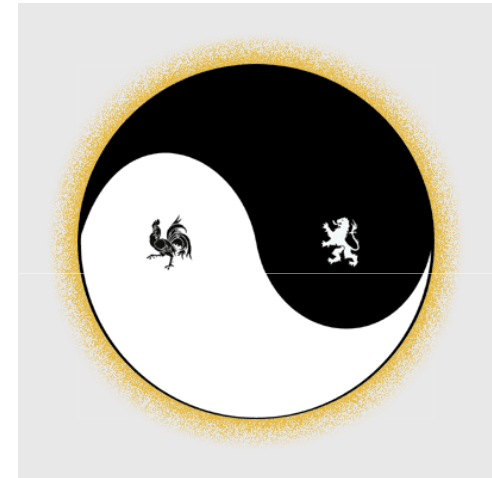


# Adatredukció, adatok bemutatásának lehetőségei, a többszörös elemzés helye az orvosi gondolkodásban és a biostatistikában



Dr. Prohászka Zoltán  
*Az MTA doktora*

Semmelweis Egyetem

Belgyógyászati és Hematológiai Klinika

2023-02-23

[Prohaszka.zoltan@med.semmelweis-univ.hu](mailto:Prohaszka.zoltan@med.semmelweis-univ.hu)

- Statistica szoftver
- Zoom link
- Honlap, házifeladat, tematika
  - [https://semmelweis.hu/kutlab/oktatas/klinikai biostatiztika 2023/](https://semmelweis.hu/kutlab/oktatas/klinikai_biostatiztika_2023/)
- Vizsga

# A véletlen szerepe az új, eredeti megfigyelésekben



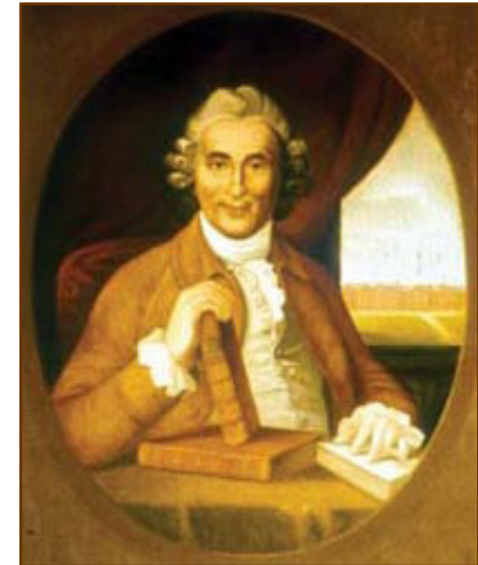
*The scene in the dining room of 52 Queen Street, Edinburgh on 4 November 1847, as Simpson, together with colleagues, Dr Matthews Duncan and Dr George Keith, discovers the effects of inhaling chloroform. Reproduced by kind permission of the Wellcome Institute for the History of Medicine.*

# Valóban hatékony?

Az első klinikai kísérlet, XVIII sz.:

„véletlen besorolásos” tanulmány a skorbut kezelésével kapcsolatban

- James Lind (1716-1794)
- *Treatise of scruvy* (1753)
- *Essay on the most effectual means of preserving the health of seamen* (1762)



## A dokumentált klinikai vizsgálatok kezdete: XIX sz. eleje

- *Pierre Charles Alexandre Louis*  
(1787-1872)
    - Előtérbe helyezte a betegségek tüneteinek értékelésén nyugvó diagnosztikát a korábbi, a betegek panaszain alapuló módszerhez képest
    - Az orvosi matematikát bevezette a gyakorlatba, és a betegek adatainak összegzett értékelése mellett tett hitet
    - Felismerte, hogy a klinikai adatok utólagos értékeléséből új ismeret, eredmény származhat
    - Bebizonyította az érvágás hatástalanságát pneumonia esetén
      - „Kutatások az érvágás hatékonyságával kapcsolatban egyes gyulladós megbetegedésekben” (1836)
    - Tanítványa: Skoda....
      - „Statisztikai-numerikus táblák”
    - ...Semmelweis
      - Két osztály halálozási statisztikáinak összevetése, trendek felismerése
- [http://mek.oszk.hu/05400/05427/pdf/Semmelweis\\_statisztika.pdf](http://mek.oszk.hu/05400/05427/pdf/Semmelweis_statisztika.pdf)



## I. TÁBLÁZAT.

Orvosok osztálya				Bábák osztálya		
<i>Év</i>	<i>Szülések száma</i>	<i>Halottak száma</i>	<i>A szülések száma úgy aránylik a halottak számához, mint 100 a</i>	<i>Szülések száma</i>	<i>Halottak száma</i>	<i>A szülések száma úgy aránylik a halottak számához, mint 100 a</i>
1841	3036	237	7·7-hez	2442	86	3·5-hez
1842	3287	518	15·8 «	2659	202	7·5 «
1843	3060	274	8·9 «	2739	164	5·9 «
1844	3157	260	8·2 «	2956	68	2·3 «
1845	3492	241	6·8 «	3241	66	2·03 «
1846	4010	459	11·4 «	3754	105	2·7 «
Összeg	20042	1989	9·92-hez	17791	691	3·38-hez

*Semmelweis Ignác Fülöp: A gyermekági láz kóroktana, fogalma és prophylaxisa, 1861*

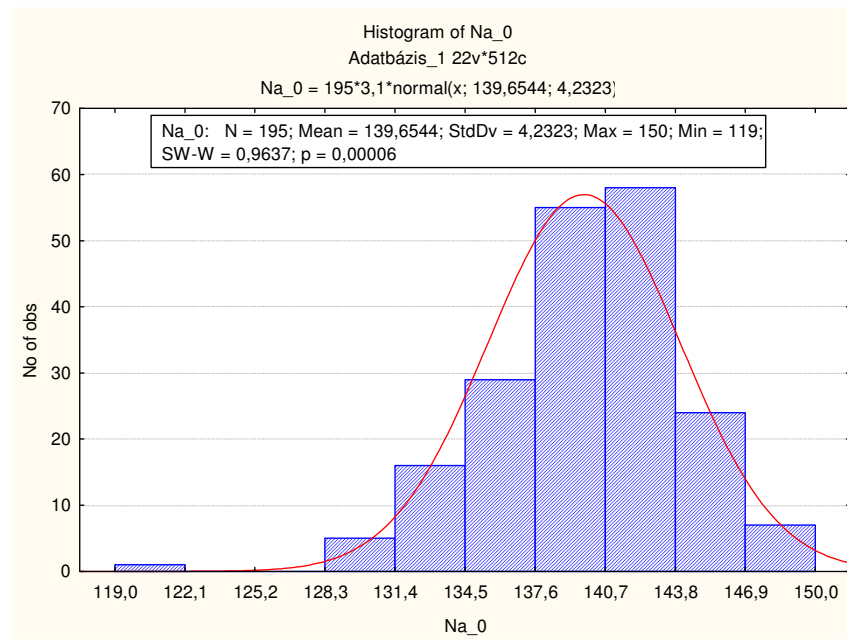
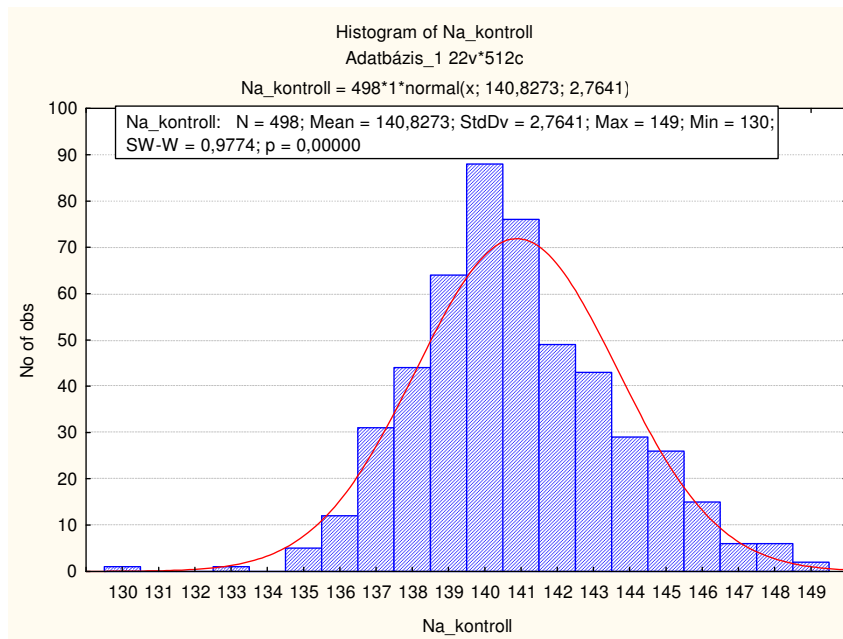
# Leíró statisztikák: folytonos változók

- Mintaterjedelelem (min-max)
- Átlag (matematikai átlag)
- Szórás
  - Adataink szétszórtsága csak akkor hasonlítható egy másik adathalmaz szóródásához, ha azonos skálán (módszer, mértékegység) történt a mérés

# Az adatok bemutatásának, összegzésének és ábrázolásának módjai

- A változó típusa és eloszlása döntő a választandó eljárás vonatkozásában
  - Használjuk a szórást (SD) nem ferdült, szimmetrikus eloszlású, folytonos változók esetén (az átlaggal).
  - Használjunk percentiliseket (leggyakrabban interkvartilis tartományt) a ferdült, nem szimmetrikus eloszlású numerikus és ordinális változók esetén (a mediánnal).
- „Korfa” jellegű ábrázolás: ordinális (kategória) változóra
- Skála-független összegzési, bemutatási módszerek
  - A variációs koefficiens használata akkor indokolt a minta szétszórtságának bemutatására, ha eltérő skálán mért értékek szóródását kívánjuk összehasonlítani (pl. azonos molekula mérése eltérő módszerekkel).
  - Ráták

