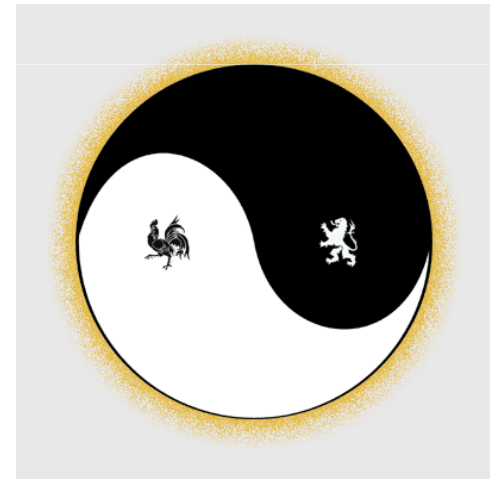


Populáció, minta és adatok: mintavételezés a klinikumban



Dr. Prohászka Zoltán
Az MTA doktora
Semmelweis Egyetem
III. Sz. Belgyógyászati Klinika
2017-09-28
prohoz@kut.sote.hu



Vázlat

- Klinikai vizsgálat típusok és az azokkal megválaszolható kérdések
- A változó fogalma, az adatok fajtái
- A minta és a populáció viszonya
- A mintavételezés gyakorlati lehetőségei
- A mintavételezés nehézségei az orvosi gyakorlatban, torzítás, szisztematikus hiba

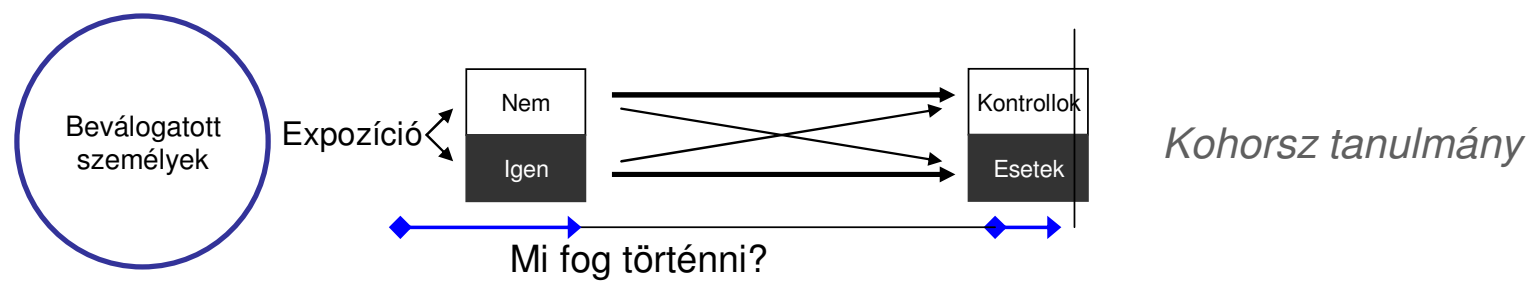
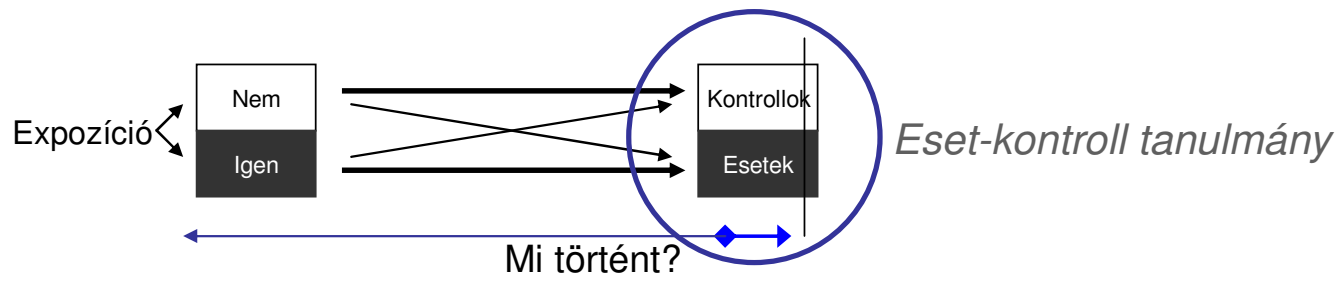
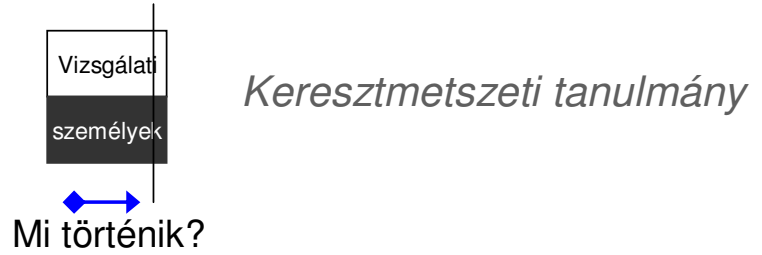
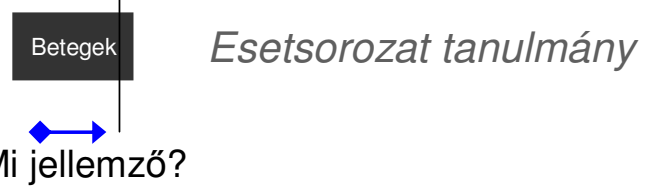
Klinikai vizsgálat típusok és az azokkal megválaszolható kérdések

- Megfigyeléses vizsgálatok, korrelációs vizsgálatok, „clinical study”
- Beavatkozással járó vizsgálatok, „clinical trials”
- Meta-analízis

