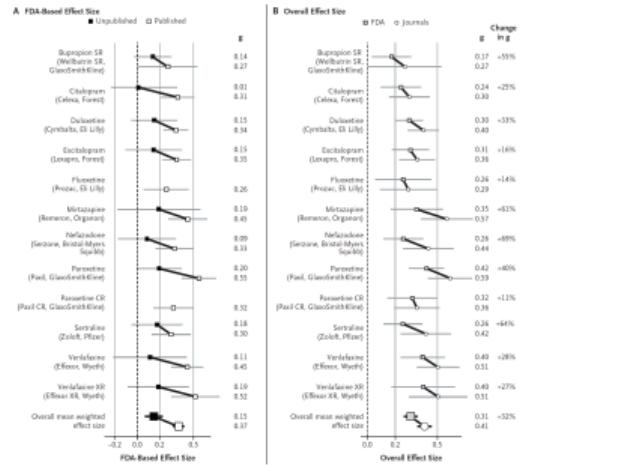
## Publikációs torzítás

Egyetlen példa (Turner et al, 2007): összes antidepresszáns anyagai az FDA-nél vs. publikálva, egy 15 éves intervallum alatt – 74 kutatásból 23-at soha, sehol nem publikáltak; a 74 kutatásból 38 volt pozitív kimenetű, ebből 37-et közöltek, a 36 nem pozitív kimenetűből viszont 3-at

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med. 2008 Jan 17;358(3):252-60.

# Publikációs torzítás

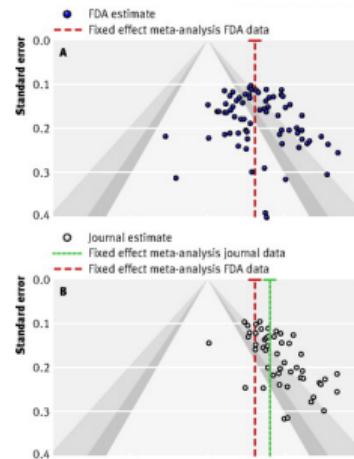
Ennek hatása:



Turner EH, Mattheus AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med. 2008 Jan 17;358(3):252-60.

# Publikációs torzítás megítélése

Antidepresszánsok (az FDA-nél és újságban publikálva):



Moreno SG, Sutton AJ, Turner EH, et al. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. BMJ. 2009 Aug 7;339:b2981.

# A funnel plot alkalmazása és problémái

- Igazából azt méri, hogy a kis kutatások eredményei eltérnek-e a nagyokétől
- Az aszimmetria megítélése szubjektív lehet
  - Egger-teszt, Rosenthal „fail-safe-N”, Trim-and-Fill...?
- Sok szubjektív választás van, ami befolyásolhatja a végeredményt

Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011 Jul 22;343:d4002.

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997 Sep 13;315(7109):629-34. Rosenthal R. The file drawer problem and tolerance for null results. Psychological bulletin. 1979; 86(3): 638-641. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000 Jun;56(2):455-63. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA. 2006 Feb 8;295(6):676-80. Peters JL, Sutton AJ, Jones DR, et al. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol. 2008 Oct;61(10):991-6. Lau J, Ioannidis JP, Terrin N, et al. The case of the misleading funnel plot. BMJ. 2006 Sep 16;333(7568):597-600.

# Kísérletek regisztráltsága

- Kísérlet regisztráltsága – természetesen előzetes regisztráltsága – kísérleti regiszterben (pl. <https://clinicaltrials.gov/> vagy <https://eudract.ema.europa.eu/>):
  - Kevésbé lehessen eltitkolni a nem kedvező eredményeket
  - Lehessen ellenőrizni, hogy tényleg azt csinálták-e, amit elterveztek
- Ez utóbbi legfontosabb oka: outcome switching
- COMParé projekt: <http://compare-trials.org/>
- Egy fontos kezdeményezés: <http://www.alltrials.net/>
- (Az utóbbi években sokat javult a helyzet; de az információhoz való hozzáférés továbbra is igen problémás)