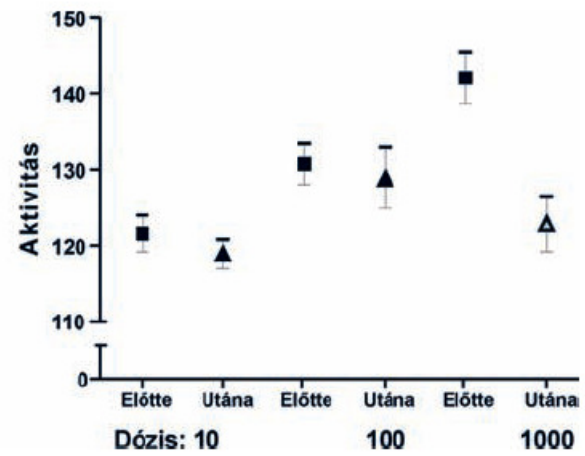
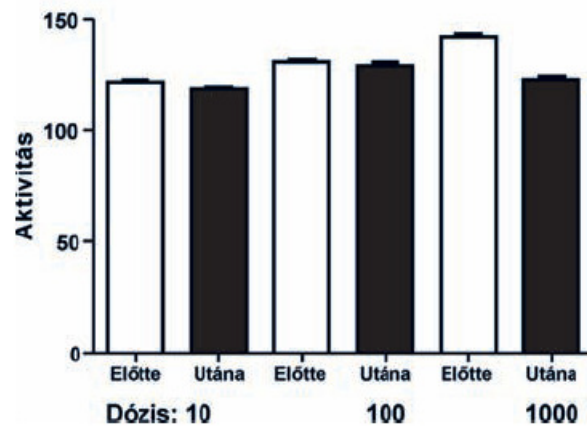
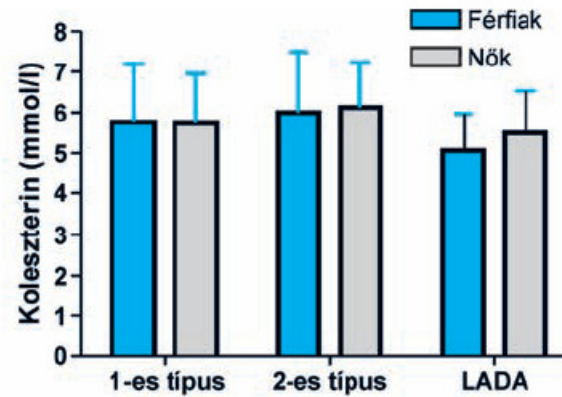
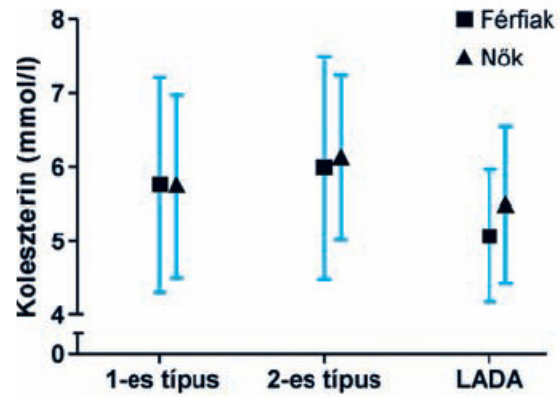


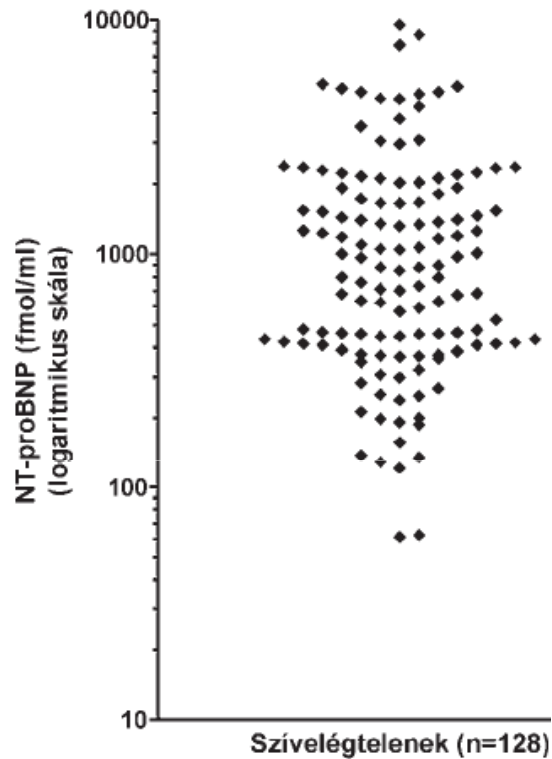


	<b>Valid N</b>	<b>Mean</b>	<b>Median</b>	<b>Std.Dev.</b>
<b>Na HF</b>	195	139,6544	140,0000	4,232340
<b>Na kontroll</b>	498	140,8273	141,0000	2,764080

Descriptive Statistics (Adatbázis\_1)

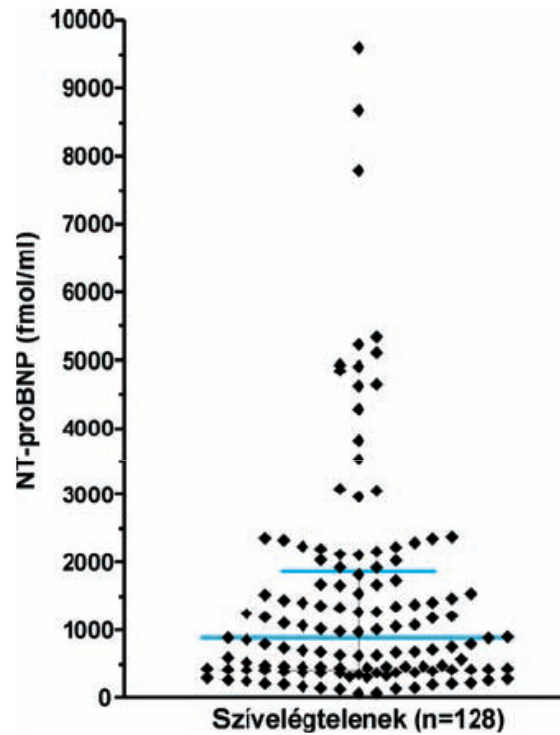
	<b>Minimum</b>	<b>Percentile - 10,00000</b>	<b>Lower - Quartile</b>	<b>Median</b>	<b>Upper - Quartile</b>	<b>Percentile - 90,00000</b>	<b>Maximum</b>
<b>Na_0</b>	119,0000	134,0000	137,0000	140,0000	142,0000	144,0000	150,0000
<b>Na_kontroll</b>	130,0000	137,0000	139,0000	141,0000	143,0000	145,0000	149,0000





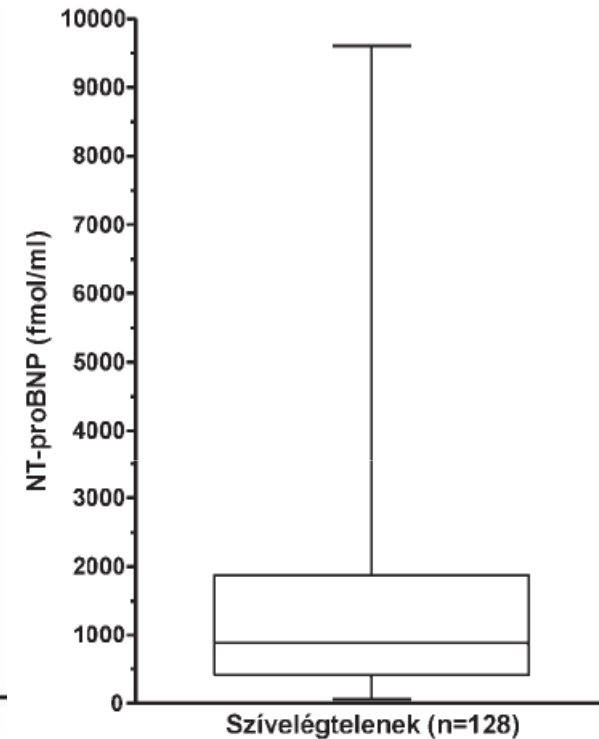
10.2. ábra

A szívégtelenségben szenvedők NT-proBNP szintjének bemutatása logaritmus skálán.



10.3. ábra

A szívégtelenségben szenvedők NT-proBNP szintjének bemutatása, a vízszintes vonal a medián, az alatta a 25-75-ös percentilisek értékét jelzik. Az adatok széles terjedelmet mutatnak.



10.4. ábra

A szívégtelenségben szenvedő betegek NT-proBNP szintje box-and-whiskers ábrázolási móddal. A vízszintes vonal a mediánt, a téglalap alsó (25) és felső (75) élei a percentiliseket, míg a pálcikák a terjedelmet jelölik.