A reprodukálhatóság krízise a tudományos szakirodalomban

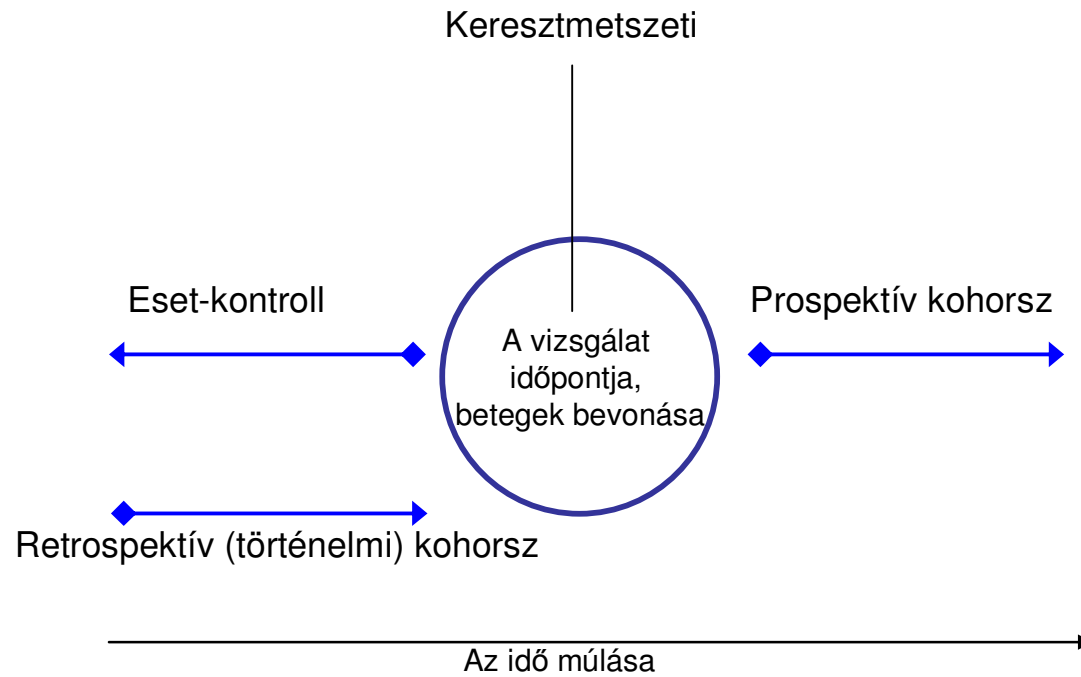
- **A jelenség**
a tudományos szakcikkeknek töredékét sikerült utánvizsgálatokban reprodukálni.
- **Az okok**
 - Tervezési hibák, alacsony elemszám és az ismeretlen „statisztikai erő”
 - Értelmezési hibák: a null hipotézis, p érték és a konfidencia intervallum fogalmában
 - Eredmények szelekciója, a negatív kutatási eredmények nem nyilvánosak
 - Rejtett (vagy eltitkolt) többszörös összehasonlítások
 - Adatok manipulálása „szignifikancia” elérésére
- **A javítás módszerei**
 - A biostatisztikai oktatása színvonalának javítása
 - Hatásméret, elemszám, reprodukálhatóság szintjének tervezése
 - Adatmanipulálási szokások tárgyalása
 - Megbízható negatív eredmények számára közlési lehetőség
 - Biostatisztikai módszertani leírás, biostatisztikai lektorálás
 - Tervezés lektorálása, nyilvánosan regisztrált kutatások elterjesztése
 - A reprodukálhatóság vizsgálatának tudományos értéként elismerése
 - Open Data közlés

Beosztás	Vizsgálat típusa	Megjegyzés
Megfigyeléses vizsgálatok		
	1. Esetsorozat vizsgálatok	Leíró, hipotézis generáló
	2. Keresztmetszeti vizsgálatok	Leíró, hipotézis generáló (Diagnosztikai döntéshozatal támogatása, súlyosság besorolása, mechanizmusok felvetése, ajánlások megfogalmazása)
	3. Eset-kontroll vizsgálatok	Leíró, hipotézist vizsgáló (Betegség okainak feltárása, rizikótényezők azonosítása)
	4. Kohorsz vizsgálatok	Követéses, hipotézist vizsgáló (Betegség okainak, rizikófaktorainak feltárása, természetes lefolyás leírása, prognosztikai tényezők azonosítása)



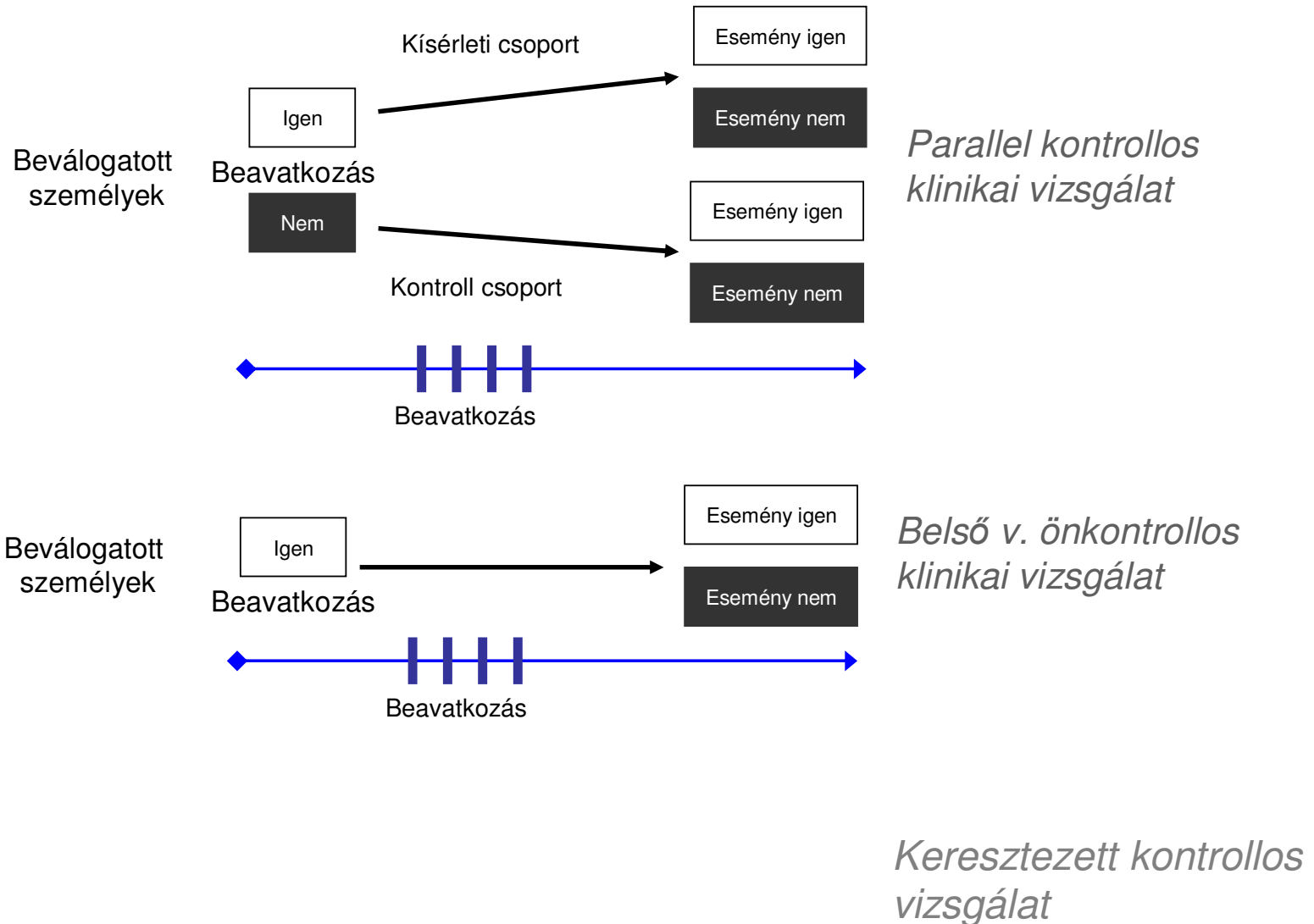
A kék nyilak a vizsgálati kérdés irányát jelölik.

