

Kérdezve vagy regisztrálva? Tanácsok és előadások a survey adatfelvételekkel és az adminisztratív adatbázisok használatával kapcsolatban

Surveyed or registered? Advice and presentations on survey data collection and the use of administrative databases.

Hibrid workshop

Delivered under CESSDA Training

November 9, 2022; 13.00–15.30



- A workshop céljai:

1, Megismertetni a közgazdászokat a survey illetve attitűd vizsgálatokkal, az adatfelvételek lehetőségeivel valamint a regiszter adatbázisokkal való összekapcsolódás lehetőségeivel.

2, Az eddig surveyt használók is megismerhetnék az adminisztratív adatbázisok felhasználásának irányvonalait.

- A workshop részei:

- Az első felében a két téma köré szerveződik egy-egy szekció, melyben a kutatók megismerkedhetnek az adott módszer előnyeivel, hátrányaival
- Adott témában meghívott előadó segítségével konkrét kutatási kérdéseikben tanácsokat kapnak.

- **1a. Survey módszer alkalmazási területei, lehetőségei és hátrányai**
- Milyen típusú társadalomkutatási kérdések megválaszolására alkalmas a survey? Milyen alapvető hibákat követhetünk el? Hogyan tudjuk ezeket elkerülni? (Pl. idő dimenziók tisztázása a kérdések során, mintavételi egységek és a kérdések inkonzisztenciája stb.). Hogyan érdemes megtervezni egy adatfelvételt, milyen részei, buktatói vannak. Milyen válaszadási arányokra számíthatunk egyéni vagy vállalati survey-nél? Milyen elvárásaink lehetnek az adatfelvételt végző cég felé, mi az, ami a kutató feladata és mit bízunk a cégre? Adatkezelési, adatvédelmi feladatok: Milyen esetekben kérjünk szakértői, jogi segítséget?
- előadó: Tóth Gergely, ügyvezető igazgató, Pharmaproject-Statiztika Bt.
- előadás 45 perc; kérdések/válaszok, konkrét kutatásokhoz kapcsolódóan is: 45 perc

- **1b. Az adminisztratív adatbázisok általános tulajdonságai, ebből adódó előnyök és korlátok**
- Hogyan épül fel egy regiszter adatbázis? Milyen elemi egységek lehetnek a kiinduló pont (egyének, háztartások, lakás, stb.)? A teljeskörűség elemzési lehetőségei, pl. kis elemszámú társadalmi csoportok, a kutatási témák sokszínűsége. Milyen módon közelítünk az adatokhoz. Már megvalósult kutatási témák bemutatása
- előadók: Köllő János, tudományos tanácsadó KRTK KTI; a KRTK Adatbank vezetője, Tir Melinda KRTK Adatbank munkatárs
- előadás 45 perc; kérdések/válaszok, konkrét kutatásokhoz kapcsolódóan is; 45 perc

- **2. Összekapcsolás lehetőségei, főbb módszerek**

- Megvalósult kapcsolások bemutatása

- Panelbeszélgetés, kérdések/válaszok; 1 óra

- Résztvevők: Koren Miklós KRTK KTI tudományos főmunkatárs, Köllő János KRTK KTI tudományos tanácsadó, Dr. Németh Ádám infokommunikációs szakjogász

- A workshop második részében közös előadáson a kétféle adattípus összekapcsolási lehetőségeiről tartanánk rövid ismertetőt konkrét példákon keresztül.



Gergely Tóth:

Survey módszerről „bővebben”

In.: Surveyed or registered? Advice and presentations on survey data collection and the use of administrative databases.

Hibrid workshop
2022. november 9. 13.00 – 15.30

Survey módszerŐL „BŐVEBBEN”

Tóth Gergely

Pharmaproject-Statisztika Bt

Mottó 1.

„Amilyen könnyű egy kérdőívet elkészíteni, olyan nehéz egy jó kérdőíves kutatás alapján minőségi adatokhoz jutni”

Gallup Story?

Ki ismeri Alf Landon nevét?

És Franklin D. Roosevel-tét?

Minta 1936: Literary Digest 2 Millió vs.