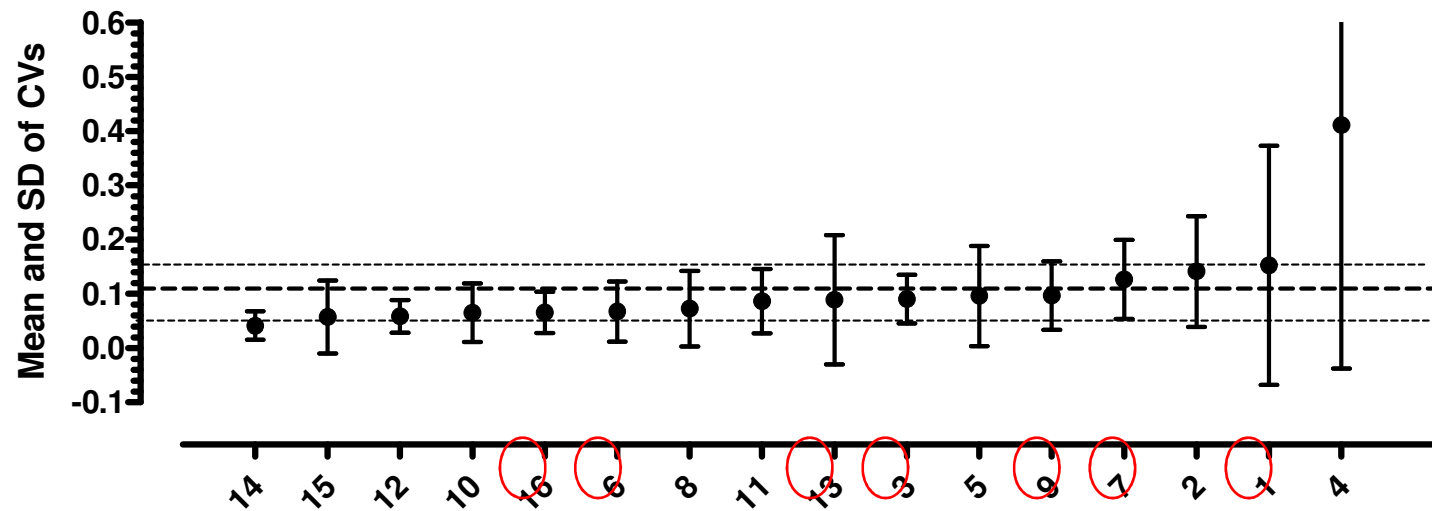
# A variációs koefficiens

- A szórás százalékos aránya az átlaghoz viszonyítva, azaz a relatív hiba. (SD/mean)
- Fő előnye, hogy dimenzió nélküli szám, így bármely adathalmaz variációs koefficiense összehasonlítható.
- Feltéve, ha az adathalmaz normál eloszlású
  - Transzformálás

# Relatív hiba a laboratóriumokban: a minőség egyfajta mérése (Precision)

- Azonos paraméterek mérése különböző laboratóriumokban, eltérő módszerekkel

Mean of CV among the laboratories: 0,11;  
95% CI 0,06-0,15



# Diszkrét eloszlású változók összegzése, bemutatása

10 tábor, két hónap

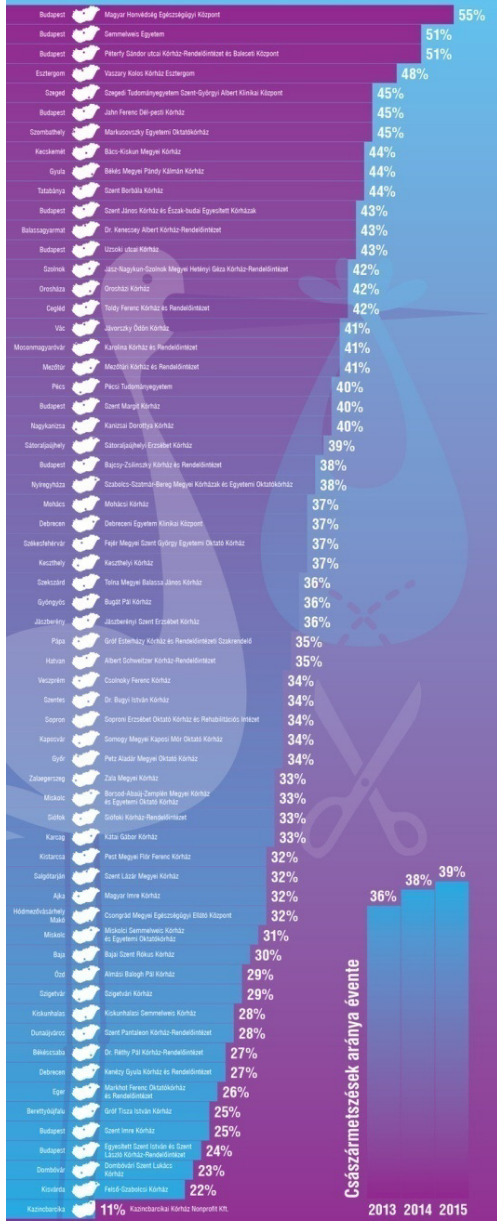
- Hasmenéses esetek száma: 35 (egy táborban)

Az incidencia és a prevalencia kiszámítását bemutató példa

	Egészséges	Megbetegedett	Összes táborozó
Összes táborban	1310	35	1345
	Incidencia:	$[(35/1345) \cdot 1000]/2$	13/1000 táborozó/hónap
Érintett táborban	110	35	145
	Prevalencia:	35/145	0,24

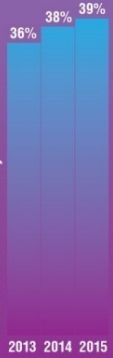
- Prevalencia (elterjedtség): Arány, viszonyszám
  - (Beteg/összes rizikónak kitett egyén)
- Incidencia = prevalencia ráta: alappal való osztás
  - (pl. halálozás/1000 ember/év, esemény/10.000 ember/év)

Császármetszések aránya a magyar kórházakban  
a 2013-as, 2014-es és 2015-ös születek alapján (átlag)



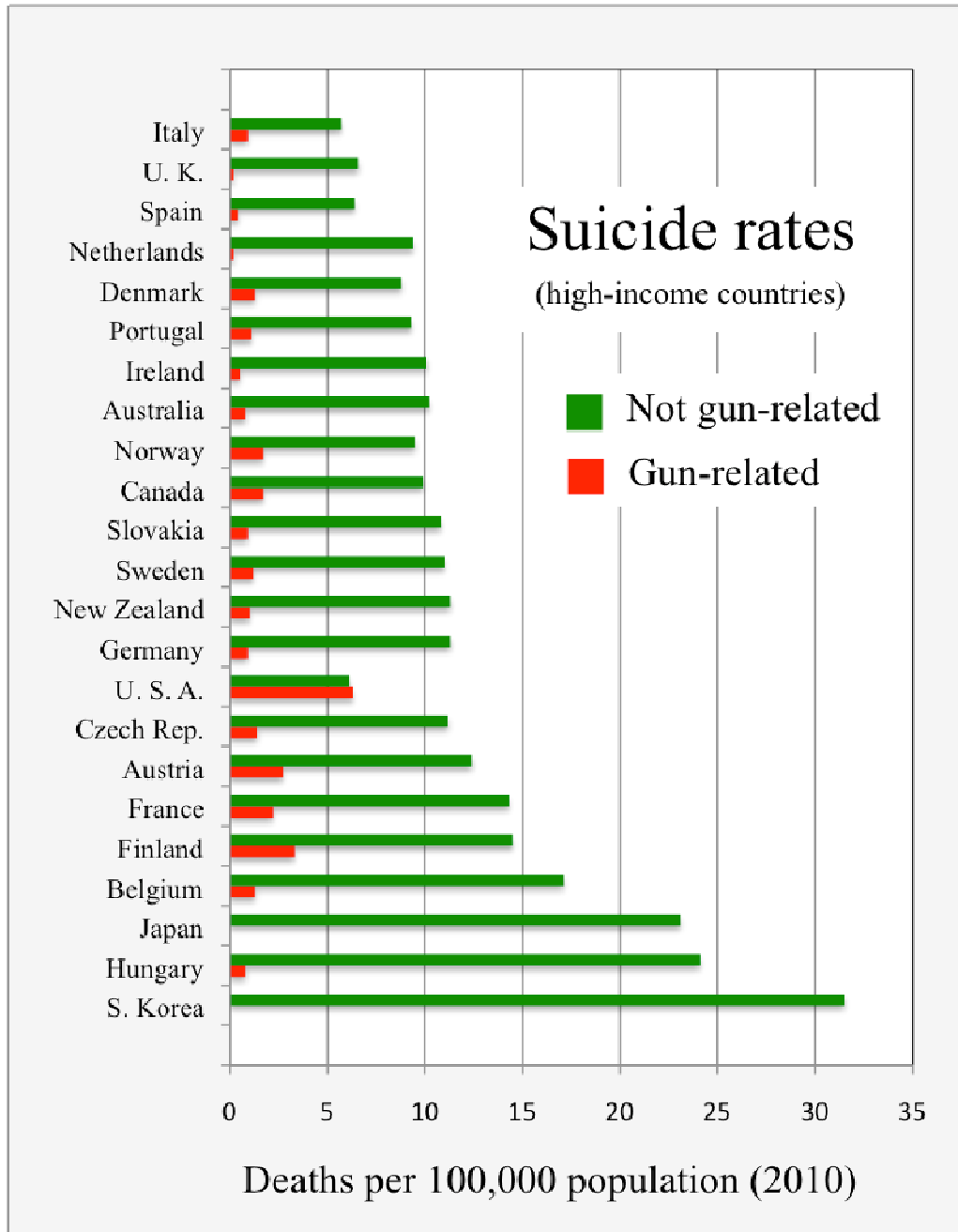
# Prevalencia

Császármetszések aránya évente





# Incidencia



# Az esemény ráta

Hipotetikus adatok a halálozási ráta direkt adjusztálásához

	Megyei kórház				Egyetemi klinika			
	Élve születések száma		Halálozás		Élve születések száma		Halálozás	
Születési súly	n	%	n	Ráta*	n	%	n	Ráta*
Összesen	2000		30	15,0	2850		61	21,4

