

A TERMÉSZETES KÍSÉRLET ALKALMAZÁSA A DEMOGRÁFIAI ÉS SZOCIOLÓGIAI KUTATÁSBAN

Moksony Ferenc

ÖSSZEGZÉS

A tanulmány célja annak a kutatástípusnak a bemutatása, amely természetes kísérletként ismert a módszertani szakirodalomban, és amelyet mind gyakrabban használnak a társadalomtudományokban. Míg a valódi kísérletben a kutató maga hozza létre a magyarázó változó különböző értékeit, és ő rendeli hozzá véletlenszerűen ezekhez az értékekhez a megfigyelési egységeket, addig a természetes kísérletben a dolgok szerencsés alakulása hoz létre olyan helyzetet, amelyben a magyarázó változó különböző értékeket vesz fel, a megfigyelések pedig gyakorlatilag – vagy akár ténylegesen is – véletlenszerűen rendelődnek hozzá ezekhez az értékekhez. A cikk első része a természetes kísérlet korai példáit, második része pedig a módszer modern alkalmazásait tekinti át. A harmadik rész azokkal a kutatásokkal foglalkozik, amelyek valamilyen lottósorsoláson alapulnak, s amelyekben ezért a randomizálás követelménye nagyobb eséllyel teljesül. Ez a rész tárgyalja azt a problémát is, amellyel akkor találkozunk, amikor a kutató által összehasonlítani kívánt csoportok bizonyos mértékben nem véletlen szelekció eredményei. A tanulmány befejező része abból a szempontból hasonlítja össze a természetes kísérletet néhány más kutatástípussal, hogy melyik módszer mennyire képes egyidejűleg biztosítani a belső és külső érvényességet.

Tárgyszavak: természetes kísérlet, kutatástervezés, randomizálás, demográfia, szociológia

Moksony Ferenc, Budapesti Corvinus Egyetem, Kommunikáció és Szociológia Intézet, 1093 Budapest, Fővám tér 8.

E-mail: fmokson@gmail.com

BEVEZETÉS

Egy kutatás minőségének megítélésekor az egyik legfontosabb szempont a kapott eredmények *belső érvényessége*. Ez a fogalom arra a kérdésre vonatkozik, hogy a magyarázó és a függő változó között megfigyelt összefüggést valóban az általunk feltételezett és az adott kutatás keretében vizsgált oksági mechanizmus hozta-e létre – nem pedig egy *azzal versengő másik* hatásmechanizmus.¹ Egy, a születések számának növelését célzó társadalompolitikai program eredményességének a vizsgálatakor például ez a kérdés úgy vetődik fel, hogy a gyermekek számának esetleges emelkedése valóban a programnak köszönhető-e, nem pedig, mondjuk, a szülőképes korú nők számában – a programmal egy időben – bekövetkezett kedvező irányú változásnak.² Minél inkább igennel válaszolhatunk erre a kérdésre, annál nagyobb a kutatásunk *belső érvényessége*.

Amint az az előzőekből is kitűnik, a *belső érvényesség* fogalma szorosan kapcsolódik az *alternatív magyarázat*éhoz: minél nagyobb a sikeresen kiküszöbölt alternatív magyarázatok száma, azaz minél kevesebb lehetséges vetélytársa marad az általunk feltételezett oksági mechanizmusnak, annál biztosabb lábakon állnak az adatokból levont következtetések, s így annál nagyobb a kutatásunk *belső érvényessége*. A *belső érvényesség* növeléséhez tehát az alternatív magyarázatok kiküszöbölésén keresztül vezet az út, ez utóbbi legjobb eszköze pedig a

¹ A *belső érvényesség* (internal validity) fogalmát Donald T. Campbell vezette be még az 1950-es években a módszertani szakirodalomba, szembeállítva azt a *külső érvényességgel* (external validity), ami a kutatási eredmények általánosíthatóságára vonatkozik (Campbell, 1957). Későbbi munkáiban (Cook and Campbell, 1979; Shadish, Cook and Campbell, 2002) megkülönböztette az érvényesség két másik típusát is: egyrészt a *statisztikai érvényességet* (statistical conclusion validity), ami azokkal a feltételekkel kapcsolatos, amelyek a változók közötti összefüggések erősségének és statisztikai szignifikanciájának a megállapításához szükségesek; másrészt pedig a *mérési érvényességet* (construct validity), ami a kutatásban használt konkrét változók és az azok által megragadni kívánt általánosabb fogalmak kapcsolatát érinti. Az érvényesség különböző fajtáinak ez a Campbell nevéhez köthető rendszerezése jelentősen eltér a bevezető módszertani vagy kutatástervezési tankönyvekben általában tárgyalt típusoktól (pl. face validity, predictive validity). Ez utóbbiak lényegében mind az elméleti fogalmak és az azokat mérni hivatott változók közötti viszonyra vonatkoznak, s ennyiben úgy tekinthetők, mint a Campbell által mérési érvényességnek nevezett kategória alesetei. Összességében a kétféle osztályozás közül a Campbell-féle az átfogóbb, egyszersmind – megítélésem szerint – a hasznosabb, termékenyebb is.

² Ez volt a helyzet például a Magyarországon 1973-ban bevezetett népesedéspolitikai intézkedéscsomag esetében: a bevezetéssel nagyjából egy időben értek szülőképes korba az 1950-es évek „Ratkó-nemzedékének” tagjai, aminek köszönhetően jelentősen megnőtt a népességben a gyermekvállalásra képes nők száma.

randomizálás, vagyis a megfigyelési egységeknek a kutató által vizsgált oksági tényezőt képviselő magyarázó változó egyes értékeihez vagy kategóriáihoz történő *véletlenszerű hozzárendelése*.³ Egy randomizált kísérletben a kutató maga hozza létre a magyarázó változó egyes értékeit, a megfigyelési egységeket (például egyéneket) pedig *véletlenszerűen rendeli hozzá* ezekhez az értékekhez. A randomizált – vagy ahogy sokszor nevezik: valódi – kísérletet ez a két sajátossága különbözteti meg egyrészt a *kvázi-kísérlettől* – amelyben a kutató maga hozza ugyan létre a magyarázó változó egyes értékeit, azonban a megfigyelési egységeket nem véletlenszerűen rendeli hozzá ezekhez az értékekhez –, másrészt a társadalomtudományokban leginkább elterjedt *megfigyeléses vizsgálatoktól*,⁴ amelyekben nem csupán a randomizálás hiányzik, de a magyarázó változó egyes értékeit sem a kutató hozza létre, hanem pusztán feljegyzi azokat (*lásd 1. táblázat*).

1. táblázat. Valódi kísérlet, kvázi-kísérlet, megfigyeléses vizsgálat

	Valódi kísérlet	Kvázi-kísérlet	Megfigyeléses vizsgálat
Kutatói kontroll a magyarázó változó fölött	Igen	Igen	Nem
Véletlen hozzárendelés	Igen	Nem	Nem

A valódi kísérlet – amint az az eddig elmondottakból is érzékelhető – a kutató aktív részvételére épül: ő teremti meg a magyarázó változó különböző értékeit, és ő rendeli hozzá véletlenszerűen ezekhez az értékekhez a megfigyelési egységeket. Néha azonban a természet⁵ veszi át mindkét téren a kutató szerepét,

³ A randomizálás – amely Ronald Fisher munkásságának köszönhetően került be az empirikus kutatás eszköztárába (Armitage, 2003; Hall, 2007) – úgy küszöböli ki az alternatív magyarázatokat, hogy *véletlenszerűen osztja szét* a vizsgált magyarázó változó egyes kategóriái között azoknak a *zavaró változóknak az értékeit*, amelyek enélkül hamis korrelációt hozhatnának létre a magyarázó és a függő változó között. Ezzel szemben az alternatív magyarázatoknak a nem kísérleti vizsgálatokban alkalmazott módszerei – így a kontrollváltozó használata, illetve a demográfiaiban elterjedt standardizálás – úgy szüntetik meg az ilyen hamis korrelációkat, hogy a zavaró változó értékét *statisztikai úton – átlagosan – azonosá teszik* a magyarázó változó egyes kategóriáiban. A randomizálásról és az alternatív magyarázatok kiküszöbölésének egyéb módszereiről bővebben lásd Moksony, 2018: 171–192.