Gallup 50 ezer

1948: Harry Truman vs. Thomas Dewey

Áttekintő

1. Alapok
2. Mintavétel
3. Kérdőíves adatfelvételek típusai
4. Kérdések típusai
5. Lehetséges hibák
6. GDPR

Alapok

- Paradigma
- Módszerválasztás:
Mit szeretnénk megtudni?
- Eredet: Leíró kutatás / várható érték / Felderítő
- Manapság: Magyarázó kutatás /
Hipotézistesztelés

Miben „lehetne” jó a kérdőíves kutatás?

1. Gyors információszerzés
2. Primer kutatás
3. Megbízhatóság
4. Általánosíthatóság

Miben NEM jó a kérdőíves kutatás?

1. Érvényesség (Kvalitatív vs. Kvantitatív)
2. Ha a megfigyelési egység eltér az elemzési egységtől „szépséghibássá” válhat (aggregációs torzítás vagy vonatkoztatási hiba)
3. Könnyen megvezethet (Könnyű adatforrás, de túl sok a csapda)

Konceptualizálás és operacionizálás fontossága



Mennyire elégedett Ön az életével?

vs.

Mennyire boldog Ön?

Idődimenzió

Keresztmetszeti vs. Longitudinális

- Az adatfelvétel idődimenziója (Pl.: választáskutatás, válság), de kérdőív hossza is...
- A kérdések vonatkoztatási idődimenziója (Jövőre, múltira?)

Retrospektív kutatások elég rosszul működnek. Az adatforrás torzít (identitás megőrzésének igénye)

Adatfelvétel típusai

- PAPI
- CAPI
- CATI
- CAWI
- Kombinált / Push-to-web

Mintavétel

Reprezentativitás VS. Valószínűségi mintavétel

- Reprezentativitás: Adott dimenziókra nézve! (nem, kor, iskola, régió) (Kvótás?)
- Valószínűségi: ismert bekerülési valószínűség
- Súlyozás
 - Design
 - Utólagos rétegzési (Válaszmegtagadási arány)
 - Projekáló
- Mintavételi hiba vs. nem mintavételi hiba (torzítás)
- „Design hatás”

Kérdések típusai

1. Nyitott vs zárt
2. Egy válaszós vs több válaszós
3. Táblázatos
4. Grafikus (HTML: „csillagos ég”)

Kérdések típushibái

1. Nem érthető / nem plauzibilis a kérdezőnek

„Hogyan értékeli Ön a jövő évi költségvetést?”

- Minta összetétele
- Próbakérdés fontossága

2. Sugalmazó kérdések

„A szankciók az élelmiszer-ellátásra is súlyos hatással vannak. Ön egyetért az élelmiszerárak emelkedését okozó szankciókkal?”

3. Dupla-triplacsövű kérdések

„Mennyire elégedett Ön a jelenlegi fizetésével és a béren kívüli juttatásokkal?”

Kérdések típushibái

4. Kerülendők a mindenre kiterjedő kérdések

Mindenki / senki / mindenkor

„Ön szokott minden reggel enni?” Igen /nem

5. Tagadó /dupla tagadó kérdések kerülendők

„Ön sem ért egyet avval az állítással, hogy nem kéne emelni a szociális kiadásokat?”

Jó kérdőív jellemzői

- Rövid
- Érthető
- „Storytelling” – íve van és élményt ad
- Nem manipulál

GDPR

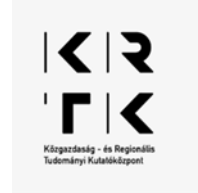
- Adatvédelmi szempontok
- Személyes adatkezelés! (Google?)
- Tájékoztatás kötelezettsége
- Tudományos kutatás (Etikai engedély? Pl.: kísérlet)

Záró mottó

Kérdőívet készíteni olyan mint házat építeni: csak befejezni lehet, de sosem készül el teljesen...

Köszönöm a figyelmet!

ppstatisztika@gmail.com



Melinda Tir:

Kérdezve vagy regisztrálva: A regiszter adatokról

In.: Surveyed or registered? Advice and presentations on survey data collection and the use of administrative databases.