\*Halálozások száma/1000 élveszületés/év

# Az esemény ráták adjusztálása

- Adjusztálás, vagyis illesztés, annak a kérdésnek a megválaszolása, hogy
  - *„Hogy alakulna a császármetszések aránya az egyes kórházakban, ha mindenhol azonos korú, kísérőbetegségben szenvedő, hasonló társadalmi helyzetű....stb. asszonyok szülnének?”*
  - *„Hogy alakulna a halálozás az egyetemi klinikán, ha pontosan ugyanolyan arányban születnének babák az egyes súly-kategóriákban, mint a megyei kórházban?”*
- Direkt és indirekt módszer
- Az egyes születési súly-kategóriákba eső esetszámokat mint súlyszámokat használva lehet kiszámítani az adjusztált rátákat
- Az elv megegyezik a többváltozós modelleknél alkalmazott adjusztálással

# Adjusztálás: direkt módszer

## Eltérő előfordulási arány

Hipotetikus adatok a halálozási ráta direkt adjusztálásához

	Megyei kórház				Egyetemi klinika			
	Élve születések száma		Halálozás		Élve születések száma		Halálozás	
Születési súly	n	%	n	Ráta*	n	%	n	Ráta*
<1500 g	30	1,5	6	200	94	3,3	19	202,1
1500 –2499 g	140	7	11	78,6	260	9,1	22	84,6
≥ 2500 g	1830	91,5	13	7,1	2496	87,6	20	8,0
Összesen	2000		30	15,0	2850		61	21,4

\*Halálozások száma/1000 élveszületés/év

- Referencia populáció: megyei kórház
- $[(202,1 \times 30) + (84,6 \times 140) + 8 \times 1830] / [30 + 140 + 1830] = 16,3$
- Az egyetemi klinikán a születéssúly-megoszlásra adjusztált halálozási ráta 16,3/1000 élveszületés /év lenne
- (Propensity score matching)

# Adjusztálás: indirekt módszer

## 9.6. táblázat

Hipotetikus adatok a halálozási ráta indirekt adjusztálásához

Születési súly	Születések száma		Specifikus halálozási ráta* egy referenciapopulációban
	Megyei kórház	Egyetemi klinika	
<1500 g	30	94	150
1500–2500 g	140	260	80
>2500 g	1830	2496	7
Összhalálozás (n)	30	61	

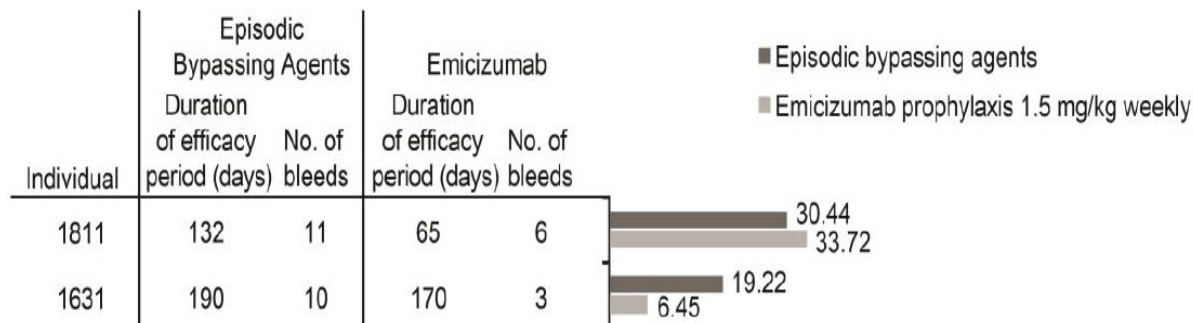
\*Halálozási ráta/1000 élveszületés/év

- Indirekt adjusztálásra akkor van szükségünk, ha nem rendelkezünk részletes adatokkal
- Ekkor egy irodalmi referenciához illeszthetjük saját megfigyeléseinket

# Ráták számítása eltérő követési idő esetén: annualizált ráta

- Esemény:
  - Egyszeri, jól definiálható (halálozás)
  - Többszörös, jól definiálható (relapszus)
  - Többszörös, minőségi (vérzés: trigger, mennyiség, lokalizáció, kimenetel...)
- Időszak:
  - Naptárhoz rögzített, fix
  - Random, fix időszakra kivetített (pl. annualizált)

**Figure S4.** Intra-Individual Comparison for Treated Bleeds in Participants Receiving Emicizumab Prophylaxis (Arm A) Versus No Previous Prophylaxis (Only Episodic Bypassing Agent Treatment).

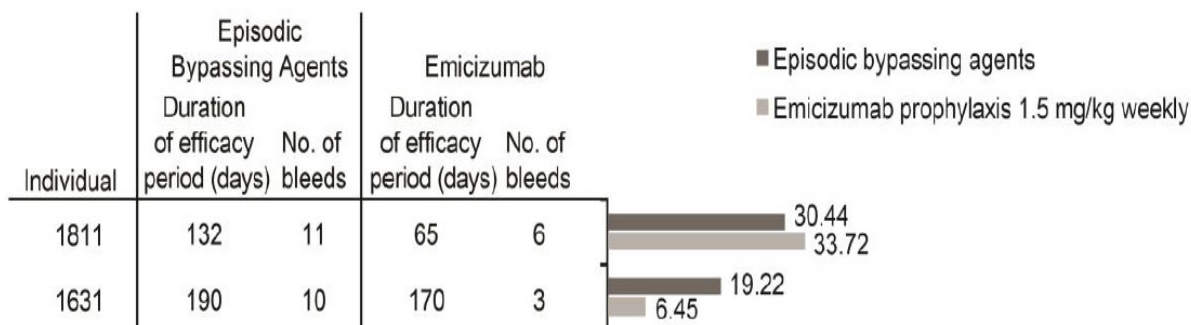


**Hogy lesz ebből annualizált ráta?**

# Ráták számítása eltérő követési idő esetén: annualizált ráta

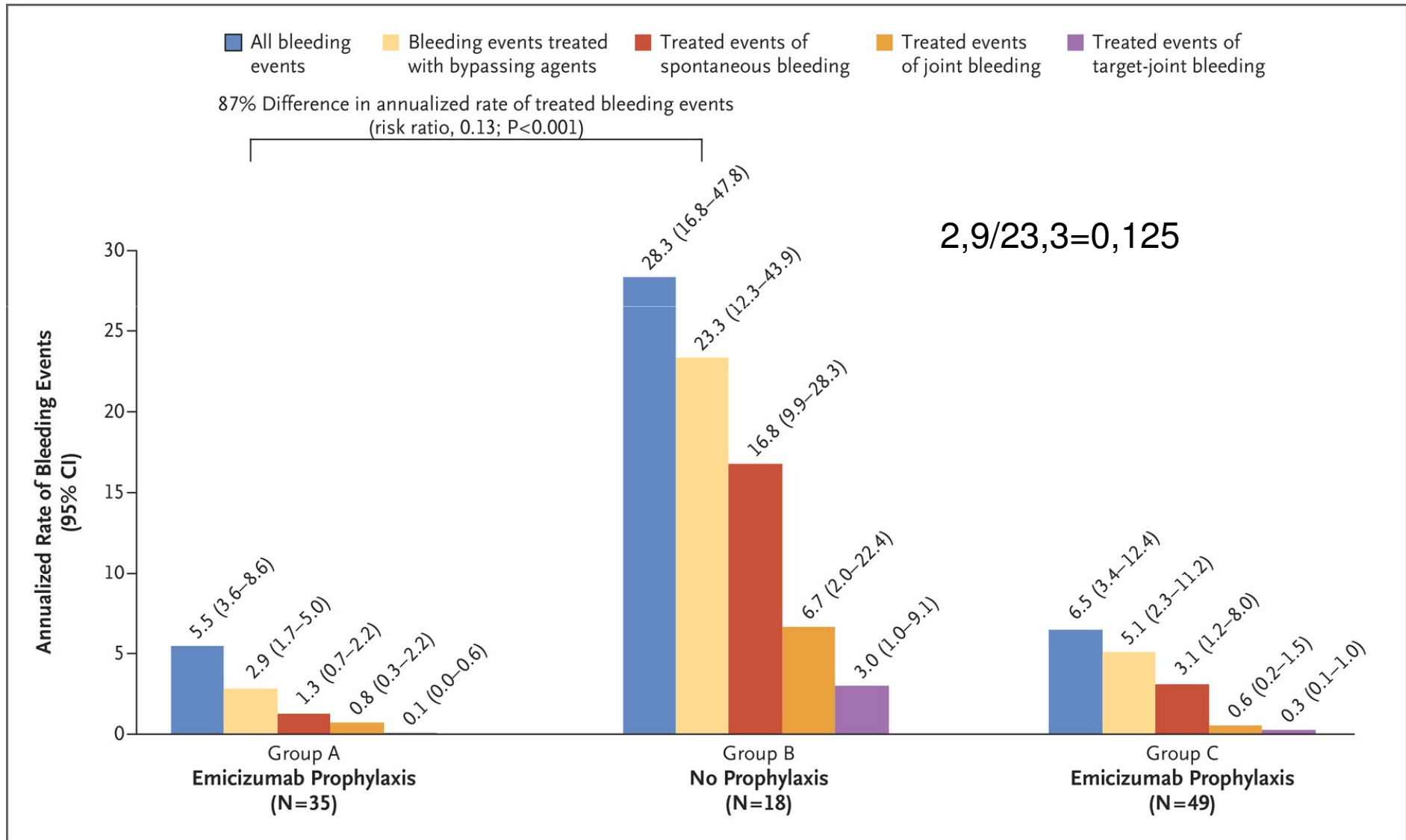
- Esemény:
  - Egyszeri, jól definiálható (halálozás)
  - Többszörös, jól definiálható (relapszus)
  - Többszörös, minőségi (vérzés: trigger, mennyiség, lokalizáció, kimenetel...)
- Időszak:
  - Naptárhoz rögzített, fix
  - Random, fix időszakra kivetített (pl. annualizált)

**Figure S4.** Intra-Individual Comparison for Treated Bleeds in Participants Receiving Emicizumab Prophylaxis (Arm A) Versus No Previous Prophylaxis (Only Episodic Bypassing Agent Treatment).