⁴ A megfigyeléses vizsgálat (observational study) elnevezés mellett a korrelációs vizsgálat (correlational study) kifejezéssel is találkozhatunk a szakirodalomban (pl. Lau and Kuziemsky, 2016). Ez azonban eléggé megtévesztő, ugyanis a korreláció a statisztikai elemzéshez tartozó fogalom, itt pedig a kutatás felépítéséről, megtervezéséről van szó. A korrelációs vizsgálat elnevezés tehát indokolatlanul összekeveri az elemzés és a kutatástervezés kérdését, ezért célszerűbb a megfigyeléses vizsgálat kifejezést használni helyette.

⁵ A „természet” szó itt *tág értelemben* szerepel, és minden, *a kutató saját aktív részvételétől független* folyamatot felölel – legyen az akár szorosan vett természeti folyamat (mint amilyen például egy földrengés vagy egy hurrikán pusztítása), akár pedig más emberek cselekedete (mint amikor például egy kormány módosít valamilyen jogszabályt, vagy amikor egy vállalat megváltoztatja az addigi működését). A lényeg, hogy a vizsgálati helyzetet *nem maga a kutató hozza létre*, ő csupán *felismeri és kihasználja* azt.

s a *dolgok szerencsés alakulása hoz létre olyan helyzetet*, amelyben a magyarázó változó különböző értékeket vesz fel, a megfigyelések pedig gyakorlatilag – vagy akár ténylegesen is – véletlenszerűen rendelődnek hozzá ezekhez az értékekhez. Ilyenkor beszélünk *természetes kísérletről* (Freedman, 2005: 6; Robinson et al., 2009; Dunning, 2012). A természetes kísérlet – szigorúan véve – nem több pusztán megfigyeléses vizsgálatnál, a kedvező körülményeknek és az azokat kihasználó képes kutató találékonyságának köszönhetően azonban belső érvényességét tekintve mégis felveszi a versenyt az „igazi” randomizált kísérlettel.

A TERMÉSZETES KÍSÉRLET KORAI ALKALMAZÁSAI

A természetes kísérlet alighanem legkorábbi, egyszersmind legismertebb példája a XIX. században élt angol orvos, John Snow nevéhez fűződik (Freedman, 1999: 244–247; Morabia, 2004: 39–42). Az 1800-as években több kolerajárvány is pusztított Angliában, ezen belül Londonban. Akkoriban az orvosok túlnyomó többsége úgy gondolta, a kolerát a rossz levegő és a rossz szagok okozzák; ez volt az ún. miazma-elmélet. Snow nem hitt ebben az elméletben; úgy vélte, a fertőzésekért a szennyezett ivóvíz a felelős. Ezt az elképzelését szigorú értelemben vett randomizált kísérlettel nyilvánvalóan nem tudta ellenőrizni – nem lehetett emberek a tudományos igazság kedvéért véletlenszerűen szennyezett víz fogyasztására kényszeríteni –; a körülmények szerencsés alakulása – és persze a saját leleményessége – folytán azonban lehetősége nyílt egy természetes kísérlet elvégzésére.

Az 1800-as évek közepe táján Londonban az emberek egy része még hagyományos kézi mozgatású kutakból nyerte az ivóvizet, más részük azonban már a maihoz hasonló vezetékes vízellátásban részesült. A vezetékes vizet több, egymással versengő társaság szolgáltatta, köztük a Southwark & Vauxhall Company és a Lambeth Company. A város egy részében a háztartások gyakorlatilag véletlenszerűen kerültek e két cég valamelyikéhez; előfordult, hogy az utca egyik oldalán álló ház az egyik, míg a túloldalon álló épület a másik társasághoz tartozott. Ezért aztán a két cég által ellátott háztartások társadalmi összetétele rendkívül hasonló volt; ahogyan maga Snow megfogalmazta: „Mindegyik társaság fogyasztói között vannak szegények és gazdagok, nagy házak és kicsik, nincs semmilyen különbség a két társaság fogyasztóinak a körülményeiben vagy foglalkozásában.” (Snow, 1855: 75)

Az 1840-es években a két társaság a Temzének ugyanarról a szakaszáról vette a vizet – mégpedig a szennyezett és tisztítatlan vizet. Ennek megfelelően a kolera

előfordulási aránya hasonló mértékű volt a két cég fogyasztói körében. 1852-ben azonban az egyik cég, a Lambeth Company áthelyezte vezetőkeit a Temze egy feljebb lévő szakaszára, ahol – mert a folyó ott még nem érte el Londont – tisztább volt a víz. Bár ennek a változásnak semmi köze nem volt Snow tudományos céljaihoz, ő felismerte a kínálókozó lehetőséget, és összehasonlította egymással a két társaság fogyasztóit a kolerahalálozás gyakorisága szempontjából. Az eredmények szerint a továbbra is szennyezett vizet szolgáltató Southwark & Vauxhall Company által ellátott háztartásokban sokkal nagyobb, csaknem 9-szer akkora volt a kolera okozta halálozás mértéke, mint azokban, amelyek a tisztább folyószakaszt használó Lambeth Company vezetőkeiből nyerték az ivóvizet (2. táblázat).

2. táblázat. Kolerahalálozás a két vízszolgáltató társaság fogyasztói körében

Vízszolgáltató társaság	Halálozások száma	10.000 házra jutó halálozás
Southwark & Vauxhall	1,263	315
Lambeth	98	37

Forrás: Snow, 1855: 86 (Table IX)

Módszertani szempontból Snow kutatása nem több pusztán megfigyelés vizsgálatnál; hiányzik belőle mind a szigorú értelmében vett randomizálás, mind pedig a magyarázó változó értékének a befolyásolása. Nem Snow döntötte el, hogy az egyik cég tisztább, a másik szennyezettebb vizet szolgáltasson, és nem rajta múlt, hogy az egyes háztartások miként kerülnek egyik vagy másik társasághoz. A körülmények szerencsés alakulása folytán azonban – mintegy a természet ajándékaként – mégis a valódi kísérlet mindkét feltétele gyakorlatilag tökéletesen megvalósult. Ahogyan maga Snow megfogalmazta: „Mivel semmilyen különbség sincs a két társaság vizét fogyasztó háztartások vagy emberek között, ... nyilvánvaló, hogy semmilyen tudatosan megtervezett kísérlet nem ellenőrizhetné alaposabban a vízszolgáltatásnak a kolera terjedésére gyakorolt hatását, mint ez, amit a körülmények készen tártak a megfigyelő elé.” (Snow, 1855: 75)

A természetes kísérlet egy másik – talán kevésbé ismert, de a mi szívünknek annál kedvesebb – példája Semmelweis Ignác nevéhez fűződik (Czeizel, 2006). Semmelweis a bécsi egyetemen szerezte orvosi diplomáját, és később ott is folytatta tevékenységét. Bécsben akkortájt két szülészeti klinika működött; az, hogy a szülő nők közül ki melyikbe került, gyakorlatilag véletlenszerű volt: egyik nap az egyik, másik nap a másik intézmény fogadta a betegeket. Az 1830-as évek végéig a két klinikán nagyjából egyforma volt a halálozási arány. 1840-től kezdődően azonban megváltozott a helyzet: amint az a 3. táblázatban is látható,

ekkortól az I. számú klinikán lényegesen nagyobb volt a halálozás mértéke, mint a másik intézményben.

Mivel magyarázható ez a jelentős változás? Mi történt 1840-ben? Ettől az évtől kezdve az I. sz. klinika csak az orvostanhallgatók, a II. sz. klinika pedig csak a bábajelöltek oktatását végezte. És mi köze volt ennek az eltérő mértékű halandósághoz? Az, hogy az orvostanhallgatók számára kötelező volt a boncolásokon való részvétel, a bábajelöltek számára viszont nem. Az I. sz. klinikán tapasztalt magas halandóságot az idézte elő, hogy az orvostanhallgatók – megfelelő fertőtlenítés hiányában – a boncteremből magukkal hurcolták a kórokozókat a szülőszobába.