Hibrid workshop
2022. november 9. 13.00 – 15.30

Kérdezve vagy regisztrálva: A regiszter adatokról





Rövid Áttekintés

- A regiszter és a kapcsolt regiszter adatokról
- KRTK Adminisztratív Adatkapcsolásainak bemutatása
- Mitől más egy kapcsolt adminisztratív panel, mint a hagyományosan rendelkezésre álló adatforrások?
- Milyen módon közelítsünk az adatokhoz?
- Néhány kutatási példa bemutatása



A regiszter és a kapcsolt regiszter adatokról

Regiszter adatok:

- Leginkább államigazgatási szereplők feladataik ellátásához kötődő nyilvántartási, engedélyezési, jogosultsági adatbázisai.
- Jellemzője a célcsoport teljes lefedettsége és egyedi azonosíthatósága.
- Példák: egyes KSH OSAP felvételek, NAV bevallások, egészségügyi, oktatási, munkanélküli regiszterek



Kapcsolt regiszter adatok:

- Alapja a 2007 CI. Törvény (visszavonva)
- Alapelve, hogy egy vagy több közös azonosító mentén összekapcsolható regisztereket/felvételeket ugyanazon mintára (max 50 %), ugyanarra az időszakra egyesíti.

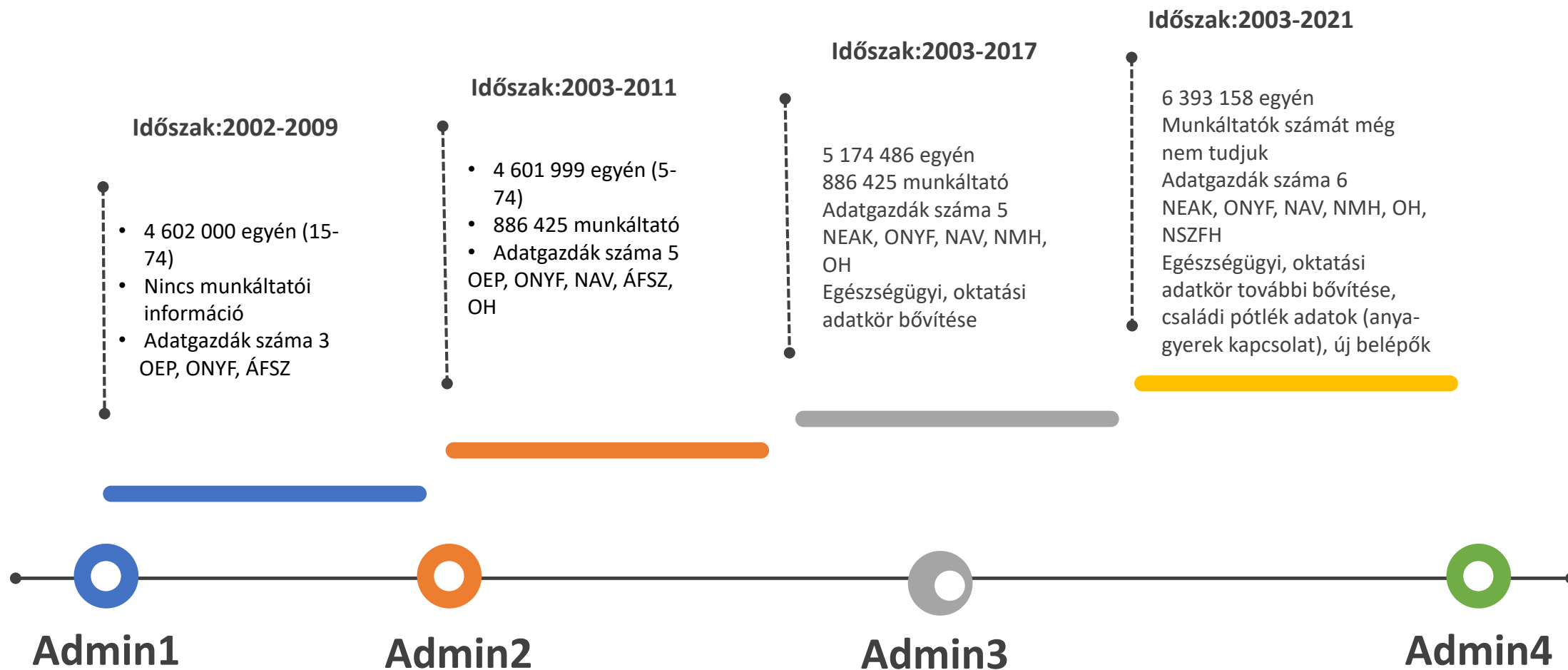
Megvalósult adatintegrációk:

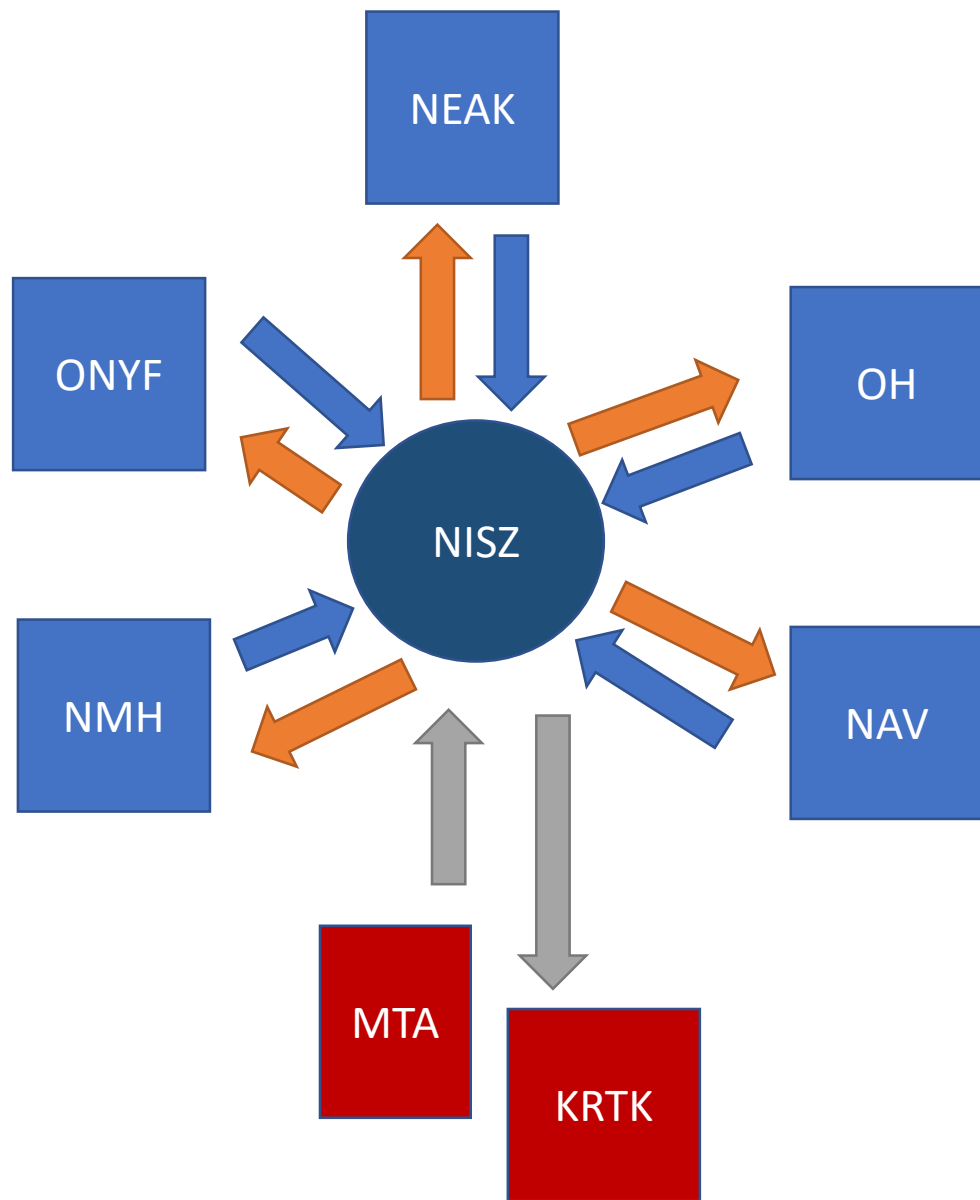
- Pénzügyminisztérium
- Oktatási Hivatal (Diplomás pályakövetés)
- MNB
- KRTK



KRTK Adminisztratív Adatkapcsolásai

Eddig megvalósult/folyamatban lévő adatkapcsolások





Admin3 példáján az adatkapcsolás folyamata (2007.CI, tv szerint)

MTA Elnök → NISZ a 2007.CI tv alapján

NISZ → adatgazdák: hash1 (TAJ), hash2 (adóazonosító) kódoló algoritmusok

NEAK: 50%-os TAJ minta (2003.01.01.).