A vizsgálat (betegbeválogatás) ideje, az idő és a megfigyeléses klinikai vizsgálatok kapcsolata. A kék nyilak a vizsgálati kérdés irányára utalnak.



Beosztás	Vizsgálat típusa	Megjegyzés
Klinikai kísérletek (beavatkozással járó vizsgálatok)		
	1. Parallel kontrollos kísérletek	Randomizált Nem randomizált
	2. Önkontrollos kísérletek	Önkontrollos Keresztezett kontrollos
	3. Külső kontrollos kísérletek	Történelmi kontroll
Meta-analízis		Több vizsgálat összegző elemzése

A parallel kontrollos és a belső (vagy ön-) kontrollos klinikai kísérletek felépítésének sémás bemutatása.



Meta-analízis

- A „meta” ebben az értelemben a későbbi, magasabb szintű ismeretekre utal.
- A meta-analízis következtetéseit már publikált tanulmányok feldolgozása, eredményeik egyesítése alapozza meg.
- Megfigyeléses és experimentális vizsgálatok is képezhetik meta-analízis alapját, azonban ezeket csak külön szabad értékelni.
- Különösen hasznos ez az adatkiértékelési forma olyan esetekben, ahol objektív akadályai vannak a nagyobb esetszámú tanulmányok szervezésének (pl. ritka betegségek), vagy kicsiny rizikó különbségeket is biztonsággal kell mérni (esetszám növelés).

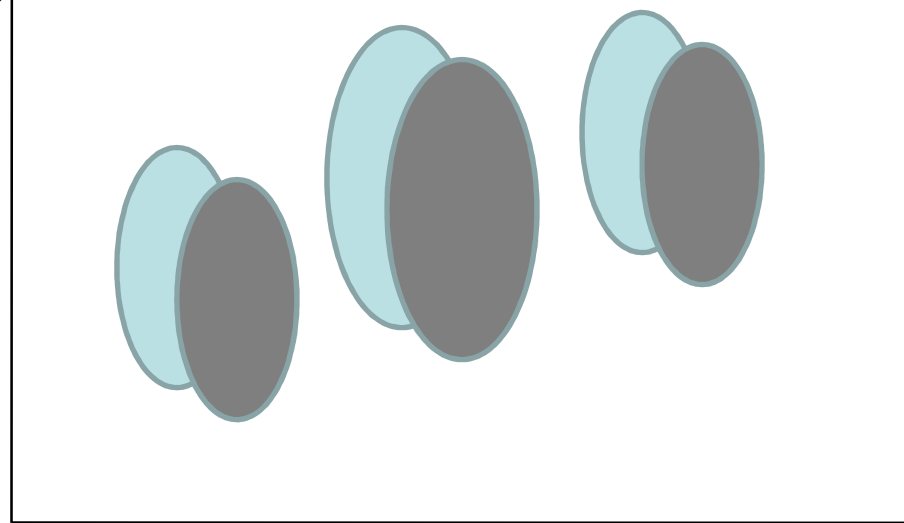
A változó fogalma, az adatok fajtái

- Az orvosi kutatások során egy **populáció** elemeinek, egyedeinek, tagjainak megfigyelésére és vizsgálatára alapozva lehet következtetéseket levonni.
- A kutatás a vizsgálat tárgyát képező **elemek** (metszetek, filmek, dokumentumok), **egyedek** (törzsek, sejtek, állatok) vagy **tagok** (egészséges vagy beteg személyek) jellegzetességeinek megfigyelését, rögzítését, mérését és feldolgozását jelenti.
- Az alkalmazott statisztikában **attribútumnak (attribute)** nevezzük a vizsgálat tárgyát elválaszthatatlanul jellemző tulajdonságokat.
- Az egyes elemeket jellemző tulajdonságok **változhatnak** (testmagasság) vagy **állandóak lehetnek** (pl. nem) a vizsgálati periódusban, valamint több-kevesebb változékonyságot mutathatnak az egyes elemek között (pl. a 14 éves fiúk testmagassága jól becsülhető tartományban szóródik).
- A gyakorlati munka során az attribútumokat sokszor **változó (variable)** névvel illetik, ami sokkal pontosabban utal ezen jellegzetességek viselkedésére vizsgálatainkban.
 - Biológiai változékonyság (személyen belül, személyek között)
 - Mérési bizonytalanság

A függő és a független változó

- A kísérletes vizsgálatokkal kapcsolatban igen lényeges megismerkedni a **függő (dependent)** és a **független (independent) változók** fogalmával.
- Egy vizsgálat során a **manipulált változót** nevezzük **független** változónak, míg azokat a változókat nevezzük függőnek, melyeket csak megfigyelünk, mérünk.
- Ha például a nők és a férfiak koleszterin szintjét kívánjuk összehasonlítani, akkor a független változó a nem (manipulált, mert csoportkialakításra használjuk), míg a függő a koleszterin szint (csak mérjük).