ARR: NBRM  
 Becslése:  
 $(11 \times 365,25) / 132 = 30,44$

# Hogyan hasonlítsunk össze annualizált esemény rátákat?

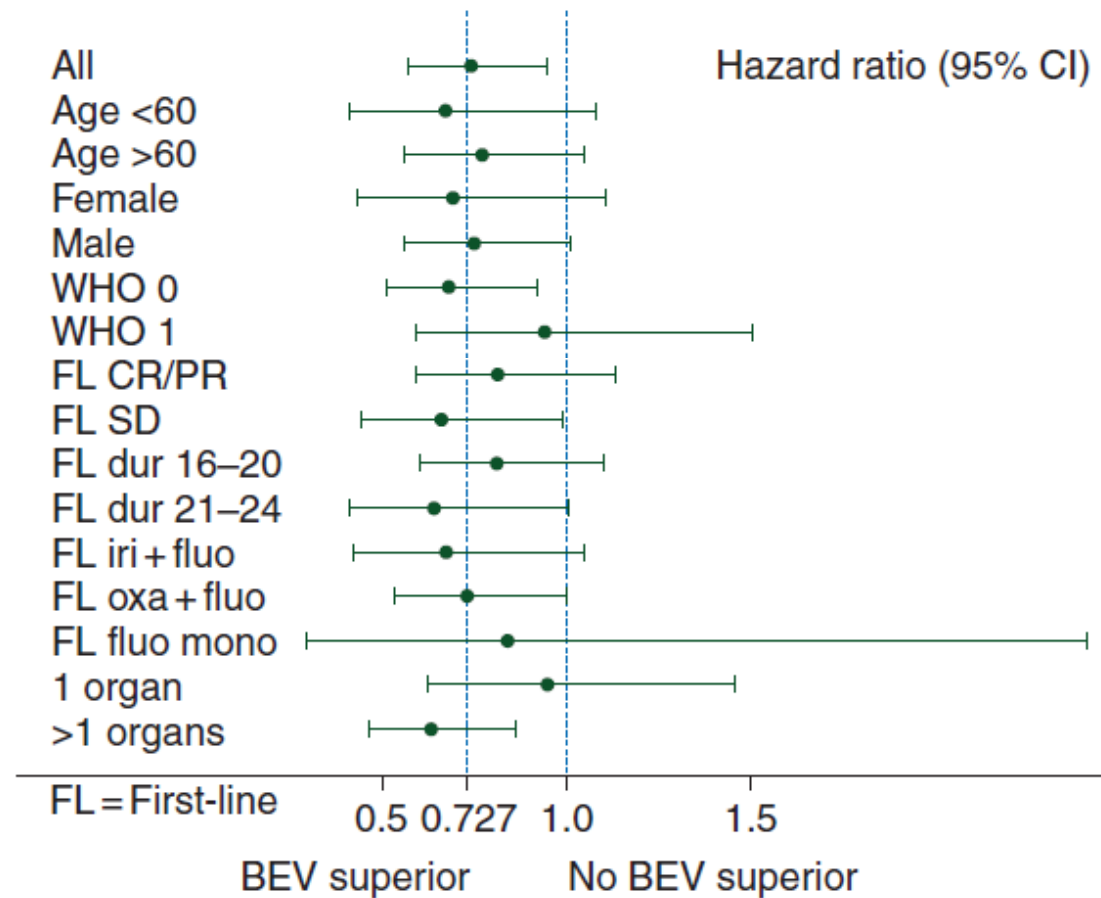




# Valószínűségi tartomány, azaz confidence interval (CI)

- Konfidencia-intervallum egy valószínűségi tartomány, az adott érték valószínűleg a jelzett korlátok közé esik
  - A populációból vett minta átlag értékével kívánjuk megbecsülni a populáció átlagát
  - A becslés pontossága (tartomány szélessége) összefügg a minta nagyságával: nagy minta, pontos becslés, szűk tartomány
- Minden mutatónak (átlag, ráta, arány, stb.) kiszámítható a CI-je
- A CI megadásakor jelezni kell a valószínűség szintjét (pl. 95%)
  - Szérum Na átlag: 141 mmol/L, az átlag 95%-os CI-je: 139,5-142,5
  - 2,5-2,5% lenne az aránya a 139,5 alatti és 142, feletti **átlagértékek** előfordulásának (ha végtelenszer megismételnénk a mintavételezést és mérést)
- A konfidencia-intervallum értelmezése közel áll a szignifikanciateszthez
  - Tartomány nem foglalja magába az 1-et (ráták, hányadosok esetén)
  - Nem fed át két tartomány (folytonos változók vagy arány esetén)
- <https://statisticasoftware.wordpress.com/2012/10/16/how-to-find-confidence-intervals-for-a-single-proportion/>

A hazard ratio (vagy más ráták, hányadosok) ábrázolásának, összehasonlításának módja:  
**Forest plot**

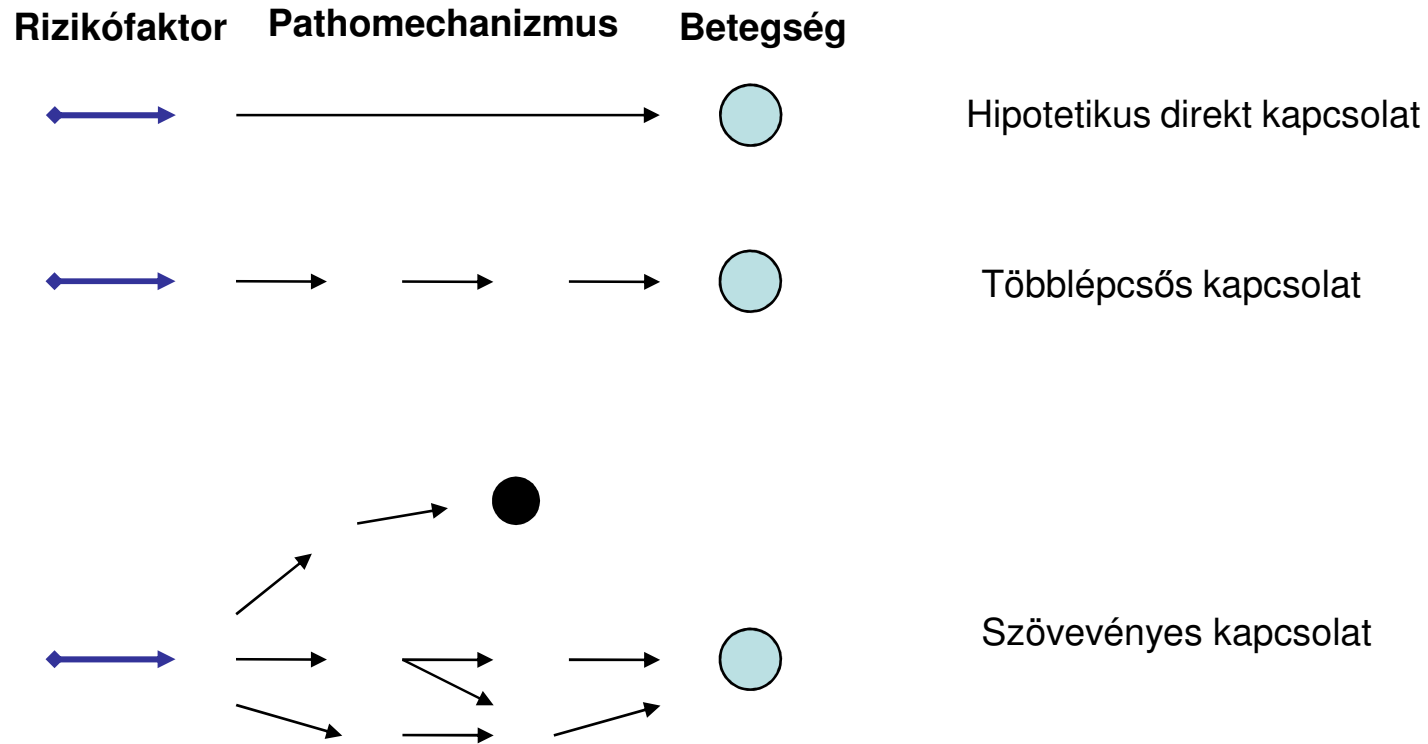


**Figure 2.** Hazard ratios and 95% confidence intervals for time to progression by patient subgroup. BEV, bevacizumab.

# A többváltozós modellek

- Az előző példák egyben jelzik azt is, hogy miért van szükség többváltozós modellekre
- A betegségek többsége multifaktoriális, még az ún. monogénes betegségek is
- A vizsgálat tárgyát képező változó mindig csak részben magyarázza a független változót ( van „reziduum”)
- Az összes magyarázó változó megtalálására csekély az esély
- A vizsgálat tervezési szakaszában dől el, hogy mekkora lesz a reziduum. Nagy lesz, ha
  - Az ismert és lényeges (irodalom!) vagy vélelmezhető (tehetségünk és gondosságunk) változók mérését/rögzítését elmulasztjuk
  - Tervezési hibát vétünk
    - Vizsgálattípus, mintavétel
    - Torzítás
    - Elemszám

# A multifaktoriális betegségek kialakulásának általános jellemzői



**Koleszterin**

**AMI**



Direkt kapcsolat?