Adatgazdák → TAJ alapján leválogatott állományok, hash1

ONYF, NAV: vállalati adóazonosítók alapján leválogatott állományok, hash2

NISZ: anonimizálás

NISZ → KRTK: kapcsolt állomány

KRTK: spell szerkezetűből fix formátumú havi panel. Javítás, tisztítás, tesztelés: 2-3 év

Admin 3

5 millió ember X 180 hónap (2003-2017). Minden hónapban tudjuk, hogy az (anonim) személy:



Adattisztítás folyamata: adatkéréstől a használatig



01

Adatkérés, az adat beérkezése

Adatgazdákkal történő részletes, mező szintű egyeztetés, szerződés és mellékleteinek véglegesítése kb. 1,5-2 év

02

Validálás, tisztítás, építés

Külső adatforrásokkal történő összevetés, nyilvántartási változások összegyűjtése. Adatkörök feldolgozása külön-külön, spell szintű formátum havi szintre alakítása, kutatási kérdéseknek megfelelő változók és adatkörök kialakítása, majd az egyes adatkörök összeépítése, keresztmetszeti és longitudinális vizsgálatok lefolytatása

03

Kollektív fázis, kutatói visszacsatolás

KRTK kutatók számára történő kijaánlás, minikutatások indítása. A fájlok még többször javításra kerülnek. A kutatók a használat során tapasztalt problémákat összegyűjtik. Ez egy iteratív 1-2 éves munka

04

Visszajelzések beépítése, adatbázis megnyitása

A visszajelzések összegyűjtése és beépítése után kb. 2 évet követően áll elő az a véglegesnek tekintett teljes adatbázis készlet és dokumentáció melynek használatát már ki tudjuk ajánlani nem KRTK-s kutató/hallgató számára is.



**Mitől más egy kapcsolt adminisztratív panel,
mint a hagyományosan rendelkezésre álló
adatforrások?**



- KSH-MEF:Részletgazdag és panel, de csak 1.5 év a követés negyedévente pontszerűen. Minta elemszám alig 1%-ka egy Adminnak.
- Bértarifa:Kiemelkedő adatsor, hosszú idősor, de csak a vállalatok követhetőek az egyének nem
- Vállalati Mérlegek:Hosszú idősor, de a vállalati szintnél mélyebbre nem látunk, nincs munkavállalói adat
- Egyéb regiszterek önmagukban: részletes nagyszámú információ, de az egyes létállapotok előzményeit és következményeit nem látjuk



Előny:

- Elemszám kellően nagy (kb. 5 millió ember)
- Egyesít több regiszter adatot, így lehetővé válik a folyamatok együttes elemzése
- Hosszú, egyéni idősor havi, de akár napi szintű adatokkal

Hátrány:

- Hosszas előkészítés
- Frissítés korlátossága
- Háztartási szint és iskolai végzettség hiánya. Csak azt tartalmazza amiről van nyilvántartás



**Milyen módon közelítsünk az
adatokhoz?**



- A megfigyelhető egység az adminisztratív státuszok változása, tehát vélemények helyett konkrét viselkedések vizsgálhatók.
- Legyünk tisztában az adat keletkezésének körülményeivel, nem a kutatásunkra jött létre!!!
- Nagy elemszám miatt akár kisebb csoportok vizsgálatára alkalmas, de az adatbázisok belső dinamizmusa, fluktuációja miatt itt is tapasztalhatók minimális eltérések
- Elemzői döntések nyomán nyerhető ki a kutatási kérdéseknek megfelelő adattartalom. Az elemzéshez használt változókat és azok számítási módját iteratív módon, kvázi menet közben érdemes véglegesíteni