3. táblázat. Gyermekágyi halálozás a bécsi I. és II. sz. szülészeti klinikán

Év	I. sz. klinika	II. sz. klinika
1840	9,5	2,6
1841	7,8	3,5
1842	15,8	7,6
1843	9,0	6,0
1844	8,2	2,3
1845	6,9	2,0
1846	11,4	2,8

(%)

Forrás: Czeizel, 2006: 34.

Ha elvonatkoztatunk a konkrét tartalomtól, és megnézzük ennek a példának a formai sajátosságait, akkor láthatjuk, hogy kísértetiesen hasonlít Snow londoni vizsgálatára. Mindkét esetben lényegében teljesül a valódi kísérlet két fő követelménye – a randomizálás és a magyarázó változó értékének aktív befolyásolása –, csak éppen nem a kutató, hanem a természet „beavatkozásának” köszönhetően. Londonban a háztartások gyakorlatilag véletlenszerűen kerültek a két vizsgálótársaság valamelyikéhez, és Bécsben a szülő nők szintén gyakorlatilag véletlenszerűen kerültek a két klinika valamelyikébe. Ugyanígy, Londonban kezdetben mindkét vizsgálótársaság a Temzének ugyanarról a szakaszáról vette a vizet, majd az egyik társaság megváltoztatta a vízvételi helyet, s ezzel ami addig konstans volt, változóvá vált. Hasonló folyamat zajlott le Bécsben: míg korábban nem különült el a más-más foglalkozásra készülő diákok képzése, 1840-től az egyik klinika csak orvostanhallgatókat, a másik pedig csak bábajelölteket oktatott.

Snow és Semmelweis vizsgálatának együttes elemzése több szempontból is hasznos. Egyrészt a két példa jól érzékelteti, hogy amennyiben a kutató képes

észrevenni és kihasználni a körülmények szerencsés alakulásában rejlő lehetőségeket, akkor egy olyan kutatás is vezethet belső érvényességüket tekintve a valódi kísérlettel felérő eredményekhez, amely pedig szigorú értelemben véve nem több passzív megfigyelésnél. Ehhez persze kell az előző mondatban említett mindkét dolog: egyrészt a körülmények szerencsés alakulása, másrészt az a kreatív kutató, aki felismeri és kiaknázza a kedvező helyzetből adódó lehetőségeket. Mert ahogyan Louis Pasteur mondta: „a véletlen csak a felkészült elméknek kedvez”. (Pasteur, [1854] 1939: 131)

Egy másik tanulság, hogy ha egy kutatás erős a *megtervezését, felépítését* („design”-ját) illetően, akkor az *adatelemzés* szakaszában viszonylag egyszerű módszerek is elegendők. Snow vizsgálatában például két halálozási arányszám összehasonlítása kellő bizonyítékot tudott szolgáltatni a szennyezett víz oksági hatására vonatkozólag. Gondoljuk meg, mit érne egy ugyanilyen összehasonlítás, mondjuk, egy hagyományos kérdőíves adatfelvétel esetében; rögtön elkezdenénk sorolni a lehetséges alternatív magyarázatokat, s azokat a tényezőket, amelyeket még be kellene vonni kontrollváltozóként az elemzésbe ezeknek az alternatív magyarázatoknak a kiküszöbölése érdekében. Ennek köszönhetően azután az egyszerű táblázatból pillanatok alatt bonyolult lineáris vagy logisztikus regressziós modell lenne. Ezek az elemzési technikák mind azt igyekeznének – több-kevesebb sikerrel – pótolni, ami a kutatás felépítéséből hiányzik: a különféle torzító tényezőktől való mentességet.

A TERMÉSZETES KÍSÉRLET MODERN ALKALMAZÁSAI

Az eddig említett példák alapján azt hihetnénk, a természetes kísérlet valamiféle muzeális dolog, amit csak a XIX. században használtak. Ez azonban nincsen így, sőt, napjainkban éppenséggel növekszik ennek a kutatástípusnak a népszerűsége. Jól látszik ez az 1. ábrán, amely a „natural experiment” kifejezés relatív gyakoriságának a Google digitalizált könyvanyagában megfigyelhető időbeli változását mutatja.

A természetes kísérlet modern alkalmazását jól szemlélteti az a kutatás, amely a rászoruló fiatalok részére nyújtott anyagi támogatásnak a továbbtanulási esélyekre gyakorolt hatását vizsgálta az Egyesült Államokban (Dynarski, 2003; ennek a kutatásnak a természetes kísérlet fogalomrendszerében történő értelmezéséhez lásd Murnane and Willett, 2011: 141–145).

1. ábra: A „natural experiment” kifejezés relatív gyakoriságának időbeli változása a Google digitálizált könyvanyagában



Megjegyzés: Ez az ábra a Google Ngram Viewer nevű keresőrendszerének segítségével készült, amely egy-egy kifejezés relatív – a cég által digitalizált több millió könyv teljes szövegéhez viszonyított – gyakoriságának időbeli változását jeleníti meg grafikus formában. Bár ennek a szolgáltatásnak a tudományos célú felhasználásáról megoszlanak a vélemények (pl. Michel et al., 2011; Zimmer, 2013; Pechenick, Danforth and Dodds, 2015; Pettit, 2016; Younes and Reips, 2019), a vitákban felmerülő problémák zöme a nagy léptékű – adott esetben több évszázadot felölelő – társadalmi-kulturális változások vizsgálatára vonatkozik, és olyan kérdésekkel (pl. a szavak hosszú távú jelentésváltozásával, a tudományos jellegű publikációknak az összes digitalizált könyvön belüli növekvő arányával – és ezáltal a mindennapi életben előforduló kifejezések látszólagos ritkábbá válásával –, vagy éppen a karakterfelismerő programok korszakokonként eltérő megbízhatóságával) kapcsolatos, amelyek csak korlátozottan érintik az Ngram Viewer itteni, alapvetően illusztratív célú alkalmazását. Így például az ábrán megjelenített időszak alatt a könyvek nyomtatási minősége aligha változott olyan mértékben, ami érdemben befolyásolná a karakterfelismerő programok pontosságát, s ezáltal torzítaná a megfigyelt tendenciát. Ráadásul ez a tendencia egybevágh azzal, ami a kifejezeten a tudományos publikációkra JSTOR adatbázisból kirajzolódik (Dunning, 2012: 2).

Az amerikai kormány 1965-től mintegy másfél évtizeden át anyagilag támogatta azoknak a 18 és 22 év közötti fiataloknak az egyetemi tanulmányait, akik elveszítették társadalombiztosítási ellátásban részesülő szülőjüket. A program hamar népszerűvé vált, és 1977-ben már csaknem 900 000 diák számára nyújtott segítséget. 1981-ben azonban a Reagan-kormány egyetlen tollvonással megszüntette ezt a támogatást. Ez a hirtelen változás jó alkalmat kínált az anyagi támogatás hatásának a vizsgálatára, méghozzá egy természetes kísérlet keretében. Az anyagi ellátás megszüntetése ugyanis két, egymástól csupán a támogatás meglétében, illetve hiányában különböző, s minden más szempontból nagyjából egyforma csoportra osztotta az 1981 körüli években ellátásra jogosulttá váló fiatalokat: azok, akik közvetlenül a változás előtt lettek jogosultak az ellátásra, még megkapták a támogatást – ők alkották a kísérleti csoportot; azok pedig, akik közvetlenül a változás után lettek jogosultak az ellátásra, már nem kapták meg a támogatást – ők alkották a kontrollcsoportot.