# A mai fogalmak szerinti első randomizált klinikai kísérlet: 1948

## BRITISH MEDICAL JOURNAL

LONDON SATURDAY OCTOBER 30 1948

### STREPTOMYCIN TREATMENT OF PULMONARY TUBERCULOSIS

A MEDICAL RESEARCH COUNCIL INVESTIGATION

The following gives the short-term results of a controlled investigation into the effects of streptomycin on one type of pulmonary tuberculosis. The inquiry was planned and directed by the Streptomycin in Tuberculosis Trials Committee, composed of the following members: Dr. Geoffrey Marshall (chairman), Professor J. W. S. Blacklock, Professor C. Cameron, Professor N. B. Capon, Dr. R. Cruickshank, Professor J. H. Gaddum, Dr. F. R. G. Heaf, Professor A. Bradford Hill, Dr. L. E. Houghton, Dr. J. Clifford Hoyle, Professor H. Raistrick, Dr. J. G. Scadding, Professor W. H. Tytler, Professor G. S. Wilson, and Dr. P. D'Arcy Hart (secretary). The centres at which the work was carried out and the specialists in charge of patients and pathological work were as follows:

*Brompton Hospital, London.*—Clinician: Dr. J. W. Crofton, Streptomycin Registrar (working under the direction of the honorary staff of Brompton Hospital); Pathologists: Dr. J. W. Clegg, Dr. D. A. Mitchison.  
*Colindale Hospital (L.C.C.), London.*—Clinicians: Dr. J. V. Hurford, Dr. B. J. Douglas Smith, Dr. W. E. Snell; Pathologists (Central Public Health Laboratory): Dr. G. B. Forbes, Dr. H. D. Holt.  
*Harefield Hospital (M.C.C.), Harefield, Middlesex.*—Clinicians: Dr. R. H. Brent, Dr. L. E. Houghton; Pathologist: Dr. E. Nassau.

*Bangour Hospital, Bangour, West Lothian.*—Clinician: Dr. I. D. Ross; Pathologist: Dr. Isabella Purdie.  
*Killingbeck Hospital and Sanatorium, Leeds.*—Clinicians: Dr. W. Santon Gilmour, Dr. A. M. Reeve; Pathologist: Professor J. W. McLeod.  
*Northern Hospital (L.C.C.), Winchmore Hill, London.*—Clinicians: Dr. F. A. Nash, Dr. R. Shoulman; Pathologists: Dr. J. M. Alston, Dr. A. Mohun.  
*Sully Hospital, Sully, Glam.*—Clinicians: Dr. D. M. E. Thomas, Dr. L. R. West; Pathologist: Professor W. H. Tytler.

The clinicians of the centres met periodically as a working subcommittee under the chairmanship of Dr. Geoffrey Marshall; so also did the pathologists under the chairmanship of Dr. R. Cruickshank. Dr. Marc Daniels, of the Council's scientific staff, was responsible for the clinical co-ordination of the trials, and he also prepared the report for the Committee, with assistance from Dr. D. A. Mitchison on the analysis of laboratory results. For the purpose of final analysis the radiological findings were assessed by a panel composed of Dr. L. G. Blair, Dr. Peter Kerley, and Dr. Geoffrey S. Todd.

#### Introduction

When a special committee of the Medical Research Council undertook in September, 1946, to plan clinical trials of streptomycin in tuberculosis the main problem faced was that of investigating the effect of the drug in pulmonary tuberculosis. This antibiotic had been discovered two years previously by Waksman (Schatz, Bugie, and Waksman, 1944); in the intervening period its power of inhibiting tubercle bacilli *in vitro*, and the results of treatment in experimental tuberculous infection in guinea-pigs, had been reported; these results were strikingly better than those with any previous chemotherapeutic agent in tuberculosis. Preliminary results of trials in clinical tuberculosis had been published (Hinshaw and Feldman, 1945; Hinshaw, Feldman, and Pfuetze, 1946; Keefer *et al.*, 1946); the clinical results in pulmonary tuberculosis were encouraging but inconclusive.

The natural course of pulmonary tuberculosis is in fact so variable and unpredictable that evidence of improvement or cure following the use of a new drug in a few cases cannot be accepted as proof of the effect of that drug. The history of chemotherapeutic trials in tuberculosis is filled with errors due to empirical evaluation of drugs (Hart, 1946); the exaggerated claims made for gold treatment, persisting over 15 years, provide a spectacular example. It had become obvious that, in future, conclusions regarding the clinical effect of a new chemotherapeutic agent in tuberculosis could be considered valid only

if based on adequately controlled clinical trials (Hinshaw and Feldman, 1944). The one controlled trial of gold treatment (and the only report of an adequately controlled trial in tuberculosis we have been able to find in the literature) reported negative therapeutic results (Amberson, McMahon, and Pinner, 1931). In 1946 no controlled trial of streptomycin in pulmonary tuberculosis had been undertaken in the U.S.A. The Committee of the Medical Research Council decided then that a part of the small supply of streptomycin allocated to it for research purposes would be best employed in a rigorously planned investigation with concurrent controls.

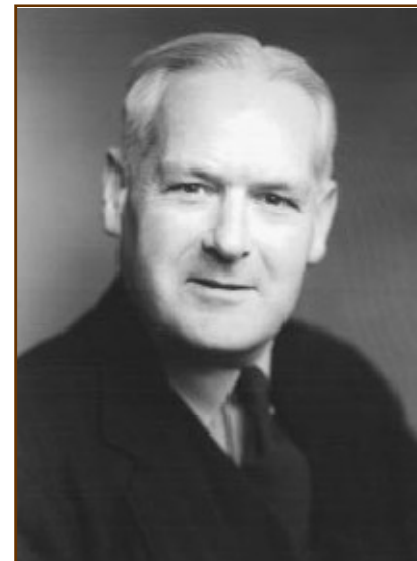
The many difficulties of planning and conducting a trial of this nature are important enough to warrant a full description here of the methods of the investigation.

#### Plan and Conduct of the Trial

##### Type of Case

A first prerequisite was that all patients in the trial should have a similar type of disease. To avoid having to make allowances for the effect of forms of therapy other than bed-rest, the type of disease was to be one not suitable for other forms of therapy. The estimated chances of spontaneous regression must be small. On the other hand, the type of lesion should be such as to offer some prospect of action by an effective chemotherapeutic agent; for this reason old-standing disease, and disease with thick-walled

4582



- **Sir Austin Bradford Hill (1897-1991)**
- **Az első randomizált klinikai kísérlet statisztikusa**
- **Később a dohányzás és a tüdőrák közötti kapcsolat feltárásáról írt munkáival vált ismertté**
- **Hill oksági kritériumok**

# Hill oksági kritériumok

## (Gondoljunk a gyermekági lázra!)

- Időszerűség (Temporal relationship)
  - Az ok megelőzi a következményt
- Erősség (Strength)
  - Ha az A és B tényezők közötti kapcsolat erős, az oki kapcsolatra utal
- Dózis-hatás összefüggés (Dose-response relationship)
  - Az expozíció fokozódásával arányosan növekszik a rizikó (Vica-versa)
  - Kimutatása nehézségekbe ütközhet (küszöbérték)
- Egybehangzóság (Consistency)
  - Ismételt vizsgálatok egybehangzó eredményeket hoznak
- Elfogadhatóság, koherencia (Plausibility, coherence)
  - A jelenség magyarázata összhangban van az érvényes törvényszerűségekkel
- Alternatív magyarázat (Alternate explanations)
  - A kutatók elvárt gondossággal ítélték meg más magyarázat lehetőségét
- Kísérletesen bizonyítható kapcsolat

**Koleszterin**

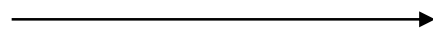


**AMI**



Direkt kapcsolat?

**CRP**



**Atherosclerosis**



Direkt kapcsolat?



European Heart Journal (2010) **31**, 2092–2097  
doi:10.1093/eurheartj/ehq242

**REVIEW**

*Controversies in cardiovascular medicine*

## **C-reactive protein is a bystander of cardiovascular disease**

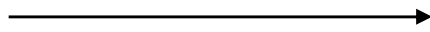
**Sonia S. Anand\* and Salim Yusuf**



- Egyváltozós modellekkel megválaszolható kérdések
  - Egy megfigyelt különbség vagy arány a véletlen hatására alakult-e ki?
- Többváltozós modellekkel megválaszolható kérdések
  - Az adott független változó(k) szignifikáns kapcsolatban vannak-e függő változóval?
  - Mekkora a független változók egymáshoz képest mutatott relatív súlya a megfigyelt különbség (arány) kialakításában?
  - Más szavakkal: segít súlyozni, megítélni, típusba sorolni a független változóinkat.

**Koleszterin**

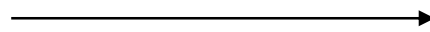
**AMI**



Direkt kapcsolat?

**Öngyújtó a zsebben**

**Tüdőrák**



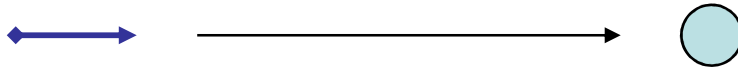
Direkt kapcsolat?