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések
  - $p$ -érték, hipotézistesztelés, és ellentmondásaik
  - A statisztika határai

# Tartalom

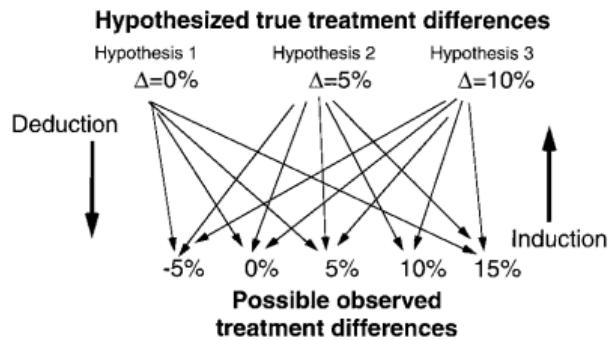
- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések
  - *p*-érték, hipotézistesztelés, és ellentmondásai
  - A statisztika határai

# Következtetéselmélet

- A megfigyelt világ és a tudásunk összekapcsolása
- Deduktív következtetés: kiindulunk egy hipotézisből (amit a tudásunk alapján fogalmaztunk meg), és ebből következtetünk arra, hogy mit fogunk látni a világban
  - Objektív (abban az értelemben, hogy minden igaz, ha a kiinduló hipotézis igaz volt)
  - Nem alkalmazható tudásunk bővítésére
- Induktív következtetés: az alapján, hogy mit látunk a világban, következtetünk arra, hogy mely hipotézis lehet igaz
  - A bizonyíték fogalma
  - Új tudásra tudunk szert tenni
  - Ám ez soha nem tud biztos lenni
- Számos kísérlet ennek feloldására (pl. Popper és a falszifikáció → megspórolható az indukció; ezt valósítja meg a hipotézisvizsgálat is, minden iskola)

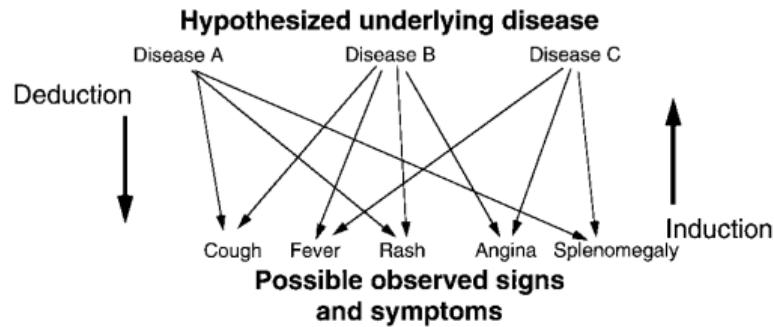
Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Statisztikai következtetéselmélet



Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# „Orvosi következtetéselmélet”



Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

## Két iskola ütközése: R. A. Fisher

- $p$ -érték: „measure of evidence”
- Mégpedig *egyetlen* hipotézisre vonatkozóan (tehát nincs ellenhipotézis)
- *Egyetlen* kísérletből akarunk dolgozni (nincs „képzeletbeli ismételt mintavétel”)
- *Nem* akarunk külső információt felhasználni (nincs prior valószínűség)
- *Nem* hibavalószínűség
- Hogyan is lehetne valószínűség → ahhoz több kísérletet kellene, legalább képzeletben, tekintenünk (ahogy annak a kijelentésnek, hogy „ez az érme 40% valószínűsséggel ad fejet” is csak úgy van értelme, ha többször feldobhatjuk – ez a valószínűség frekvencionista értelmezése)
- Inkább informális index, esetleg az a szabály, de ez sem mechanikusan, hogy adott  $p$  alatt elutasítunk: „significance testing” (*nem* „hypothesis testing”!)