Snow és Semmelweis vizsgálatához hasonlóan ezúttal is a kutatótól független, külső tényezők révén jött létre egy olyan helyzet, amely gyakorlatilag megfelelt a valódi kísérlet követelményeinek. Egyrészt a Reagan-kormány döntése „hozta létre” a magyarázó változó két értékét (részeseül anyagi támogatásban – nem részeseül anyagi támogatásban) – ugyanúgy, ahogyan a londoni példában az egyik vízszolgáltató cég döntése „hozta létre” ezt a két értéket (szennyezett ivóvizet kap – tiszta ivóvizet kap); Semmelweis esetében pedig az egyetem döntése (1840-től az egyik klinikán csak orvostanhallgatókat, a másikon csak bábánövendékeket oktatnak). Másrészt a magyarázó változónak ehhez a két értékéhez gyakorlatilag véletlenszerűen kerültek a megfigyelések (egyetemi hallgatók) – megint csak ugyanúgy, mint a korábbi történeti példákban. Úgy, ahogyan a szülő nők különféle sajátosságai (életkoruk, társadalmi státuszuk stb.) aligha korreláltak azzal, hogy konkrétan melyik napon indult meg a szülés, és ennek következtében a két klinika melyikére kerültek – az egyetemisták, illetve családjuk különféle sajátosságai is aligha korreláltak azzal, hogy konkrétan mikor veszítették el szüleiket, s ennek révén konkrétan mikor váltak jogosulttá a támogatásra – akkor, amikor még kaphattak segílyt, vagy akkor, amikor már nem.

Hogyan befolyásolta az anyagi támogatás megszüntetése a fiatalok továbbtanulását? Erre a kérdésre keresett választ az imént említett kutatás, amely az 1979 és 1981 közötti években, tehát közvetlenül a megszüntetés előtt jogosulttá vált, s így még támogatásban részesültek egy csoportját (137 fő) hasonlította össze azzal a másik csoporttal, amelybe 54 olyan fiatal került, akik 1982-1983-ban, tehát közvetlenül a megszüntetés után váltak jogosulttá, s így már nem kaptak támogatást. A 4. táblázat azok arányát mutatja a két csoportban, akik 23 éves korukig bezárólag megkezdték felsőfokú tanulmányaikat (college). Látható, hogy a kísérleti csoportban – vagyis a még éppen támogatásban részesülők körében – jóval nagyobb (56%) a továbbtanulási arány, mint a kontrollcsoportban, vagyis azok körében, akik már éppen kimaradtak a támogatásból (35%).

4. táblázat. Szülőjüket elvesztő fiatalok anyagi támogatásának hatása a továbbtanulásra

Csoport	Évjárat	n	Továbbtanulási arány, %
Kísérleti (még éppen kap támogatást)	1979-1981	137	56
Kontroll (már éppen nem kap támogatást)	1982-1983	54	35

Forrás: Murnane and Willett, 2011: 143 alapján

Ha ezt a példát megnézzük kicsit közelebbről, és mindenekelőtt annak logikai szerkezetére fordítjuk figyelmünket, akkor láthatjuk, hogy a *megfigyelési*

egységek egy olyan kontinuumon – az idő tengelyen – helyezkednek el, amelyet mintegy kettévág a kormányzati beavatkozás, nevezetesen a támogatás megszüntetése. Azok, akik ettől a vágásponttól közvetlenül balra – az időben visszafelé – helyezkednek el, alkotják a kísérleti csoportot, azok pedig, akik attól közvetlenül jobbra – az időben előre – vannak, alkotják a kontrollcsoportot. Erre a logikára épült az a kutatás is, amely a szülők által igénybe vehető gyermekgondozási szabadság meghosszabbításának a további gyermekvállalásra gyakorolt hatását vizsgálta Ausztriában (Lalive and Zweimüller, 2005). Itt a szabadság hosszát egy évről két évre növelő, 1990. július 1-től érvényes jogszabály-módosítás vágta ketté az idő tengelyét, létrehozva egyrészt a kisbabájukat közvetlenül az új szabályozás bevezetése előtt világra hozó – és ezért annak előnyeit még éppen nem élvező – anyák alkotta kontrollcsoportot, másrészt pedig a gyermeküket közvetlenül az új jogszabály életbe lépése után megszüülő anyák alkotta kísérleti csoportot, akikre már épp az új, kedvezőbb feltételek vonatkoztak. A vizsgálat eredményei szerint az új törvény mind rövid-, mind hosszabb távon növelte a további gyermekek megszületésének valószínűségét, valamint kedvezően befolyásolta a szülők munkába való visszatérését is.

Az a kontinuum, amelyen a megfigyelési egységek elhelyezkednek, s amelyet a beavatkozás két részre bont, nem csupán időbeli lehet. Egy amerikai kutatás (Black, 1999) a helyi iskola minőségének a lakásárakra gyakorolt hatását vizsgálta olyan lakások árainak az összehasonlítása révén, amelyek egymás közvetlen közelében, de pont egy iskolakörzet-határ két ellentétes oldalán álló házakban voltak. Ezek a lakások – épp közelségük miatt – gyakorlatilag minden szempontból (környezetüket, fizikai jellemzőiket, az ott élők társadalmi összetételét stb. tekintve) hasonlóak voltak – egyetlen dolgot kivéve, éspedig annak az iskolának a minőségét, ahova a lakók gyerekei tartoztak. Ezért a közöttük az árak tekintetében megfigyelt különbséget viszonylag nyugodt szívvel lehetett az oktatási színvonal számlájára írni. Szemben az előző két példával, ahol a kontinuum, amelyen a megfigyelési egységek (a valamelyik szülőjüket elvesztő fiatalok, illetve a gyermeküket szülő édesanyák) elhelyezkedtek, időbeli volt, itt a kontinuum térbeli – ezt a térbeli kontinuumot vágta ketté az iskolai körzethatár. Azok az otthonok, amelyek ettől a vágásponttól az egyik irányba – éspedig a jobb minőségű iskola irányába – estek, alkották a kísérleti csoportot, azok pedig, amelyek a másik irányban – éspedig a gyengébb minőségű iskola irányában – helyezkedtek el, alkották a kontrollcsoportot. Az eredmények szerint a vevők átlagosan 2,5%-kal többet fizettek azokért a lakásokért, amelyek a jobb minőségű oktatást nyújtó iskola körzetéhez tartoztak.⁶

⁶ Hasonló eredményre jutott egy későbbi francia kutatás (Frack and Grenet, 2010), amely ugyanakkor azt is kimutatta, hogy az iskola minőségének a lakásárakra gyakorolt hatása függ a környékbeli magániskolák hozzáférhetőségétől.

TÉNYLEGES RANDOMIZÁLÁSON ALAPULÓ TERMÉSZETES KÍSÉRLETEK

A természetes kísérlet eddig tárgyalt példáiban a kutatók olyan helyzeteket, csoportokat hasonlítottak össze egymással, amelyek a körülmények szerencsés alakulásának köszönhetően *gyakorlatilag randomizáltak* voltak tekinthetők – de szigorúan véve mégsem tényleges véletlen hozzárendelés eredményeként jöttek létre.⁷ John Snow vizsgálatában a két vízszolgáltató cég ügyfelei között nem volt semmilyen szisztematikus különbség – de az egyes fogyasztók mégsem valódi randomizálással kerültek egyik vagy másik társasághoz. Ugyanígy, Semmelweis esetében a két klinika betegei között sem volt semmilyen szisztematikus különbség – de a szülő nők mégsem valódi randomizálással kerültek egyik vagy másik klinikára.

Előfordul azonban, hogy a kutató olyan helyzetekre bukkan, ahol az összehasonlítható csoportok *ténylegesen véletlen hozzárendeléssel* – rendszerint valamilyen sorsolással – jöttek létre. Az ilyen helyzetek – amennyiben a kutatók észreveszik és kihasználják őket – különösen alkalmasak olyan oksági tényezők szerepének a vizsgálatára, amelyek egyébként rendszerint *nem véletlen szelekció* eredményei, s ezért hatásuk más módszerekkel nehezen vizsgálható.

Képzeljünk el egy olyan kutatást, amelynek középpontjában a migráció egészségre gyakorolt hatása áll. Az, hogy ki vállalkozik hazája elhagyására és ezzel egy új országba költözésre, általában nem véletlenszerű, hanem egy sor olyan tényezőtől függ (például a nemtől, az életkortól, az anyagi helyzettől), amelyek egyszersmind az egészségi állapotot – azaz a függő változót – is befolyásolják. Így aztán könnyen lehet, hogy ami első pillantásra a migráció egészségre gyakorolt hatásának tűnik, az pusztán a migrációhoz vezető nem véletlen szelekció következménye.