A független változó			
Típusa	Jellegzetességei	Felismerési lehetőség	Példa
Zavaró faktor (confounder)	Kapcsolatban áll a rizikótényezővel és okozati összefüggésben a kimenetellel	Rétegzett analízis Többváltozós analízis A statisztikai analízis azonban önmagában nem elegendő, hogy a kettő között különbséget lehessen tenni.	Öngyújtó, dohányzás, tüdőrák
Közbeeső változó (intervening variable)	Kapcsolatban áll a rizikótényezővel és okozati összefüggésben a kimenetellel, a rizikótényező általa fejt ki hatását		Doh, cotinine, toxikus metabolitok, tüdőrák
Gátló faktor (suppressor)	Kapcsolatban áll a rizikótényezővel és a kimenetellel, a rizikófaktor hatását elfedi	Rétegzett analízis a supresser változó szerint Többváltozós analízis, adjusztálás (illesztés) a supresser változóra	Aspirin, MI, halálozás Születés helye, perinat halálozás; születési súly
Hatásmódosító vagy interakciós változó (interaction variable)	Hatásmódosítás. A rizikótényező hatása a kimenetelre egy harmadik változó értékétől függ.	Rétegzett analízis az interakciós változó szerint. Többváltozós analízis „with interaction terms”	Expozíció, inokuláció mértéke, betegség súlyossága

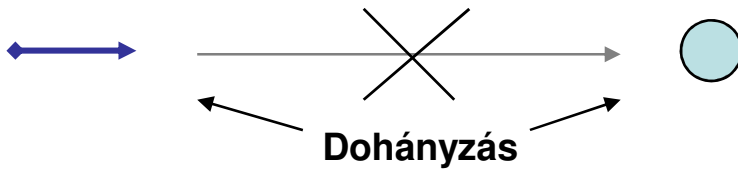
# Confounder

Öngyújtó a zsebben

Tüdőrák

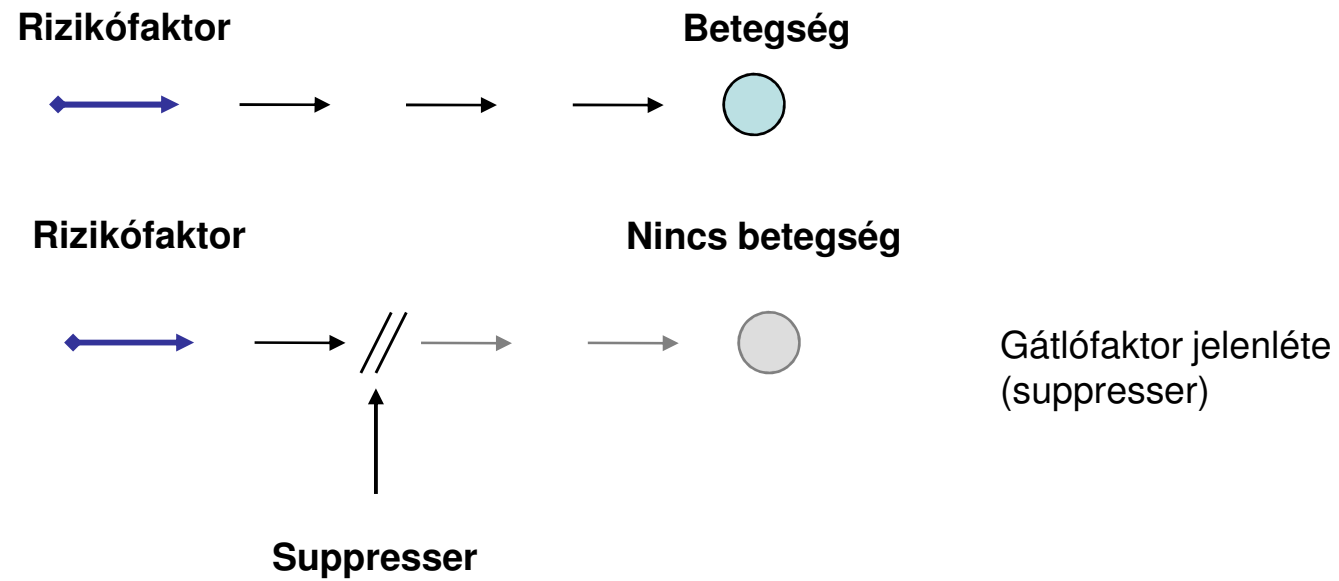


Direkt kapcsolat?



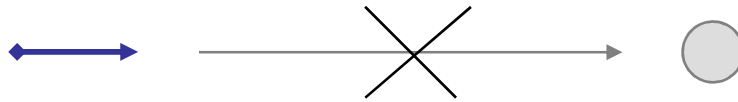
Dohányzás, mint zavaró tényező

# Confounder, mint suppresser (betegség van-nincs)



**Aspirin**

**Halálozás**



Egyváltozós analízis eredménye

**Aspirin**

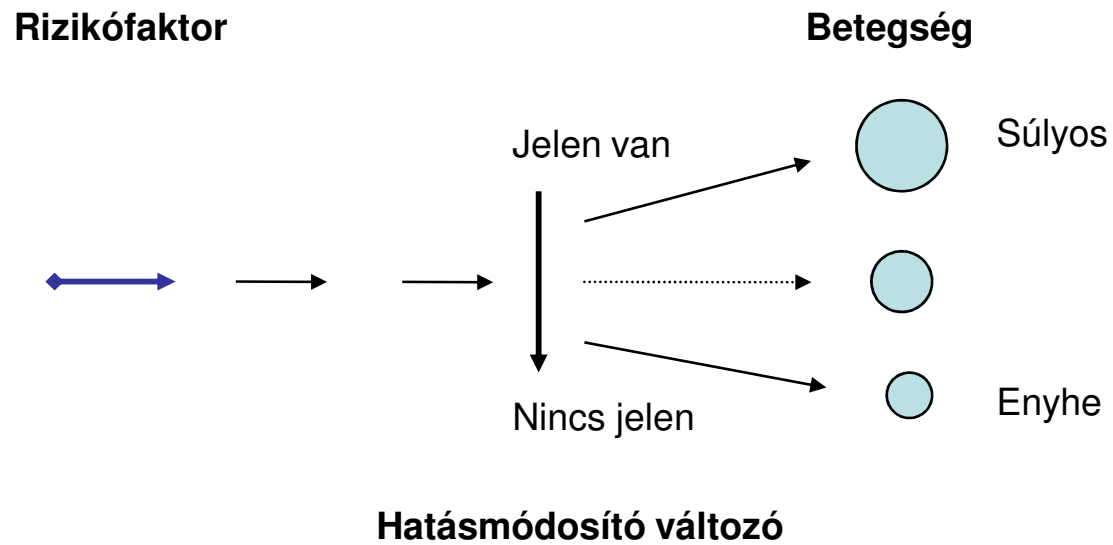
**Halálozás**



Többváltozós, a gátlófaktor jelenlétére illesztett analízis eredménye

**AMI az anamnézisben**

# Hatásmódosító tényező (betegség típusának, súlyosságának változása)



Rizikófaktor



Betegség

## Zavaró tényezők

*Pozitív:* a kapcsolat túlbecslését okozzák

*Negatív:* a kapcsolat alulbecslését okozzák

Ugyan hasonlóan viselkednek, de nem tartoznak a zavaró tényezők közé:

*Hatásmódosító tényezők (interakció)*

*Közbeeső változók (oki tényezők)*

## Zavaró tényezők jellegzetességei

*Kapcsolatban állnak a rizikófaktorral*

*Kapcsolatban állnak a betegséggel a rizikófaktor hiányában is*

*Nem közbeeső változók*

*Eltérő gyakorisággal fordulnak elő a vizsgálati csoportokban*

## A zavaró tényezők ellenőrzés alatt tartásának lehetőségei

*Tervezési fázisban:* a lehetséges zavaró tényezők felmérése, megismerése, vizsgálatba építése

*Vizsgálati szakaszban:* véletlenszerű, korlátozott vagy kapcsolt betegbeválogatás

*Analízis során:* rétegzett vagy többváltozós analízis



ma



Logisztikus regresszió



# Hipotetikus példa a zavaró tényező hatásának bemutatására

Születési sorrend

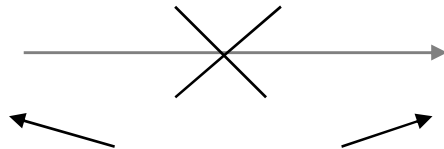
Down-szindróma



Egyváltozós analízis eredménye

Születési sorrend

Down-szindróma



A zavaró tényezőre illesztett rétegzett analízis eredménye

Anyai életkor

# Hipotetikus példa a zavaró tényező hatásának bemutatására

Basic statistics/Tables>>Frequency tables

Frequency table: terhesség száma (Down\_példa)

Szülés sorszáma	Count	Cumulative - Count	Percent	Cumulative - Percent
<b>1</b>	39	39	28,05755	28,0576
<b>2</b>	50	89	35,97122	64,0288
<b>3</b>	50	139	35,97122	100,0000
<b>Missing</b>	0	139	0,00000	100,0000

Frequency table: Down (igen/nem) (Down\_példa)

Down igen/nem	Count	Cumulative - Count	Percent	Cumulative - Percent
<b>0 nem</b>	121	121	87,05036	87,0504
<b>1 igen</b>	18	139	12,94964	100,0000
<b>Missing</b>	0	139	0,00000	100,0000

# Egy változós analízis eredménye

Statistics: terhesség száma(3) x Down (igen/nem)(2) (Down\_példa)

2-Way Summary Table: Observed Frequencies (Down\_példa) Marked cells have counts > 10

Szülés sorszáma	Down nem	Down igen	Row - Totals	Incidencia/ 100 szülés
1	37	2	39	2/39 =0,05
2	44	6	50	6/50 =0,12
3	38	12	50	12/50 =0,24
<b>Totals</b>	119	20	139	20/139=0,143