Dependent
(függő) változó (k)



Independent (független) változó

A folytonos és a diszkrét változók

- Az adott tulajdonságot kifejező változó lehet diszkrét (discrete) vagy folytonos (continuous).
- A diszkrét változó csak valamilyen ismert halmazból vehet fel értékeket, ezzel szemben a folytonos változók valamely mért skála értékeit vehetik fel.
- Diszkrét változók a nem, a város, egy állat végtagjainak száma; míg folytonos változók a hőmérséklet, a magasság.
- A diszkrét változók (melyeket szoktak kategória változónak is nevezni) esetében az egyes kategóriákat jelölheti szöveg vagy kódszám.
- A diszkrét változók különleges esetét jelenti a binomiális (vagy dichotóm) eloszlás, amikor a változó értéke csupán kétféle, egymást kizáró kategória lehet (jelölése általában igaz/hamis, 1 vagy 0).

Folytonos változók, mérési skála

- A folytonos változók valamely mért skála értékeit vehetik fel
- Igen fontos további tulajdonsága egy mért jellegzetességnek a **mérési skála** (measurement scale), ami alapvetően meghatározza a változó információtartalmát. A mérési skála tulajdonságai alapján a változókat 4 csoportba oszthatjuk: nominális, ordinális, intervallum és hányados.
- A **nominális skálán mért változók** csak minőségi információt hordoznak, arra alkalmasak, hogy egymástól jól elkülönülő kategóriákat hozzanak létre. Tipikus nominális változó a rassz, a város, a kezelés helyszíne.
- Az **ordinális skálán rögzített változók** segítségével a vizsgálat elemei sorrendbe állíthatók, meg lehet mondani, hogy kinek van több, avagy kevesebb az adott jellegzetességből. Nem lehet azonban kifejezni, hogy mennyivel több. Tipikus ordinális skálán mérhető jellegzetesség pl. a betegség súlyosság, a társadalmi-gazdasági helyzet.
- Az **intervallum skálán felvett változók** segítségével nemcsak az mondható ki, hogy kinek van több, avagy kevesebb az adott jellegzetességből, hanem ez a különbség pontosan, számszerűen ki is fejezhető. Tipikus intervallum skálán mért változók a hőmérséklet, a magasság, a súly.
- A **hányadosok** alapvetően az intervallum változók tulajdonságaival bírnak, azonban további jellegzetességük, hogy valamely kiindulási ponthoz viszonyítottak, így kifejezetten alkalmasak változások egységesített összehasonlítására.

A változó neve	A változó eloszlása	A változó típusa	A mérési skála	A változó értékei	Megjegyzés
Életkor	Diszkrét	Kategória/ numerikus	Ordinális/ intervallum	20 és 87 közötti egész számok	Év
Nem	Diszkrét	Binomiális	Nominális	1 és 2	1 (férfi) és 2 (nő)
Cukorbetegség típusa	Diszkrét	Kategória	Nominális	1, 2, és 3	1=I-es típus, 2=II-es típus, 3=LADA típus*
Testsúly	Folytonos	Numerikus	Intervallum	46,0 és 136,0 közötti számok	1 tizedes pontossággal mért testsúly (kg)
Retinopátia jelenléte	Diszkrét	Kategória	Ordinális	0, 1, 2	0=nincs, 1=egyoldali, 2=kétoldali
Ischaemiás szívbetegség az anamnézisben	Diszkrét	Binomiális	Nominális	0, 1	0=nincs, 1=van

*LADA: **L**ate-onset **a**utoimmune **d**iabetes of **a**dults

A minta és a populáció viszonya

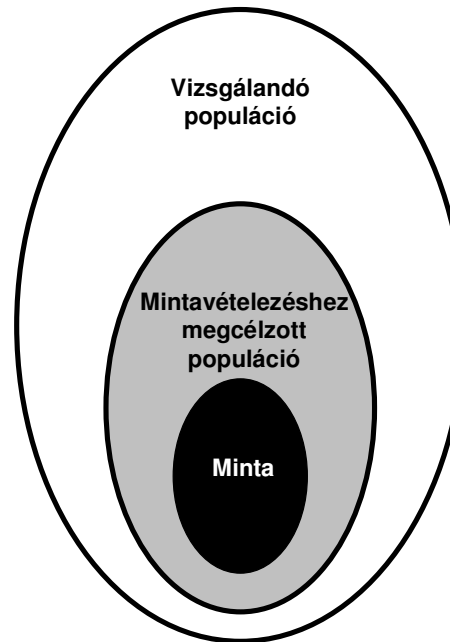
- A statisztikai analízis során egy **minta** adataiból (egyedeiből, elemeiből) kívánunk az egész **populációra** (sokaság, az elemek jól definiált összessége) következtetést levonni.
- A populáció statisztikai értelemben egy olyan sokaság, amelynek van valamely **jól megfogható közös tulajdonsága**.
- Orvosi kutatások esetén legtöbbször betegek, állatok, törzsek vagy más élő szervezetek összességére utal ez a fogalom, de élettelen tárgyak gyűjteménye is jelentheti a sokaság elemeit (pl. szöveti blokkok, radiológiai felvételek, anyakönyvi kivonatok, stb.).
- A populációt **meghatározott módszerekkel kiválasztott** részhalmazát nevezzük **mintának**.