Hogyan lehetne ezt a nem véletlen szelekciót kiküszöbölni, azaz hogyan lehetne a migráció és az azt befolyásoló számtalan tényező (nem, életkor, anyagi helyzet stb.) közötti szisztematikus kapcsolatot megszüntetni? Az embereket aligha lehet a tudomány érdekében akaratuk ellenére költözésre kényszeríteni, vagy a költözés lehetőségétől megfosztani, következésképpen a kutató által végrehajtott tényleges randomizáláson alapuló valódi kísérletre nincs mód. Itt lép be a képbe a természetes kísérlet. Előfordul, hogy a költözni kívánók jóval

⁷ Az ilyen esetekre használják az angol nyelvű szakirodalomban az „*as if*” *randomization* kifejezést (Dunning, 2008), arra utalva ezzel, hogy az összehasonlítható csoportok olyanok, „mintha” randomizáltak lennének – igazából azonban mégse teljesen azok.

többen vannak, mint amennyit a célpontként választott ország befogadni képes; ilyenkor gyakran valamilyen lottószerű sorsolással döntik el, hogy a bevándorlásra jelentkezők közül ki kapja meg a beköltözés esélyét, és ki az, aki elesik ettől a lehetőségtől. Ez a sorsolás a jelentkezőket véletlenszerűen osztja két csoportra, következésképpen a két csoport tagjai sem nem, sem életkor, sem anyagi helyzet, sem semmilyen más tényező szempontjából nem különböznek szisztematikusan egymástól. Ezért aztán a sorsolásból vesztesként kikerülőket megalapozottan tekinthetjük úgy, mint akik a rendszerint „counterfactual”-ként emlegetett „mi lett volna, ha nem történt volna semmi” (a konkrét példában: „mi lett volna, ha nem került volna sor a migrációra”) hipotetikus helyzetét képviselik, a sorsoláson nyertesek és vesztesek egészségi állapota között tapasztalt különbséget pedig viszonylag nyugodt szívvel írhatjuk a migráció számlájára.⁸

Egy ilyen bevándorlási lottó kínálta lehetőséget használt ki az a kutatás, amely az óceániai Tonga-szigetokről Új-Zélandra költözők egészségi állapotát vizsgálta (Gibson et al., 2013). Mivel a jelentkezők száma jelentősen meghaladta a rendelkezésre álló helyekét, az új-zélandi kormány lottósorsolással döntötte el, hogy a tongaiak közül kik élhetnek ténylegesen a bevándorlás lehetőségével. Ez a sorsolás véletlenszerűen osztotta két részre – nyertesekre és vesztesekre – az Új-Zélandra költözni kívánók csoportját. Ezt a két részt összehasonlítva a kutatók megállapították, hogy a nyerteseknek átlagosan magasabb volt a vérnyomása, nagyobb arányban voltak vérnyomásproblémáik, ezt a különbséget pedig úgy értelmezték, mint a migrációnak az egészségi állapotra gyakorolt tényleges ok-sági hatását.⁹

A migráció hatásának vizsgálatához hasonló nehézségbe ütköznek azok a kutatók, akik az anyagi helyzet egészségre gyakorolt hatását igyekeznek megállapítani. Miként a bevándorlóvá válás, úgy a gazdaggá vagy szegénnyé válás sem véletlenszerű, hanem ez is egy szelekciós folyamat eredménye. Az, hogy ki lesz gazdag, és ki szegény, egy sor olyan tényezőtől függ (nem, életkor, lakóhely stb.), amelyek egyszersmind az egészségi állapotot is befolyásolják. Ezért ami első pillantásra az anyagi helyzet hatásának tűnik, az könnyen lehet, hogy pusztán a gazdagsághoz, illetve szegénységhez vezető nem véletlen szelekció következménye.

⁸ A dolog annyiban bonyolultabb, hogy a nyertesek közül nem mindenki él ténylegesen a költözés lehetőségével, és az, hogy melyik *nyertesből* lesz *tényleges* költöző, már *nem teljesen véletlenszerű*; itt már érvényesül a *szelekció*. Erre a problémára hamarosan visszatérünk egy másik példa kapcsán.

⁹ Az előző lábjegyzetben említett szelekciós problémát a kutatók statisztikai módszerekkel igyekeztek kiküszöbölni – ezekről a módszerekről nemsokára részletesebben is lesz szó.

Hogyan lehetne ezt a nem véletlen szelekciót kiküszöbölni? A migráció esetéhez hasonlóan a kutató által végrehajtott tényleges randomizálás ezúttal is aligha jöhet szóba – a természetes kísérlet azonban itt is megoldást jelenthet. Egy svédországi vizsgálatban (Lindah, 2005) például a lottón nyert pénzt tekintették olyan jövedelemnek, aminek a megszerzése gyakorlatilag véletlenszerűnek tekinthető, s ami ezért aligha korrelál az egyének olyan tulajdonságaival (nemével, életkorával, lakóhelyével stb.), amelyek az egészségüket is befolyásolják. Az eredmények szerint a lottón nyert pénz javítja az egészségi állapotot, s ezt a hatást – a vizsgálat felépítésének, természetes kísérlet jellegének köszönhetően – a kutatók viszonylag nyugodt szívvel tekinthették valóban oksági jellegűnek.

A legismertebb – és a kutatók számára később igazi „aranybányának” bizonyuló – lottósorsolást a vietnami háború idején tartották az Egyesült Államokban annak eldöntésére, hogy a hadköteles korú férfiak közül kiket sorozzanak be katonának. A sorsolás keretében születési dátumokat húztak ki véletlenszerűen egymás után; minél hamarabb húzták ki valakinek a születési dátumát, annál nagyobb volt a behívás kockázata. A hadsereg munkaerő-szükséglete alapján a kormány meghatározta a kihúzási sorszám „plafonértékét” – akinek a születésnapjához ennél alacsonyabb sorszám tartozott, azt ténylegesen is besorozták, míg akinek a születési dátumát ennél később – magasabb sorszámmal – húzták ki, az mentesült a behívás alól.

Bár a sorozási lottó eredeti célja az volt, hogy a katonai behívás teljesen pártatlan, elfogulatlan legyen, ne a protekción vagy a jobb anyagi körülményeken, hanem kizárólag a vakszerencsén múljon, kinek kell végül Vietnámban harcolnia,¹⁰ ezen túlmenően egy természetes kísérlet lehetőségét is kínálta. A születésnapok véletlenszerű kiválasztása ugyanis úgy hozta létre a katonai szolgálatra kötelezettek és a szolgálat alól mentesülők két csoportját, hogy azok sem iskolázottság, sem anyagi helyzet, sem etnikai hovatartozás, sem semmilyen más hasonló változó szempontjából nem különböztek szisztematikusan – a véletlen határain túlmenően – egymástól. Ez komoly módszertani előnyt jelentett a hagyományos megfigyeléses vizsgálatokhoz képest, ahol az összehasonlítható csoportok rendszerint nem véletlen szelekció eredményei, következésképpen összetételük

¹⁰ Paradox módon a legelső, 1969 decemberében rendezett sorozási lottó épp ennek a kiválasztásnak nem tett maradéktalanul eleget. Amint azt statisztikusok (Fienberg, 1971) kimutatták, az év egyes napjain születetteknek nem volt teljesen egyforma esélyük a besorozottnak, illetve a szolgálat alól mentesülők közé kerülésre. A dátumokat tartalmazó golyókat ugyanis egymás után tették az urnába, így az év későbbi napjait tartalmazó golyók nagyobb eséllyel maradtak az urna felszínén, és ezért hamarabb is húzták ki őket, vagyis az év későbbi hónapjaiban születettek esetében nagyobb volt a besorozás kockázata. Azonban a *születési dátum aligha korrelált a sorozási lottót később természetes kísérlet céljára felhasználó kutatók által vizsgált különféle jelenségekkel*, következésképpen ez a probléma aligha torzította az ezekben a kutatásokban kapott eredményeket.