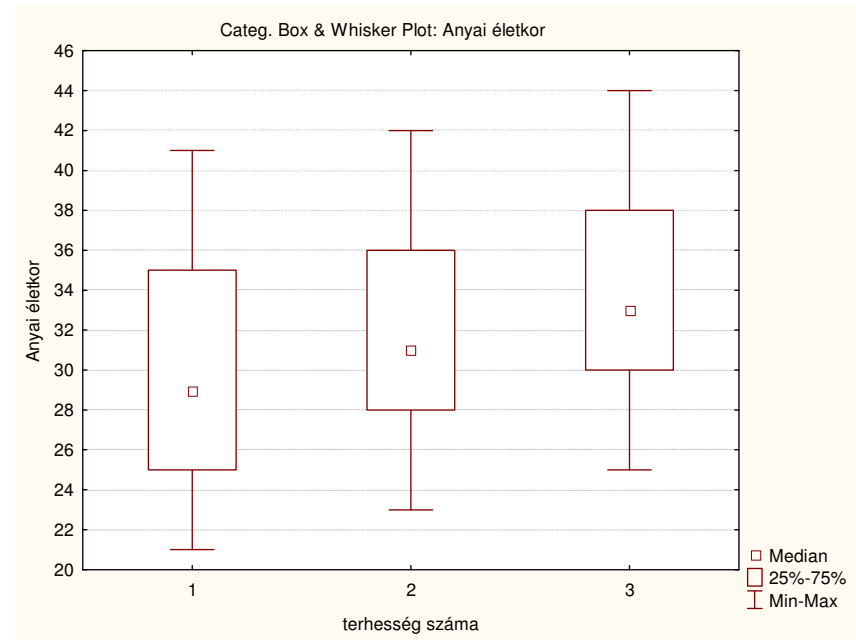
	Chi-square	df	p
<b>Pearson Chi-square</b>	6,696320	df=2	p=,03515

Vajon a ráták közötti különbség a véletlen hatására alakult-e ki?  
Hogy vizsgálhatjuk az anyai életkor szerepét (confounding)?

# Vajon minden zavaró (confounding) tényezőre figyelemmel voltam-e?

Graphs>>2D graphs>> boxplot (advanced)

- A születési sorrend kapcsolatát mutat az anyai életkorral
- Egy-szemponos ANOVA  
 $p=0,008$



Rétegzett analízis eredménye:  
összehasonlítás születési sorrend szerint, figyelembe véve a korcsoportot

	szülés sorszáma	Kor_kategória (<25, 25-34, >=35)	Down nem	Down igen
	1	<25	8	0
	1	25-34	20	1
	1	>34	9	1
	2	<25	2	0
	2	25-34	31	0
	2	>34	11	6
	3	<25	0	0
	3	25-34	25	4
	3	>34	13	8
	All Groups		119	20

Down nem fordult elő

Rétegzett analízis eredménye:  
összehasonlítás születési sorrend szerint, figyelembe véve a korcsoportot

	szülés sorszáma	Kor_kategória (<25, 25-34, >=35)	Down nem	Down igen
	1	<25	8	0
	1	25-34	20	1
	1	>34	9	1
	2	<25	2	0
	2	25-34	31	0
	2	>34	11	6
	3	<25	0	0
	3	25-34	25	4
	3	>34	13	8
	All Groups		119	20

P value 0,1377

Rétegzett analízis eredménye:  
összehasonlítás születési sorrend szerint, figyelembe véve a korcsoportot

	szülés sorszáma	Kor_kategória (<25, 25-34, >=35)	Down nem	Down igen
	1	<25	8	0
	1	25-34	20	1
	1	>34	9	1
	2	<25	2	0
	2	25-34	31	0
	2	>34	11	6
	3	<25	0	0
	3	25-34	25	4
	3	>34	13	8
	All Groups		119	20

P value 0,1498

Rétegzett analízis eredménye: összehasonlítás korcsoport szerint,  
figyelembe véve a születési sorrendet

	szülés sorszama	Kor_kategória (<25, 25-34, >=35)	Down nem	Down igen
	1	<25	8	0
	1	25-34	20	1
	1	>34	9	1
	2	<25	2	0
	2	25-34	31	0
	2	>34	11	6
	3	<25	0	0
	3	25-34	25	4
	3	>34	13	8
	All Groups		119	20

Chi-square test for trend  
P value 0,3362

P value 0,0007

P value 0,0470



# Megoldás logisztikus regresszióval:

- 2023-05-04

Valós, biológiai kapcsolat

vagy

zavar a vizsgálatban?

### •**Hatásmódosító tényező (interaction)**

- Biológiai kapcsolaton alapul*
- A kapcsolat eltérő különféle alcsoportokban*
- Felismerésével a folyamatok biológiai alapja jobban megismerhető*
- Segít megmagyarázni megfigyelésünket*
- Sok esetben egyszerű és gyakorlatban is hasznos jelenséget tár fel*

### •**Zavaró tényező (confounder)**

- A vizsgálathoz tartozik*
- Az egy- és többváltozós esélyhányadosok különböznek*
- A valós hatást elferdíti*
- Az adatokat összezavarja*
- Megelőzhető (tervezés)*
- Ellenőrzés alatt tartható (adatanalízis)*

- A torzítás a változók mérése vagy rögzítése során fellépő szisztematikus hiba
  - Beválogatás, mintavételezés, mérési metodika, adathiány, stb...(ld. első kurzus [előadása](#))

Köszönöm a kitartó figyelmet!

Logisztikus regresszió:  
Független változó terhesség sorszáma,  
függő változó: Down igen/nem

Statistics>>Advanced linear/nonlinear modles>> Nonlinear estimation>>Quick logit regression

	Const.B0	Independent var: szülés sorszáma (1-2-3)
<b>Estimate</b>	-3,755809	0,869951
<b>Standard Error</b>	0,9023546	0,3552801
<b>t(137)</b>	-4,162231	2,448634
<b>p-level</b>	0,00005538191	0,01560353
<b>-95%CL</b>	-5,540153	0,167409
<b>+95%CL</b>	-1,971464	1,572493
<b>Wald's Chi-square</b>	17,32417	5,995809
<b>p-level</b>	0,00003159763	0,01434523
<b>Odds ratio (unit ch)</b>	0,02338154	<b>2,386794</b>
<b>-95%CL</b>	0,003925927	<b>1,182238</b>
<b>+95%CL</b>	0,1392528	<b>4,818645</b>
<b>Odds ratio (range)</b>		5,696784
<b>-95%CL</b>		1,397686
<b>+95%CL</b>		23,21934

# Egyváltozós modell eredménye

## Függő változó: Down (i/n)

	Const.B0	Anyai életkor (év)
<b>Estimate</b>	-8,945886	0,2094651
<b>Standard Error</b>	1,981592	0,05442469
<b>t(137)</b>	-4,514493	3,848716
<b>p-level</b>	0,00001354875	0,0001814117
<b>-95%CL</b>	-12,86435	0,1018441
<b>+95%CL</b>	-5,027423	0,3170862
<b>Wald's Chi-square</b>	20,38065	14,81261
<b>p-level</b>	0,000006369446	0,0001189694
<b>Odds ratio (unit ch)</b>	0,0001302721	<b>1,233018</b>
<b>-95%CL</b>	0,000002588717	<b>1,107211</b>
<b>+95%CL</b>	0,006555684	<b>1,373121</b>
<b>Odds ratio (range)</b>		123,6801
<b>-95%CL</b>		10,40632
<b>+95%CL</b>		1469,949

# Egyváltozós modell eredménye

## Függő változó: Down (i/n)

	Const.B0	Kor_kategória (<25=0, 25-34=1, >=35=2)
<b>Estimate</b>	-6,716664	1,978871
<b>Standard Error</b>	1,479602	0,5424259
<b>t(137)</b>	-4,539507	3,648186
<b>p-level</b>	0,00001222425	0,0003743813
<b>-95%CL</b>	-9,642475	0,9062608
<b>+95%CL</b>	-3,790853	3,051481
<b>Wald's Chi-square</b>	20,60713	13,30926
<b>p-level</b>	0,000005659103	0,0002645174
<b>Odds ratio (unit ch)</b>	0,00121057	7,234569
<b>-95%CL</b>	0,00006491218	2,475051
<b>+95%CL</b>	0,02257634	21,14663
<b>Odds ratio (range)</b>		52,33899
<b>-95%CL</b>		6,125876
<b>+95%CL</b>		447,1801

Többváltozós modell eredménye: a zavaró tényezőre adjusztált modell  
 Függő változó: Down (i/n)

	Const.B0	Anyai életkor	születési szám
<b>Estimate</b>	-9,947693	0,1961575	0,6421229
<b>Standard Error</b>	2,138264	0,05530873	0,3761015
<b>t(136)</b>	-4,652228	3,546592	1,707313
<b>p-level</b>	0,000007697286	0,000535909	0,09004588
<b>-95%CL</b>	-14,17624	0,08678115	-0,1016407
<b>+95%CL</b>	-5,719146	0,3055339	1,385887
<b>Wald's Chi-square</b>	21,64323	12,57832	2,914917
<b>p-level</b>	0,000003296801	0,0003908051	0,0877736
<b>Odds ratio (unit ch)</b>	0,00004783786	1,216719	<b>1,900511</b>
<b>-95%CL</b>	0,0000006971672	1,090658	<b>0,9033541</b>
<b>+95%CL</b>	0,003282514	1,35735	<b>3,998369</b>
<b>Odds ratio (range)</b>		91,0695	3,611943
<b>-95%CL</b>		7,359312	0,8160486
<b>+95%CL</b>		1126,961	15,98696

- Vajon milyen viszonyban van a két független változó ebben a kérdésben?
- Az adjusztált modellben csak az anyai életkor mutat kapcsolatot a függő változóval, a születési sorszám nem
- A statisztikai analízis önmagában nem elegendő annak megállapítására, hogy a feltételezett biológiai kapcsolatot melyik változó okozza (hatás vagy zavar)