A véletlen szerepe, a valószínűség becslése

- A populációra vonatkozó következtetéseinket a minta vizsgálatával és a **valószínűség becslésével** alapozzuk meg, ezért mintavételezésünk során nagy gondossággal kell eljárunk, hogy helyes következtetést vonhassunk le.
- Statisztikai analízisre az a minta a legalkalmasabb, ami a populációból **véletlenszerűen** kerül kiválasztásra, ugyanis ebben az esetben sérül legkevésbé a valószínűség helyes becslése.

Véletlenszerű mintavételezés (random sampling)

- **A véletlenszerű mintavételezéshez** azonban ismernünk kell a teljes populációt (annak **összes** elemét, egyedét, pl. az orvosi kutatásokban egy betegcsoport összes tagját), és speciális módszereket kell alkalmaznunk, hogy a sokaság minden tagjának **egyenlő esélye** legyen a mintába kerülésre.
- Az általában használatos statisztikai tesztek azon az alapvetésen nyugszanak, hogy a minták **véletlenszerűen** lettek kiválasztva egy **sokkal nagyobb** populációból



Magyarországon élő, diabetológiai ambulanciákon kezelt cukorbetegek csoportja

A véletlenszerűen kiválasztott 10 diabetológiai ambulancián kezelt betegek

A 10 diabetológiai ambulancián kezelt betegek közül a vizsgálatba véletlenszerűen bevont betegek

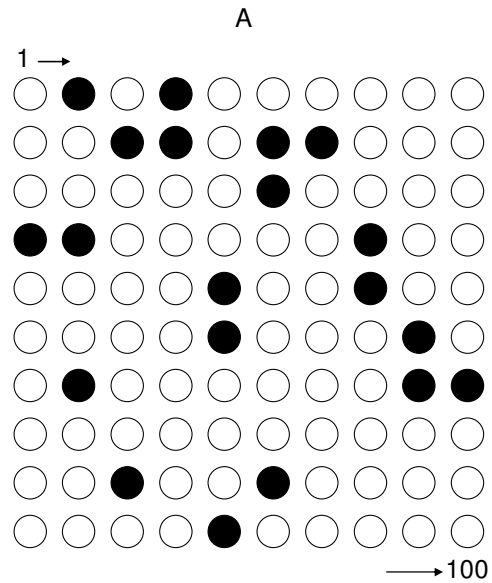
Miért nem dolgozunk a populációval, miért kell mintát vennünk?

- 1, A mintákat **hamarabb** tudjuk vizsgálni, mint a populációkat. A *gyorsaság* sokszor kiemelt szempont egy orvosi vizsgálat kivitelezésénél (pl. új oltóanyag hatásosságának felmérése).
- 2, A minták vizsgálata jelentősen **olcsóbb**, mert kevesebb elemet/egyedet kell megvizsgálni. A költségek különösen nagy vizsgálatoknál jelentenek korlátot.
- 3, Sok esetben a **teljes populáció vizsgálata nem is lehetséges** (pl. a kezelt állatok egy részét meghatározott időnként fel kell áldozni szövettani vizsgálat miatt; vagy pl. egy jövőbeni eseményt akarunk vizsgálni, és ma még nem tudjuk, hogy kinél alakul ki az adott állapot, így mintát sem tudunk az érintettektől vizsgálatra eltenni).
- 4, A kisebb elemszámú minták mérése sokszor **pontosabban** hajtható végre, mint a nagyobb elemszámúaké. A kevés mérést kevesebb, ám speciálisan képzett személyzet végezheti; kevés minta mérésére alkalmazhatunk drágább (sokszor pontosabb) mérést, stb..
- 5, Megfelelő mintavételezés esetén a valószínűség becslésén alapuló módszereket alkalmazhatunk a **statisztikai hiba becslésére**. Mint látni fogjuk, ez kiemelt szempont vizsgálatunk tervezésekor (power-analízis).
- 6, A mintavételezés pontos megtervezésével lehetőség van a **heterogenitás (biológiai variabilitás) csökkentésére**. Tegyük fel, hogy szisztémás lupus erythematosus (SLE)-ben szenvedő betegeket akarunk vizsgálni. A betegség alapvető jellegzetessége a kifejezetten heterogén klinikai megjelenés. Ha csupán a betegség egyes aspektusait akarjuk vizsgálni (pl. veseszövődmény kialakulása), helyes, ha a mintavételezés során eleve csökkentjük a heterogenitást.

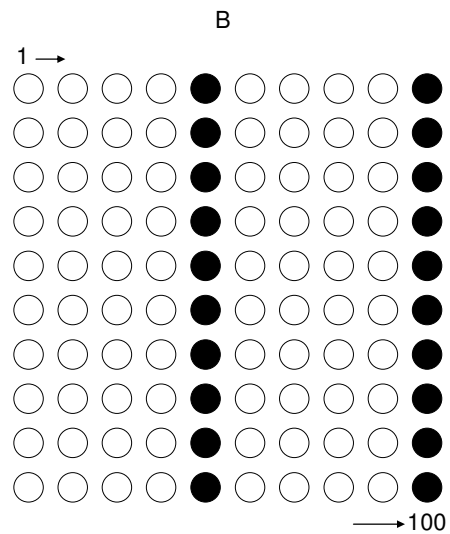
Mintavételezési módszerek

- ***Valószínűségi mintavételezés (probability sampling)***
 - Véletlenszerű mintavételezés (random sampling)
 - Szisztematikus (konszekutív) véletlenszerű mintavételezés
 - Véletlenszerű, rétegzett mintavételezés (Stratified sampling)
 - Véletlenszerű, csoportos (vagy többlépcsős) mintavételezés (Cluster sampling)
 - Véletlen besorolásos vizsgálatok (random assignment)
- ***Nem valószínűségi mintavételezés (non-probability sampling)***
 - Kényelmi mintavételezés (Convenience sampling)
 - Kvóta minta (Quota sample, vagy reprezentatív mintavételezés)