## Két iskola ütközése: J. Neyman és E. Pearson

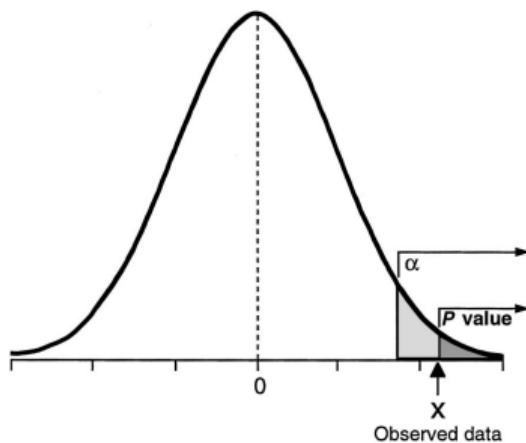
- Döntés: „hibavalószínűség”
- Két hipotézis között döntünk
- Kísérletek sorát tételezzük fel...
- ...és a célunk, hogy *hosszú távon* biztosítsunk adott első- és másodfajú hibaarányt
- Egy adott, konkrét hipotézisről úgysem dönthető el, hogy mennyire igaz, a kérdés csak hosszú távon értelmezhető
- Informális, szubjektív index helyett objektív, döntési szempontú („viselkedés, nem következtetés”) megközelítés
- Elfogadási és elutasítási tartomány, *mindegy* hogy azon belül hová esik az empirikus tesztstatisztika (nincs bizonyíték-erősség)
- A határt eredetileg még úgy gondolták, hogy a kétféle hiba súlya alapján határozzák meg a kutatók
- „Hypothesis testing” (*nem „significance testing”*)

# A két iskola inkonzisztenciája

- Döntési határ van (és mindegy, hogy hová esik az empirikus tesztstatisztika) vagy nincs határ, és egy folytonos mértékünk van a bizonyíték erősségeire?
- „Az érdekel minket, hogy az adott gyanúsított ártatlan-e, vagy az, hogy hosszú távon csak kevés ártatlant küldünk börtönbe?”
- A legfontosabb baj: nem lehet egyszerre beállítani a hosszútávú hibavalószínűséget és egyúttal az egyedi eredmény bizonyítőerejéről is nyilatkozni!
- A jelenlegi orvosi statisztikai gyakorlat azért keveri össze – teljesen inkonzisztensen – a kettőt, hogy megteremtse ennek az *illuszióját*

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004. Goodman SN. p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. Am J Epidemiol. 1993 Mar 1;137(5):485-96; discussion 497-501. Gigerenzer G. The superego, the ego, and the id in statistical reasoning. In: A handbook for data analysis in the behavioral sciences: Methodological issues. 1993:311-339. Lehmann EL. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?. In: Selected Works of EL Lehmann. 2012:201-208.

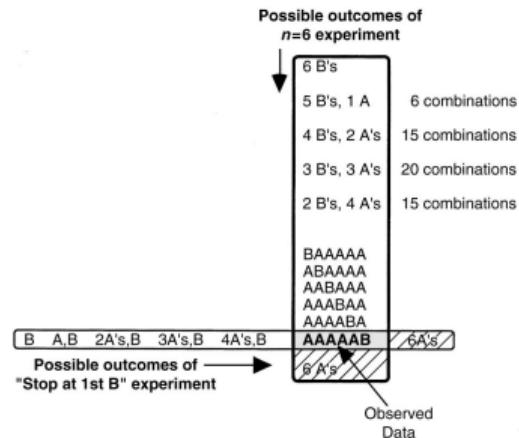
# A kompromisszum elkerülésének illúziója



Úgy tűnhet, hogy a  $p$ -érték is egyfajta hibavalószínűség  
„15. legjobb vagyok az évfolyamban”

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# A kompromisszum elkerülésének illúziója