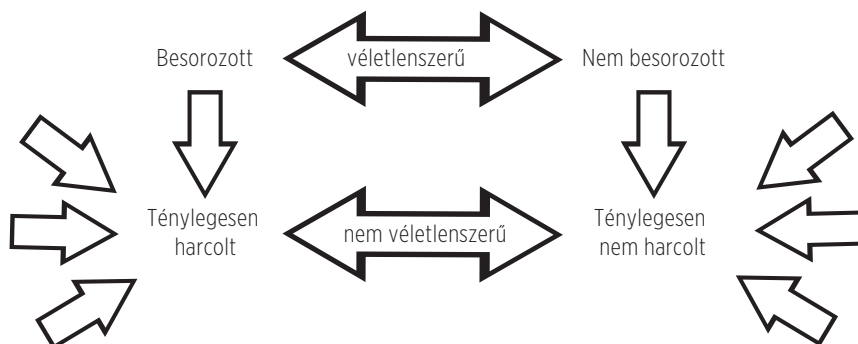
általában már eleve eltérő (amit aztán a kutatók kontrollváltozók bevonásával igyekeznek – több-kevesebb sikerrel – ellensúlyozni).

A sorozási lottóban rejlő lehetőséget számosan igyekeztek kiaknázni. Az egyik legkorábbi ilyen vizsgálat (Hearst, Newman, Hulley, 1986) a katonai szolgálatnak a későbbi halálozásra gyakorolt hatásával foglalkozott. Az eredmények szerint a születési dátumuk alapján besorozott férfiak halandósága jelentősen felülmúlta azokét, akik mentesültek a behívás alól. Különösen nagy volt a különbség az erőszakos halálozások tekintetében: az öngyilkosság kockázata például 13%-kal, a halálos kimenetelű autóbaleseteké pedig 8%-kal volt nagyobb a besorozottak körében, mint a behívást „megúszók” csoportjában. Egy másik kutatás (Angrist, 1990) a háborús veteránok későbbi anyagi helyzetét hasonlította össze a katonai szolgálat alól mentesülőkével, és azt találta, hogy a fehér bőrű férfiak körében a veteránok jövedelme még tíz évvel a harcok lezárulta után is jelentősen alulmúlta azokét, akik kimaradtak a háborúból, vélhetően azért, mert a harctéren szerzett tapasztalataik nem ellensúlyozták azt a veszteséget, amit a munkaerőpiacról való távollétük miatt elszenvedtek. Végül Erikson és Stoker (2011) vizsgálatának középpontjában a besorozás veszélyének a politikai nézetekre gyakorolt befolyása állt. Míg Hearst és munkatársai, illetve Angrist a tényleges katonai szolgálat hatására volt kíváncsi, Erikson és Stokert a lehetséges harctéri bevetés kockázatának a szerepe érdekelte. Eredményeik szerint azok, akik a születésnapjuk kihúzási sorszáma alapján a besorozottak – és ezáltal esetlegesen a háborúban harcolók – közé kerültek, a későbbiekben erősebben háborúellenesek voltak, inkább képviseltek liberális nézeteket, és nagyobb arányban szavaztak a demokrátaakra a választásokon, mint azok, akik megúszták a besorozást.

Ezekkel a kutatásokkal kapcsolatban fontos megjegyezni, hogy a sorozási lottó véletlenszerű jellegéből adódó módszertani előnyök *csak a besorozottak és a sorozás alól mentesülők csoportjainak az összehasonlítására* vonatkoznak; csak ennek a két csoportnak az esetében igaz az, hogy az egyének véletlenszerűen kerültek egyikbe, illetve másikba, következésképpen csak ezt a két csoportot tudjuk úgy szembeállítani egymással, hogy a sorozáson kívül semmilyen más szisztematikus különbség ne legyen közöttük. A sorozás azonban nem feltétlen jelent *tényleges* háborús bevetést – ugyanúgy, ahogyan a sorozás alól való mentesülés sem okvetlenül jelenti a háborús bevetés *tényleges* elmaradását. A besorozottak egy része valóban harcolt, más része azonban nem, és ugyanígy a nem besorozottak egy része valóban kimaradt a háborúból, más része azonban nem (ők önkéntesnek jelentkeztek). Az pedig, hogy egy besorozott (vagy épp nem besorozott) férfiből hogyan lesz valóban harcoló (vagy épp nem harcoló) katona, *már nem véletlenszerű* – itt már ugyanolyan *szelekció* érvényesül, mint egy hagyományos kvázi-kísérletnél, ahol vagy ma-

guk az egyének döntik el, hogy részt vesznek-e valamilyen programban, vagy pedig a kutató már létező – és ezért egy sor szempontból már eleve eltérő – természetes csoportokat (például iskolákat vagy osztályokat) hasonlít össze egymással. A magasabb iskolázottságú vagy kedvezőbb anyagi helyzetű, esetleg jobban kamatoztatható társadalmi kapcsolatokkal rendelkező férfiak például másokhoz képest alighanem könnyebben tudtak kibújni a tényleges harctéri szolgálat alól. Ennek megfelelően a ténylegesen harcolók és a ténylegesen nem harcolók összehasonlítására a sorozási lottó véletlen jellegéből adódó módszertani előnyök *már nem* vonatkoznak (lásd 2. ábra, ahol az oldalsó nyilak a tényleges háborús bevetést vagy annak elkerülését a besorozáson túlmenően még befolyásoló olyan egyéb oksági tényezőket jelölik, mint amilyen az iskolázottság vagy a jövedelmi helyzet).

2. ábra: A nem véletlen szelekció problémája



Mit lehet tenni ebben a helyzetben? Az egyik lehetőség, hogy maradunk a besorozottak és a sorozás alól mentesülők összehasonlításánál, nem törődve azzal, hogy a két csoport tagjai közül kik harcoltak ténylegesen, és kik nem. Vagyis az összehasonlítást *arra korlátozzuk, amire a véletlenszerű hozzárendelés valóban kiterjed*. Ezt a stratégiát hívják *intention-to-treat* elemzésnek (Hollis and Campbell, 1999), arra utalva ezzel az elnevezéssel, hogy ebben az esetben azokat hasonlítjuk össze, akiket a beavatkozás, illetve annak elmaradása *szándékunk szerint* érint (és akik esetében a randomizálás valóban érvényesül), függetlenül attól, hogy ez a szándék megvalósul-e vagy sem. Ez a megoldás bizonyos esetekben teljesen megfelelő lehet; Erikson és Stoker imént említett kutatása (Erikson and Stoker, 2011) például eleve nem a tényleges háborús részvétel, hanem csupán a besorozás általi – és a születésnap kihúzási sorszámától függő – fenyegetettség hatására irányult. Más esetekben azonban az *intention-to-treat* elemzés azzal a hátránnyal járhat, hogy nem a bennünket igazából érdeklő kér-

désre kapunk választ: a sorozás hatását tudjuk csak megállapítani, holott mi valójában a tényleges háborús részvétel következményeire lennénk kíváncsiak.

Ilyenkor jelenthet megoldást az a módszer, amelynek a természetes kísérletekben történő alkalmazása elsősorban Joshua Angrist nevéhez fűződik, és amely az ún. *instrumentális változó* használatára épül. Ez a módszer két lépésből áll: először a randomizálással létrejött eredeti besorolás (a példában: besorozták – nem sorozták be) alapján megbecsüljük a későbbi, már nem randomizálással kialakult besorolást (harcolt – nem harcolt), utána pedig az így kapott becslést használjuk magyarázó változóként az eredeti magyarázó változó helyett (például a sorozás alapján becsült háborús részvételt használjuk a tényleges részvétel helyett). Ennek a kétlépcsős eljárásnak az értelmét az adja, hogy mivel a sorozás (szemben a tényleges katonai szolgálattal) véletlenszerű volt, az *független minden olyan egyéb, általunk figyelembe nem vett oksági tényezőtől*, amely a vizsgált függő változót (például a halandóságot, a jövedelmeket vagy a politikai nézeteket) befolyásolja. Következésképpen ha mi a sorozás alapján megbecsüljük a tényleges katonai szolgálatot, akkor az így kapott becsült érték – lévén az a sorozás tökéletes függvénye – *szintén független lesz* minden olyan egyéb oksági tényezőtől, amely a vizsgált függő változót befolyásolja – ellentétben a katonai szolgálat tényleges értékével, ami már nem véletlen hozzárendelés eredménye, s ezért jó eséllyel korrelál ezekkel a tényezőkkel. Ha tehát a katonai szolgálat *tényleges* értéke *helyett* annak a sorozás alapján *becsült* értékét használjuk magyarázó változóként, akkor ez az új magyarázó változó már nem fog korrelálni a függő változót befolyásoló egyéb, a modellből kihagyott és a hibatényezőbe „számozított” oksági tényezőkkel, hiszen tökéletes függvénye a sorozásnak, ami pedig véletlenszerű.