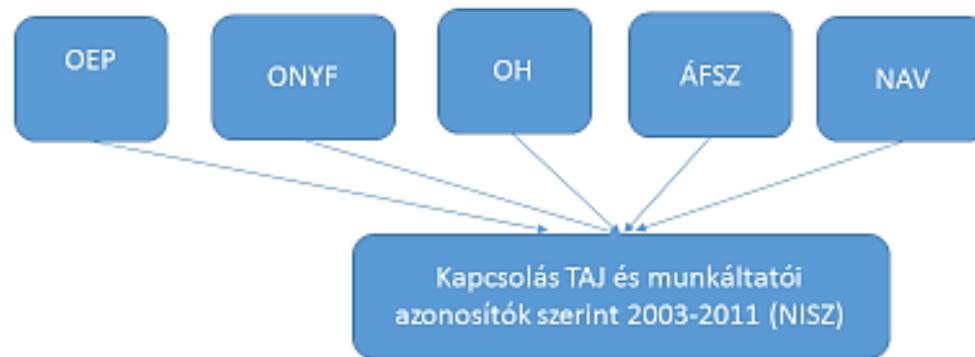


Néhány Kutatási példa bemutatása

Mekkora az orvoselvándorlás és visszavándorlás? Varga Júlia (2017), Magyarország

Probléma: a kivándorlási szándékfelmérések és engedély-kiváltások nem mérvadóak, a visszavándorlásról pedig végképp semmit sem mondanak.

Feltevés: ha egy orvos nyomtalanul eltűnik a rendszerből (és nem halt meg), az nagy valószínűséggel külföldön van

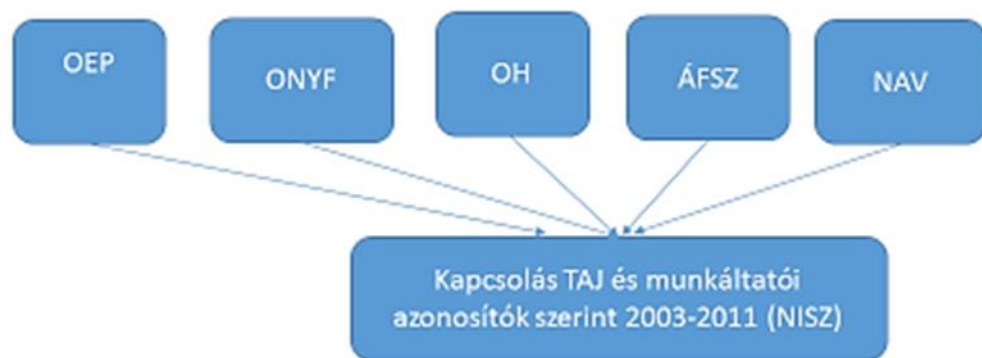


Eredmény: 2003 január és 2011 december között a praktizáló orvosok 12%-a hagyta el az országot (az inaktívvá válókat leszámítva 14%). Nagyobb hullámok: 2010 tavasza, 2013 (A, D). 50% visszatért. Ennél is nagyobb veszteséget okozott a belföldi pályaelhagyás.

Többet keresnek-e a multiknál dolgozók?

Probléma: a multiknál dolgozók – esetleg – a belföldi cégeknél is többet keresnének. És lehet, hogy az előnyük múlékony, megszűnik, ha átlépnek a belföldi szektorba.

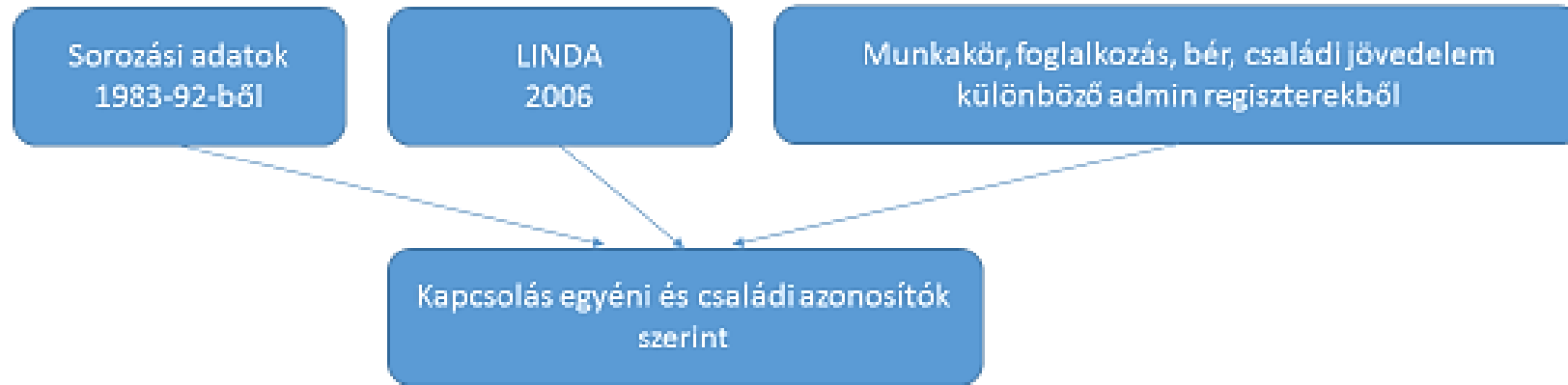
Kulcs: a meg nem figyelt heterogenitást kiszűrő fix hatások használata, hasonló belépők összehasonlítása a hazai szektorban, a külföldiektől érkezettek arányának hatása az „inkumbens” kollégáikra.



