

# A confounding megoldásai: megfigyelés és kísérlet

Ferenci Tamás  
tamas.ferenci@medstat.hu

2018. május 16.

## A megoldási lehetőség

- A feladat tehát: elérni, hogy a csoportok *csak* a vizsgált expozícióban térjenek el! Hogyan biztosítható ez?
- Törekedni sokféleképp lehet rá...
- ... de *biztosra* menni csak egyféleképp: ha az alanyaink *véletlenszerűen* választott része kapja meg az expozíciót
- Ez a *randomizálás* gondolata (R. A. Fisher)
- Figyelem: ehhez szükséges, hogy mi *irányítsuk* az expozíciót, mi határozzuk meg, hogy ki kap ilyen
- Ezt a kutatási módszert, amelyben a kutatók *aktívan befolyásolják*, hogy ki kap expozíciót *kísérletes (experimentális) vizsgálatnak* nevezzük

## A kísérlet

- 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
- 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

## A megfigyelés

- 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...?
- Tenni tehetünk a confounding ellen, de tökéletesen nyugodtak soha sem lehetünk
- Ugyanis ahhoz, hogy tegyünk 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 végezhetőek 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 legjóhízeműbb szervezés esetén is egy „steril” csoportot jelent a kísérletbe bekerülők halmaza, kérdés, hogy az így nyert ismeretek mennyire általánosíthatóak az összes alanyra
- Néha ezek miatt még akkor is csinálnak megfigyelést, ha lehetne (vagy épp történt is) kísérlet!

### Kauzalitás vizsgálata megfigyeléses adatokból

- Annyit mondtunk, 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!
- Plusz, a módszertől függően különböző további feltevések jöhetnek be
- (Megjegyzés: ezek a confounding analitikai fázisban történő korrigálásnak eszközei – tehetünk tervezési fázisban is lépéseket)

## Rétegzés

- Lényege: az expozíciót és a végpontot a confounder(ek) egyes lehetséges értékei (illetve azok kombináció) szerinti kategóriákban, ún. rétegekben külön-külön vizsgáljuk
- Előnyök:
  1. Teljesen univerzális, olyan értelemben, hogy mindenképp megoldja a confoundingot (a figyelembe vett confounderek erejéig, természetesen!), semmilyen további, módszerből adódó feltevést nem tesz szükségessé
- Hátrányok:
  1. Ahogy nő a confounderek száma, egyre nehezkesebb (hiszen nő a rétegek száma), pláne, ha kombinatorikusan nézzük
  2. Még ha nem is száll el nagyon a kombinációk száma, akkor is bizonytalanabb lesz a becslés, hiszen minél több confounder van, annál kevesebb alany jut egy cellába
  3. Nem egyetlen végeredményt kapunk, hanem többet, adott esetben sokat
  4. Probléma van akkor, ha a confounder folytonos

## Standardizáció

- Lényege: úgy indul, mint a rétegzés, de aztán minden expozíció szerint kategóriában egyetlen számot képezünk, úgy, hogy az egyes rétegekbeli értékeket ugyanazokkal a súlyokkal súlyozzuk össze
- Előnyök:
  1. Jól ismert, széles körben alkalmazott, és bizonyos területeken kifejezetten klasszikus módszer (pl. mortalitás összehasonlítása országok között)
- Hátrányok:
  1. Ahogy nő a confounderek száma, egyre nehezkesebb (hiszen nő a rétegek száma), pláne, ha kombinatorikusan nézzük
  2. Még ha nem is száll el nagyon a kombinációk száma, akkor is bizonytalanabb lesz a becslés, hiszen minél több confounder van, annál kevesebb alany jut egy cellába
  3. Kérdéses, hogy hogyan választjuk meg a standardot: egyrészt befolyásolhatja a végeredményt, másrészt sok esetben jó lenne valami egyezményest választani, de ez csak egy-két esetre létezik (pl. nem és életkor)
  4. Probléma van akkor, ha a confounder folytonos

## 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 confoundert 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:

1. A regressziós modellhez nagyon hasonló (pontos összevetésük jelenleg is aktív vita tárgya)