TERMÉSZETES KÍSÉRLET ÉS MÁS KUTATÁSTÍPUSOK

A természetes kísérletet korábban olyan kutatástípusként jellemeztük, amelyben egyrészt nem maga a kutató, hanem tőle és a kutatás céljától is független, külső erők hozzák létre a magyarázó változó egyes értékeit vagy kategóriáit, másrészt pedig a megfigyelések gyakorlatilag – vagy ténylegesen is – véletlenszerűen kerülnek ezekbe a kategóriákba. E két sajátossága alapján a természetes kísérlet jól elhatárolható az empirikus kutatás olyan más fajtáitól, amelyeket sokszor összekevernek vele (Ditlmann and Paluck, 2015: 129).

Ezek egyike a *terepkísérlet* (Baldassarri and Abascal, 2017). Míg a természetes kísérletben a kísérleti helyzet a kutató tényleges részvétele nélkül jön létre, és az ő szerepe alapvetően ennek a tőle függetlenül kialakult helyzetnek a

találékony kihasználásában áll, addig a terepkísérletben a kutató *aktívan részt vesz magának a kísérleti helyzetnek a kialakításában*: ő hozza létre a magyarázó változó egyes értékeit (például elindít valamilyen társadalmi programot, megteremtve ezzel a programban résztvevők és nem résztvevők kategóriáját), a megfigyelési egységeket (egyéneket vagy csoportokat) pedig szintén ő rendeli hozzá véletlenszerűen ezekhez a kategóriákhoz.

Egy másik kutatástípus, amit gyakran természetes kísérletként jellemeznek, noha szigorúan véve nem az, a *természetes kvázi-kísérlet*. Míg a természetes kísérlet és a terepkísérlet között a különbség abban állt, mennyire játszik a kutató közvetlen, aktív szerepet a kísérleti helyzet létrehozásában, itt az eltérés abból fakad, mennyire tekinthetők a magyarázó változó egyes kategóriáihoz tartozó megfigyelések úgy, mint amelyek valamilyen, a randomizálásnak megfelelő, azzal egyenértékű véletlenszerű folyamat eredményeként kerültek egyik vagy másik csoportba. Ellentétben a természetes kísérlettel, ahol a körülmények ismeretében általában viszonylag hihetően érvelhetünk amellett, hogy a véletlenszerűség követelménye valóban teljesült, a természetes kvázi-kísérletnél az összehasonlítható csoportok rendszerint még csak megközelítően sem mondhatók randomizáltak. Ez volt a helyzet például abban az egyébként nagy fokú kutatói leleményességről tanúskodó kanadai vizsgálatban, amelynek középpontjában a televíziózásnak – egyebek között – a gyerekek agresszivitására gyakorolt hatása állt (MacBeth, 2001). Egy völgyben fekvő, hegyekkel körülvett kisvárosban – épp a kedvezőtlen domborzati viszonyok miatt – egészen az 1970-es évekig nem lehetett fogni a TV-műsorokat. Aztán 1973-ban létesítettek egy átjátszóállomást, ami már lehetővé tette a TV-adások vételét. Ez az átjátszóállomás természetesen nem azért épült, hogy a televíziózásnak az emberek magatartására – ezen belül a gyerekek agresszivitására – gyakorolt hatását vizsgálják, azonban, mintegy mellékesen, lehetőséget kínált erre is. A kutatók a játszótéri veszekedések, verekedések gyakoriságával mérték az agresszivitást, és két időpontban nézték annak erősségét: a TV-állomás létesítése előtt, illetve az után. Az így megfigyelt változást pedig összehasonlították azzal a változással, amit két olyan településen tapasztaltak, ahol már régóta fogni lehetett a TV-műsorokat. Bár ezeket a kontrollként használt településeket igyekeztek úgy kiválasztani, hogy minél hasonlóbbak legyenek a kísérleti településhez, ennek ellenére minden bizonnyal voltak olyan szisztematikus különbségek a kísérleti és a kontrolltelepülések között, amelyek hatása összekeveredhetett a televíziózás megjelenésének hatásával. Ennek megfelelően ezt a kutatást – az arról beszámoló tanulmány címével ellentétben – indokolt természetes kísérlet helyett inkább természetes kvázi-kísérletnek nevezni.

ZÁRÓ GONDOLATOK

A cikk legelső mondatában annak fontosságát hangsúlyoztuk, hogy kutatásunk belső érvényessége minél magasabb legyen, vagyis minél biztosabbak lehessünk abban, hogy az adatainkban megfigyelt összefüggést valóban az általunk feltételezett oksági mechanizmus hozta létre – nem pedig valamilyen egészen más hatásmechanizmus. Láttuk, hogy ennek a célnak az eléréséhez az alternatív magyarázatok kiküszöbölésén keresztül vezet az út, ez utóbbi legjobb eszköze pedig a randomizálás, azaz a megfigyelési egységeknek a magyarázó változó egyes kategóriáihoz történő véletlenszerű hozzárendelése. Ezért tekintik az oksági következtetések megalapozottsága szempontjából a randomizáláson alapuló valódi kísérletet általában a módszerek „királynőjének”.

Ez azonban az éremnek csupán az egyik oldala. Amennyire lényeges, hogy következtetéseink okságilag minél biztosabb lábakon álljanak, annyira lényeges az is, hogy ezek a következtetések minél inkább *általánosíthatók* legyenek, vagyis, hogy minél magasabb legyen a kutatásunk *külső* érvényessége. Ez utóbbi tekintetben pedig a valódi kísérlet – különösen ha azt laboratóriumi keretek között végzik – sokszor nem teljesít különösebben fényesen. Egyrészt a vizsgálatba bevont egyének gyakran nem reprezentálják megfelelően azt a tágabb sokaságot, amelyre megállapításainkat általánosítani szeretnénk, másrészt a kísérleti körülmények sokszor igencsak eltérnek attól, amit a mindennapi életben tapasztalunk, következésképpen kétséges, hogy amit a kísérlet során találunk, megfigyelhető-e természetes körülmények között is. Nem ritkán maga a vizsgálat ténye is befolyásolja az abban résztvevők viselkedését, amint azt a Western Electric távközlési cég hawthorne-i üzemében végzett kutatás már az 1900-as évek első harmadában kimutatta, és ami ennek nyomán „Hawthorne-hatás”-ként vonult be a módszertani szakirodalomba.¹¹

A természetes kísérlet egyik előnye, hogy segíthet megteremteni az *egyen-súlyt a belső és a külső érvényesség*, az eredmények oksági megalapozottsága és általánosíthatósága között. Ebben rokon egy másik, korábban már említett kutatástípussal, a terepkísérlettel, azonban bizonyos fokig túl is megy azon an-

¹¹ A Hawthorne-hatás (és a vele sok szempontból rokon placebohatás) tulajdonképpen speciális esete annak az általánosabb problémának, amit úgy fogalmazhatunk meg, hogy *a kutatás – beavatkozás*. Beavatkozás, ami megváltoztathatja magát a vizsgált jelenséget. Bár ezzel a problémával a természeti jelenségekkel foglalkozó tudományokban is találkozhatunk, igazán súlyossá az emberekkel foglalkozó tudományokban válik. Az emberek ugyanis tudatosan reagálnak arra, hogy vizsgálják őket, és ez a reakció összekeveredhet a konkrétan vizsgált oksági tényező hatásával. Nem véletlenül tekintette a kiváló gazdaságtörténész, Alexander Gerschenkron a természet- és a társadalomtudományok közötti egyik legfőbb különbségnek azt, hogy – mint fogalmaz – „a társadalomtudós és vizsgálatának tárgya kölcsönösen befolyásolhatják egymást”. (Gerschenkron, 1984: 159–160)

nak köszönhetően, hogy lehetőséget nyújt a *beavatkozásmentes* vizsgálatra, csökkentve ezzel annak veszélyét, hogy maga a kutatás ténye befolyásolja az abban résztvevők magatartását. A beavatkozásmentes kutatás egyéb formái közül ugyanakkor – mint amilyen például a tartalomelemzés vagy a korábban mások által, más célból gyűjtött adatok újbóli feldolgozásán alapuló másodelemzés – azáltal tűnik ki, hogy az alternatív magyarázatoknak rendszerint szélesebb körét teszi kiküszöbölhetővé, s ily módon a kapott eredmények és az azok nyomán megfogalmazott következtetések okságilag általában megalapozottabbak.