Eredmény: A multiknál dolgozók többet keresnek, de nem 112%-kal, hanem csak 25%-kal, ha kontrollálunk a megfigyelhető és meg nem figyelhető tulajdonságaikra. Kilépés után a hazai cégeknél dolgozva is többet keresnek 14%-kal. Egy szórás egységgel több multitól érkezett kolléga → 3%-kal magasabb bér.

Milyen hatást gyakorolnak a kognitív és nem kognitív készségek a munkaerőpiaci kimenetekre? Lindquist & Vestman 2011, Svédország

Probléma: a sikeres munkaerőpiaci karrier maga is fejleszti a ~ készségeket. Korábbi információ kell.

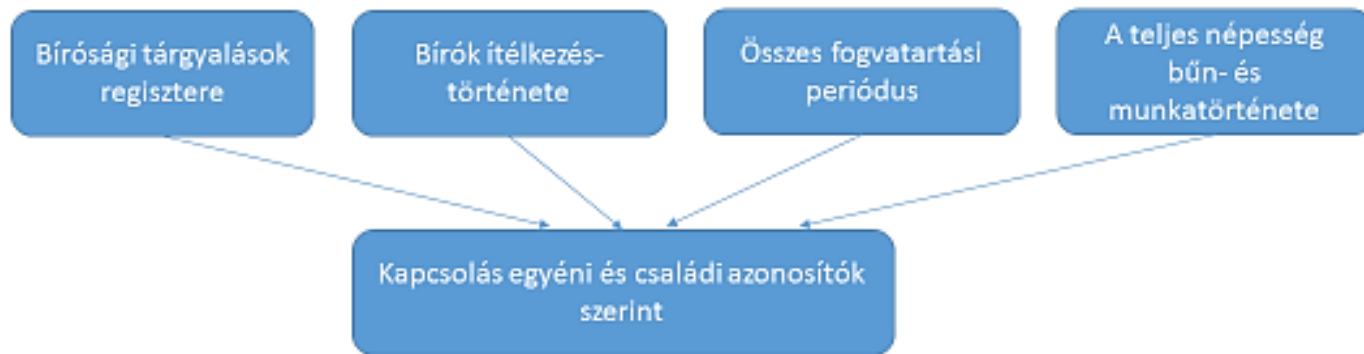


Eredmény: a nem kognitív készségek erősebb szerepet játszanak, különösen a munkaerőpiac „alsóbb” régiójában + (kapcsolódó kutatásból) a szerepük időben rendkívüli mértékben erősödött.

Van-e visszatartó ereje a börtönbüntetésnek az elítélt büntársaira és rokonaira? Bhuller et al. 2018, Norvégia

Probléma: bűnözésre hajlamosító környezet → annak, hogy A-t letöltendőre ítélték és társa, B később bűnt követ el, közös oka van → OLS: pozitív kapcsolat A bebörtönzése és B későbbi bűnelkövetése között → megfelelő instrumentumra (IV) van szükség

IV: az ítéletet meghozó bíró *átlagos* ítélezési szigora (Kling 1999, Bhuller et al. 2016)