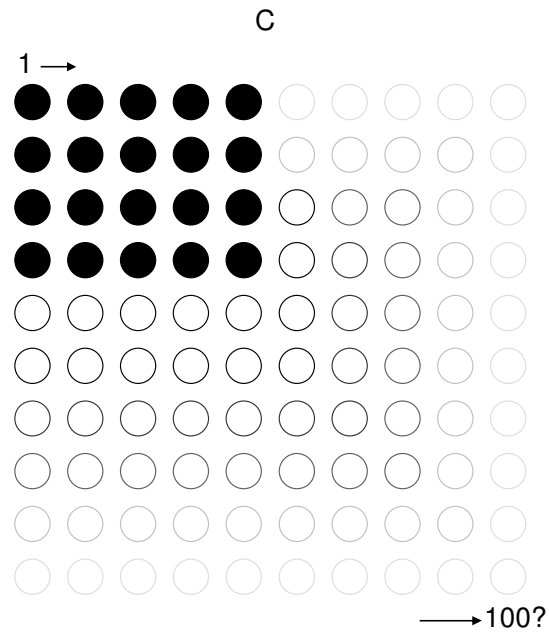
A véletlenszerű minta viszonya a populációhoz



A konszekutív minta viszonya a populációhoz



A kényelmi minta viszonya a populációhoz



Néhány szempont a definíciókhoz

- Véletlenszerű mintavételezés (random sampling)
 - Az eljárás alkalmazásához rendelkezésünkre kell állnia egy listának, amely az ismert elemek (tagok) egyedi azonosítóját tartalmazza.
 - Ezt a listát nevezzük **mintavételi listának (sampling frame)**.

Néhány további szempont a definíciókhoz

Szisztematikus (konszekutív) véletlenszerű mintavételezés

- 1970 és 1990 közötti időszakban szeretnénk felmérni különféle balesetek előfordulásának gyakoriságát (tegyük fel, hogy elérhető a teljes baleseti statisztika): *nézzük meg minden 3. és 9. hónap adatait*

• **Véletlenszerű, rétegzett** mintavételezés (Stratified sampling)

- Először az *alcsoportok* (rétegek, strata) kerülnek kialakításra (azonosításra, definiálásra), majd minden alcsoportból véletlenszerűen kerül kiválasztásra a megkívánt számú minta. *Balesettípusok szerinti alcsoportok vizsgálata.*

• **Véletlenszerű, csoportos** (vagy többlépcsős) mintavételezés (Cluster sampling)

- A populáció tagjait földrajzi elhelyezkedésük alapján *csoportokra* bontják. *Nézzük meg a balesetek előfordulását földrajzi elhelyezkedésük szerint.*

• **Véletlen besorolásos** vizsgálatok (random assignment)

- A vizsgálatokban első lépésben az előzetesen megállapított kritériumok szerint kiválasztják a betegeket a vizsgálathoz, majd véletlenszerű eljárással sorolják a betegeket egyik vagy másik csoportba

Néhány szempont a definíciókhoz

Kényelmi mintavételezés (Convenience sampling)

- Nem ismert a megcélzott populáció, nem becsülhető meg az egyes elemek (személyek) előfordulási gyakorisága és mintába kerülésének esélye, így az sem, hogy a minta valóban reprezentálni fogja-e a megcélzott populációt.

- **Kvóta minta** (Quota sample, vagy **reprezentatív** mintavételezés)

- Reprezentatív mintavételt alkalmaznak, ha a populáció nem teljesen ismert, azonban **előzetes adatokkal** rendelkeznek a mintavételezéshez megcélzott populáció egyes adatairól, tulajdonságairól és a mintavételezést ennek megfelelően kívánják végrehajtani.
- A reprezentatív ebben az értelemben arra utal, hogy *törekednek* a betegek minden csoportjának egyenlő esélyt adni a vizsgálatba való bekerülésre.

A mintavételezés gyakorlati lehetőségei

- **Példa.** Az aktuális ajánlások szerint a kezelt II-es típusú cukorbetegségben szenvedő betegek esetén az éhomi vércukorszint célértéke a 3,89-7,2 mmol/l tartományba esik. Kutatásunkban fel kívánjuk mérni, hogy Magyarországon ezt a célértéket a betegek hány %-a éri el?

A feladat megoldásának módja a *populáció vizsgálatával*

- Az ajánlásoknak megfelelően ***minden*** cukorbeteg (n=750000, hipotetikus adat) egy évben egy alkalommal részt vesz a kötelező diabetológiai felülvizsgálaton egy diabetológiai szakambulancián.
- Ha biztosítani lehetne, hogy ennek a felülvizsgálatnak az eredményei (pl. az éhomi vércukorszint) összesítésre kerüljenek (**kötelező bejelentés**), akkor lehetővé válna a teljes populáció adatainak elemzése.

A példa megoldása *véletlenszerű* mintavétellel

- Példánk esetében a véletlenszerű mintavételezéshez rendelkezünk kellene az összes Magyarországon élő cukorbeteg adatait tartalmazó listával.
- Az összes kezelt beteg azonosítóját tartalmazó lista lenne a sampling frame ($n=750000$)
- Véletlenszám generátorral 1 és 750000 közötti egész számokat rendelünk a listához, és kiválasztjuk a megkívánt számú mintát.

Szisztematikus mintavételezés példánk esetében

- Példánk esetében az összes beteg adatait tartalmazó listából minden 100. beteget választanánk ki
- Veszélyes, ha a listában ciklikusság fordul elő!

A példa megoldása csoportonkénti véletlenszerű mintavételezéssel

- Csökkenthetjük a mintavételezésre fordítandó energiáinkat akkor, ha két lépésben oldjuk meg a feladatot. Először készítsünk egy listát az **összes** diabetológiai gondozóról (tegyük fel $n=100$)
- Válasszunk ki véletlenszerűen közülük tizet, és csak ennek a tíz központnak a betegeiből készítsünk teljes listát.
- A második lépésben ennek a tíz központnak a betegeiből válasszunk ki véletlenszerűen 7500 beteget, és a vizsgálatot a korábban leírtak szerint fejezzük be.