Hosszútáv vs. rövidtáv

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

# Tartalom

- 1 Az empirikus orvosi kutatások alapvonalai
- 2 Konkrét vizsgálati módszerek: előnyök, hátrányok, alkalmazási területek
- 3 Mintavételi hiba, következtető statisztika
- 4 A statisztikai modellek alapjai
- 5 Rendszerszintű problémák
- 6 Filozófiai kérdések
  - $p$ -érték, hipotézistesztelés, és ellentmondásaik
  - A statisztika határai

# Korábbi információ vs. kontextus

- A legtöbb orvosi cikkben kevés szerepet kapnak a korábbi eredmények, biológiai megfontolások stb. (maximum kvalitatíve a diszkusszióban)
- A filozófiai problémákból adódó illúzió: tudunk döntést hozni csak az adott mintára támaszkodva...?
- Tipikus helyzet: a konklúziót már a diszkusszió legelső mondatában levonják...
- ...és az igazából nem más, mint egy  $p$ -érték „nyelvi kifejtése”
- Út a bayes-i iskola felé (prior: ún. szubjektív komponens, Bayes faktor: adat („bizonyíték”), poszterior: hibavalószínűség)

Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med. 1999 Jun 15;130(12):995-1004.

## Egyedi beteg kezelése vs. átlagos beteg kezelése

Az empirikus vizsgálat látott módszertana megköveteli a csoportosítást (ezt már Louis is felismerte), az átlagra koncentrálunk

- Ez *inherensen* így van, a mintavételi ingadozás – ha csak nincs időgépünk – kizárolag így csökkenthető
- Persze, már ott is úgy fogalmaztunk, hogy „kiátlagolódnak” az ingadozások
- Azaz: az ingadozást – teljesen szándékoltan – *eltüntetni* akarjuk, nem *építeni* rá
- Inherensen ellentmond a mintavételi ingadozás csökkentése és az egyénre következtetés

# Egyedi beteg kezelése vs. átlagos beteg kezelése

- Ez viszont szemben áll azzal, hogy az orvosnak *adott* beteget kell kezelnie
- (Persze nyilván van módszertani kezelési lehetőség is, kezdve a már tárgyalt alcsoport-analízisekkel)
- De a probléma ott marad (személyre szabott orvoslás (???) vs. (?) bizonyítékokon alapuló orvoslás?)

Bereczki D. Personalized medicine: a competitor or an upgrade of evidence-based medicine? Personalized Medicine. 2012; 9(2): 211-221. de Leon J. Evidence-based medicine versus personalized medicine: are they enemies? J Clin Psychopharmacol. 2012 Apr;32(2):153-64. Kumar D. From evidence-based medicine to genomic medicine. Genomic Med. 2007;1(3-4):95-104. Goldberger JJ, Buxton AE. Personalized medicine vs guideline-based medicine. JAMA. 2013 Jun 26;309(24):2559-60.

## Például: „n-of-1 trial”

- Mintha mégis lenne időgéünk: ugyanazon a betegen próbáljuk ki többször!
- Annyiban nem újdonság, hogy az orvosok máskor is szokták „beállítani” az beteg gyógyszerelését, de itt ezt egy klinikai kísérlet szigorúságával tesszük
- (Persze bejön az a feltételezés, hogy ezek egymástól független, „tiszta lappal indulunk”, esetleg wash-out periódus emiatt)
- Sajnos csak akkor működik, ha a betegség 1) krónikus 2) olyan kezelés van ellene, ami belátható időn belül mérhető változást hoz
- De ha mindenek megvalósulnak, akkor mégiscsak tudunk *adott* betegre következtetni

Lillie EO, Patay B, Diamant J, et al. The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Per Med. 2011 Mar;8(2):161-173. Guyatt GH, Keller JL, Jaeschke R, et al. The n-of-1 randomized controlled trial: clinical usefulness. Our three-year experience. Ann Intern Med. 1990 Feb 15;112(4):293-9. van der Greef J, Hankemeier T, McBurney RN. Metabolomics-based systems biology and personalized medicine: moving towards n = 1 clinical trials? Pharmacogenomics. 2006 Oct;7(7):1087-94.