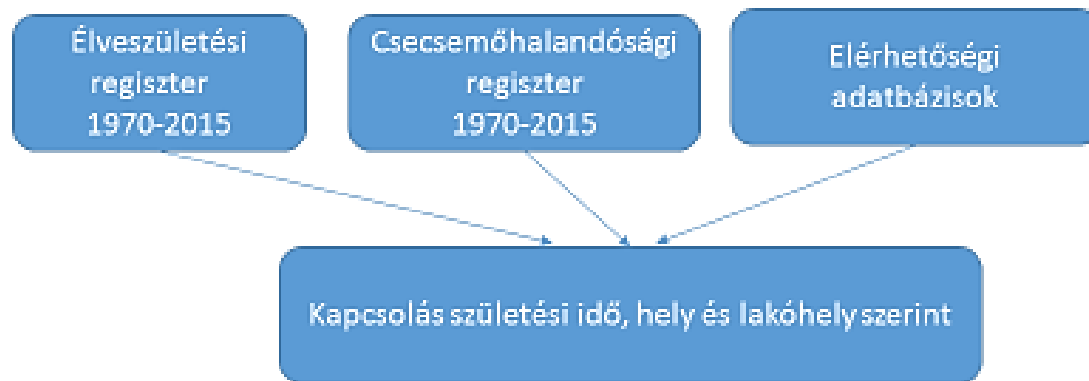


Eredmény: tettestársak későbbi bűnelkövetése: -51%., Fiatalabb fiútestvér: - 32%, 4 éves időtávon.

Csökkentették-e a csecsemőhalálozást a Perinatális Intenzív Centrumok? Hajdu et al. (2018), Magyarország

Probléma: nem biztos, hogy csökkentették, empirikus kérdés.

Kulcs: a PIC-III hálózat fokozatosan épült ki. Kihasználható információ a PIC elérhetőségének változása.

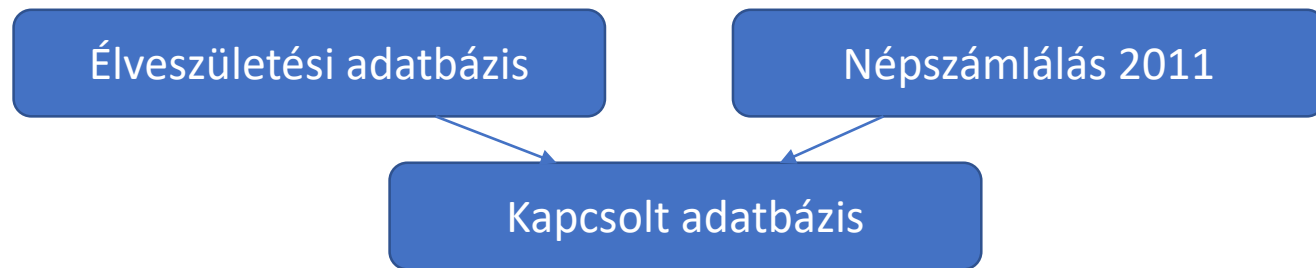


Eredmény: a csecsemőhalandósági ráta 2011-2015 között e hatás nélkül a ténylegesen megfigyelt érték közel kétszerese lenne. Különösen erős hatás a nagyon kis súllyal születettekre.

Hatott-e a kötelező iskolázási kor felemelése (16→18) a roma lányok serdülőkori fertilitására? (Adamecz – Scharle 2018)

Probléma: a naptári idő is telik, a fertilitás változása esetleg csak egybeesik a szabályozási változással („akkor, tehát azért”)

Kulcs: a közvetlenül 1991 júniusa előtt illetve után születettek (nagy valószínűséggel) a 16 illetve 18 éves korhatár vonatkozott. Ők lényegében azonos korúak, a környezet hasonlósága kontrollokkal biztosítható, ebben viszont erősen különböznek.



Eredmény: A korhatáremelés 6,8 százalékponttal (26%) csökkentette annak a valószínűségét, hogy egy roma serdülő 18 éves kora betöltése előtt anya lesz. Az emelés az iskolaév során csökkentette a teherbe esés valószínűségét, a nyári és a téli szünetekben nem befolyásolta azt.

Anna Adamecz-Völgyi – Ágota Scharle (2018): The Effects of Increased Compulsory School Leaving Age on the Teenage Fertility of Roma Women, a Disadvantaged Ethnic Minority, BWP 2018/2.