A példa megoldása rétegzett (alcsoportoknak megfelelő) mintavételezéssel

- Tegyük fel, hogy különösen gondosan kívánunk eljárni a betegség súlyosságának vonatkozásában, és biztosítani akarjuk, hogy a legsúlyosabb stádiumú, különféle szövődményekkel sújtott betegek adatai is valós súlyuk arányában kerüljenek be a mintába
- A vizsgálat megkezdése előtt fel kell mérni az egyes szövődmények/stádiumok előfordulási adatait a megcélzott populációban
- Ehhez a tíz kiválasztott gondozó betegregiszterének adatait 2 évre visszamenően áttekintjük
- A 7500 beteg véletlenszerű kiválasztását ezek után az alcsoportoknak megfelelően hajthatjuk végre biztosítva azt, hogy a kiválasztott mintában pontosan olyan arányban forduljanak elő a szövődményes betegek, mint a teljes megcélzott populációban
- Esetsorozat-tanulmányok eredményei nagyon hasznosak lehetnek ilyenkor!

A példa megoldása kényelmi mintavételezéssel

- Gyorsan és minimális ráfordítással kell adatokhoz jutnunk a cukorbetegek vércukor célértékeinek vizsgálatához
- Lehetőségünk összesen annyi, hogy egyetlen centrumban egy hónapig rögzítsük az éhomi vércukoradatokat minden beteg vonatkozásában (***konszekutív mintavételezés***)
- ***Reprezentatív mintavételezés***: ha nem ismert az összes diabetológiai ambulancia listája, akkor is *törekedhetünk* arra, hogy a kiválasztott ambulanciák jól reprezentálják a teljes országot
- További fontos szempont, hogy az egyes ambulanciákon a betegek valós előfordulási arányaiknak megfelelően kerüljenek a vizsgálatba, és ne a legszorgalmasabb ambulancia adja a betegek zömét (***kvóta minta***).

A mintavételezés nehézségei az orvosi gyakorlatban, torzítás, szisztematikus hiba

- Az orvosi kutatások során **torzításról** akkor beszélünk, ha a mintagyűjtés (betegbeválogatás), mintavétel (vizsgálati anyag vétele), az adatrögzítés vagy a mérés fázisában szisztematikus, vagyis állandó, egy irányba ható hiba lép fel.
 - 25 fok felett a műszer 50%-al nagyobb értéket mér

A véletlen hiba

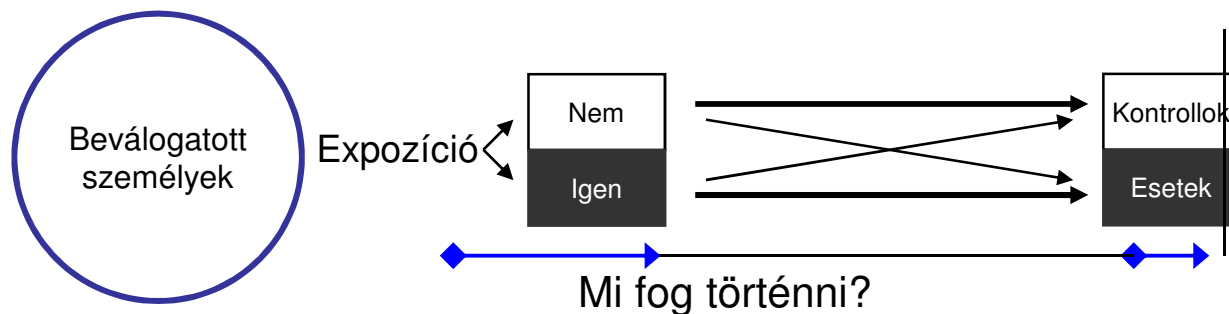
- Mérésünknek, eredményeinknek „csak” a megbízhatósága romlik
 - Például 25 fok feletti laborhőmérséklet esetén a műszer pontatlanná, megbízhatatlanná válik, egyaránt alá és fölé mérhet, az eredmények nagyobb tartományban szóródnak.

Torzítás a betegség prevalenciája/incidenciája miatt

- **Prevalencia** torzítás akkor léphet fel, ha egy kórállapot eredete (kóroki tényezők behatása) és a betegség megjelenése/diagnosztizálása (vizsgálatba való beválasztása) között lényeges idő telik el, és a legsúlyosabb betegek addigra meghalnak.
- Ilyen betegségek esetén torzításmentes vizsgálatot a ***kohorsz*** vizsgálatok adhatnak
- ***Eset-kontroll*** vizsgálatokat helyes úgy szervezni, hogy csak a frissen diagnosztizált betegek kerüljenek be a vizsgálatba.

A prevalencia torzítás bemutatása: kohorsz vizsgálatból származó eredmény

Betegek	Betegek száma a kohorsz vizsgálatban ma, a 10 éves követés végén		
	Életben levők koszorúér betegséggel	Myocardiális halálozás	Életben levők szívbetegség nélkül
Nem cukorbeteg	80	20	900
Cukorbeteg	50	250	700



Helyes következtetésként levonható, hogy a cukorbetegek esélye lényegesen nagyobb az AMI kialakulására mint a nem cukorbetegéké (300 a 100-hoz képest), és lényegesen nagyobb az esélye a halálos AMI-nak cukorbetegekben, mint nem cukorbetegekben (250 a 20-hoz képest).