IRODALOM

- Angrist, J. D. (1990). Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *American Economic Review*, 80(3), 313–336. <https://www.jstor.org/stable/2006669>
- Armitage, P. (2003). Fisher, Bradford Hill, and randomization. *International Journal of Epidemiology*, 32, 925–928. doi: 10.1093/ije/dyg286
- Baldassarri, D. and Abascal, M. (2017). Field Experiments Across the Social Sciences. *Annual Review of Sociology*, 43, 41–73. <https://doi.org/10.1146/annurev-soc-073014-112445>
- Black, Sandra E. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics*, 114(2), 577–599. <https://doi.org/10.1162/003355399556070>
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297–312. <https://doi.org/10.1037/h0040950>
- Cook, T. D., and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Czeizel, E. (2006). *Tudósok – gének – tanulságok*. Budapest: Galenus Kiadó.
- Ditlmann, R. and Paluck, E. L. (2015). Field Experiments. In: James D. Wright, eds., *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, Vol 9. Oxford: Elsevier. pp. 128–134. ISBN: 9780080970868
- Dunning, T. (2012). *Natural Experiments in the Social Sciences*. Cambridge, etc.: Cambridge University Press. ISBN-10: 1107698006, ISBN-13: 978-1107698000.
- Dunning, T. (2008). Improving Causal Inference: Strengths and Limitations of Natural Experiments. *Political Research Quarterly*, 61(2), 282–293. doi: 10.1177/1065912907306470
- Dynarski, S. (2003). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review*, 93(1), 279–288.
- Fack, G. and Grene, J. (2010). When do better schools raise housing prices? Evidence from Paris public and private schools. *Journal of Public Economics*, 94, 59–77. doi:10.1016/j.jpubeco.2009.10.009
- Fienberg, S. E. (1971). Randomization and Social Affairs: The 1970 Draft Lottery. *Science*, 171: 255–61. doi: 10.1126/science.171.3968.255
- Freedman, D. A. (2005). *Statistical models. Theory and practice*. Cambridge etc.: Cambridge University Press. ISBN-10 : 0521743850, ISBN-13 : 978-0521743853
- Freedman, D. A. (2005). From Association to Causation: Some Remarks on the History of Statistics. *Statistical Science*, 14(3), 243–258.
- Gerschenkron, A. (1984). *A gazdasági elmaradottság történelmi távlatból*. Budapest: Gondolat Kiadó.
- Gibson, J., Stillman, S., McKenzie, D. and Rohorua, H. (2013). Natural experiment evidence on the effect of migration on blood pressure and hypertension. *Health Economics*, 22, 655–672. <https://doi.org/10.1002/hec.2834>
- Hall, Nancy S. (2007). R. A. Fisher and his advocacy of randomization. *Journal of the History of Biology*, 40, 295–325. doi: 10.1007/s10739-006-9119-z
- Hearst, N., Newman, T. B. and Hulley, S. B. (1986). Delayed Effects of the Military Draft on Mortality. *New England Journal of Medicine*, 314(10), 620–624.
- Hollis, S. and Campbell, F. (2009). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 319 (7211), 670–674. doi: 10.1136/bmj.319.7211.670

- Lalive, R. and Zweimüller, J. (2005). *Does Parental Leave Affect Fertility and Return-to-Work? Evidence from a „True Natural Experiment“*. IZA Discussion Paper No. 1613. <https://dx.doi.org/10.2139/ssrn.731766>
- Lau, F. and Kuziemsky, C. (Eds.) (1972). *Handbook of eHealth Evaluation: An Evidence-based Approach*. Victoria: University of Victoria.
- Lindahl, M. (2005) Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income. *Journal of Human Resources*, 40(1), 144–168.
- MacBeth, T. M. (2001). The impact of television: A Canadian natural experiment. In C. McKie & B.D. Singer (Eds.) *Communications in Canadian Society*, 5th ed. (pp. 196–213). Toronto: Thompson Educational Publishing.
- Michel, J-B. et al. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176–182. doi: 10.1126/science.1199644.
- Moksony, F. (2018). *Gondolatok és adatok. Társadalomtudományi elméletek empirikus ellenőrzése*. Budapest: Budapesti Corvinus Egyetem.
- Morabia, A. (2004). *A history of epidemiologic methods and concepts*. Basel: Birkhäuser Verlag.
- Murnane, R.J. and Willett, J.B. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Oxford, etc.: Oxford University Press.
- Pasteur, L. ([1854] 1939). Discours prononcé à Douai, le 7 décembre 1854, à l'occasion de l'installation solennelle de la Faculté des lettres de Douai et de la Faculté des sciences de Lille. Reprinted in: Pasteur Vallery-Radot, ed., *Oeuvres de Pasteur*. Paris, France: Masson and Co., vol. 7.
- Pechenick E. A., Danforth C. M. and Dodds P. S. (2015) Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution. *PLoS ONE* 10(10), e0137041. doi: 10.1371/journal.pone.0137041
- Pettit, M. (2016). Historical time in the age of big data – Cultural psychology, historical change, and the Google Books Ngram Viewer. *History of Psychology*, 19(2), 141–153.
- Robinson, G., McNulty, J. E. and Krasno, J. S. (2009). Observing the Counterfactual? The Search for Political Experiments in Nature. *Political Analysis*, 17, 341–357. doi:10.1093/pan/mpp011
- Shadish, W. R, Cook, Th. D. and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Snow, J. (1855). *On the Mode of Communication of Cholera*. 2nd ed. London.
- Younes N., Reips U-D. (2019) Guideline for improving the reliability of Google Ngram studies: Evidence from religious terms. *PLoS ONE* 14(3), e0213554. <https://doi.org/10.1371/journal.pone.0213554>
- Zimmer, B. (2013). Is English ‘cooling’? A new claim by physicists triggers a turf war. Boston Globe, February 10. (Internetes elérhetőség: <http://benzimmer.com/wpd/wp-content/uploads/2013/03/cooling.gif>; letöltés időpontja: 2020. december 27.)

THE USE OF NATURAL EXPERIMENTS IN DEMOGRAPHIC AND SOCIOLOGICAL RESEARCH

ABSTRACT

This paper reviews the type of research known as natural experiment, which has become increasingly popular in the social sciences over the past several decades. While in true experiments, scholars themselves form experimental and control conditions and assign people or other units to them by the explicit use of some chance mechanism, in natural experiments, external forces outside of the control of the researcher create groups that for all practical purposes can be treated as if they were the products of randomization. The first part of the article discusses early examples of this design, whereas the next one treats some modern applications that demonstrate the potentials of this method. The third part is devoted to studies that are based on lotteries of some form and where, therefore, the criterion of random assignment is more likely to be met. This part also considers the problem that arises when the groups the researcher is really interested in comparing are partly the result of some non-random selection, such as when drafted soldiers escape actual military service. Two methods to solve this problem are discussed: intention-to-treat analysis and the use of instrumental variables. The paper concludes with some general remarks on the ability of natural experiments to strike a good balance between internal and external validity.