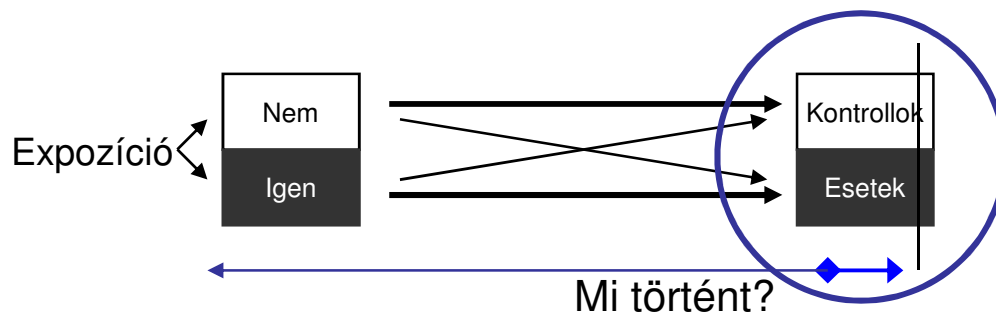
A prevalencia torzítás bemutatása: eset-kontroll vizsgálatból származó eredmény (ha a beválogatást nem korlátozzuk a frissen diagnosztizált cukorbetegekre).

Betegek	A betegek száma eset-kontroll vizsgálatban	
	Életben levők koszorúér betegséggel	Életben levők szívbetegség nélkül
Cukorbeteg	50	700
Nem cukorbeteg	80	900

Nem szerepel:

250

20



Az eredményekből azt a helytelen következtetést vonnánk le, hogy a cukorbetegség védő faktor az AMI kialakulásával szemben

Az orvoshoz fordulás gyakoriságának eltéréseiből származó torzítás

- Példaként olyan vizsgálatok említhetők, amikor kórházban kezelt betegeket vizsgálunk, és egyik csoportunk tagjai (rizikófaktor pozitív betegek) sokkal gyakrabban kapnak kórházi kezelést, mint a másik csoportéi (rizikófaktor negatív betegek, vagy rizikófaktor pozitív egészségesek).
- Az orvoshoz fordulás gyakoriságának eltéréseiből eredő torzítást csoportjaink gondos megválasztásával (orvoshoz utalás rendje ismert protokoll alapján, stb.), esetenként több kontrollcsoport (többféle beteg kontroll és egészséges kontroll) kialakításával küszöbölhetjük ki.

A válaszadás megtagadásából és az önkéntességből eredő torzítás

- A részvétel megtagadásából eredő torzítás alapja az, hogy a kieső személyek egyes jellegzetességei alapvetően eltérhetnek a mintavételezéshez megcélzott sokaság jellegzetességeitől.
- Más szavakkal kifejezve sérül a véletlenszerű mintaválasztás elve, mert a megcélzott sokaság fontos alcsoportja nem vesz részt a vizsgálatban.
- Tágabb értelemben ezt a fajta torzítást kifejezhetjük úgy is, hogy az ***adatok hiányából eredő hiba***.
- Önkéntesség, válaszmegtagadás, mérési hiba

A résztvevők besorolásából eredő torzítás, selection bias

- Nem helyes az a gyakorlat, ha egy kutató az egyik napon kontrollnak (hagyományos kezelés), másik napon a vizsgálati csoportba (alternatív kezelés), így nem biztosítható a véletlenszerű kiválasztás és besorolás (**ciklikusság**).
- Nagy esély van arra, hogy egy súlyosabb állapotúnak tartott beteg részletesebb kivizsgálásban, átgondoltabb terápiában vagyis összességében gondosabb ellátásban részesül, mint egy enyhébb klinikai képet mutató társa (**case-selection bias**).
- Ez a hiba csak előre rögzített beválogatási feltételek és csoportba sorolási rend kialakításával előzhető meg. A vizsgálati protokollnál törekedni kell objektív, mérhető, dokumentálható feltételek megadására, és kerülni kell a beteg szubjektív panaszain vagy a vizsgáló saját benyomásán alapuló tényezőket.
- **Index event bias**
 - Ismétlődő események torzító hatása: pl. FV mutáció DVT rekurrenciát fokozó hatása nem igazolható, ha csak korábban trombózison átesett személyeket vizsgálunk.

A ciklikusságból eredő torzítás

- A *sampling frame* ciklikusságán alapul.
- Évszakok, hónapok, napok ciklikussága, vizsgálatra érkezés sorrendje, mintavétel sorrendje...
- Elkerülhető, ha a ciklikusság ismert, vagy aktívan keressük.

Rejtett összefüggésekből eredő torzítás

- A vizsgálatokat torzító tényezők legnehezebben kontrollálható csoportját a rejtett összefüggések képezik.
- Ha kellő alaposággal megismerkedünk a vizsgált betegséggel, a mérendő molekulákkal, alkalmazni kívánt eljárásokkal, jelentősen csökkenthető a rejtett összefüggések visszamaradása a megtervezett vizsgálatban.

(Bio)statisztikai hibák és orvoslásuk a jobb reprodukálhatóságért

- **Tervezés**
 - **Hatásméret, megfelelő elemszám és a „statisztikai erő” (power)**
 - **A tervezési folyamatban** szerepeljen a reprodukálhatóság becslése
 - **A szórás és a hatás méret** szerepeljen a paraméterek között
 - **A feltáró és a megerősítő vizsgálatok megkülönböztetése**
 - Adattisztítás szabályainak rögzítése
- **Biostatisztikai szemlélet**
 - A hipotézis vizsgálat, a p érték és a konfidencia intervallum gondos értelmezése
 - A fals pozitív esély becslése, értelmezése
 - A szignifikancia szó használatának mellőzése
- **Többszörös összehasonlítás a kutatási folyamatban:**
 - Adathalászat sok mérés között
 - a „szignifikáns” esetek feltüntetése tervezett vizsgálatként
- **Adat manipulálás módszerei „szignifikancia” elérésére**
 - Kedvezőtlen adatok kihagyása a szórás csökkentésére
 - **Menet közbeni esetszám növelés a $p < 0,05$ eléréséig**
- **A negatív kimenetelű kutatási eredmények kezelése**
 - A pozitív eredmények közlésében a negatív vizsgálatok is megjelenítendőek
 - A kutatási hipotézisek negatív eredményű tesztelésének közlése

Vizsgálettervezés

- Irodalomkutatás
- Módszertan kiválasztása
- Power-analízis
- Betegbeválogatás, definíciók, biológiai mintavétel
- Adatfeldolgozás
- Statisztikai elemzés
- Interpretálás

