

Dr. Drótos Gergely ■ Dr. Janka Zoltán

## Akadályfutás a számokkal – a tudományos közlés buktatói

A tudományos vizsgálatok publikálását megelőző kutatói tevékenység rengeteg buktatót rejt magában. A legtöbb probléma gyökerét a vizsgálat megtervezésének szakasza jelenti, amely magába foglalja az alapkoncepció kidolgozását, a minta és a mérőeszközök kiválasztását, ám a jó alapok kidolgozását és a mérést a további módszertani és statisztikai hibák lehetőségét magukban hordozó adatelemzés és értelmezés lépései követik. Közleményünk a teljesség igénye nélkül foglalkozik a kutatás és a publikációs folyamat során felmerülő néhány hibával.

**KULCSSZAVAK:** MÓDSZERTAN, STATISZTIKA, VIZSGÁLTATÉTERVEZÉS, PUBLIKÁLÁS

A statisztika társadalmi megítélése meglehetősen változatos, amit mi sem bizonyít jobban, mint a Mark Twain által sokat idézett bölcsesség (amely nagy valószínűséggel Benjamin Disraelitől származik): „Háromféle hazugság létezik: a hazugság, a megátalkodott hazugság és a statisztika”. A kétes megítélés a szándékos megtévesztésen kívül több körülményre vezethető vissza, egyrészt az adatok elemzésével kapcsolatos hibákra, másrészt a helyes eredmények esetleges téves értelmezésére. Tudományos körökben az említett hibák kiküszöbölésére irányuló próbálkozások egyike a különböző módszertani iránymutatások megalkotása, mint pl. az American Physiological Society kiadványa,<sup>1</sup> illetve a számtalan cikk, amely a témában megjelent.<sup>2</sup> Tanulmányozva a nemzetközi szakirodalmat, elmondhatjuk, hogy egy kutatás során a legtöbb hiba a vizsgálat megtervezésével, a statisztikai elemzéssel, értelmezés-



**Dr. Drótos Gergely**

sel, valamint a publikálással kapcsolatos problémákból ered.

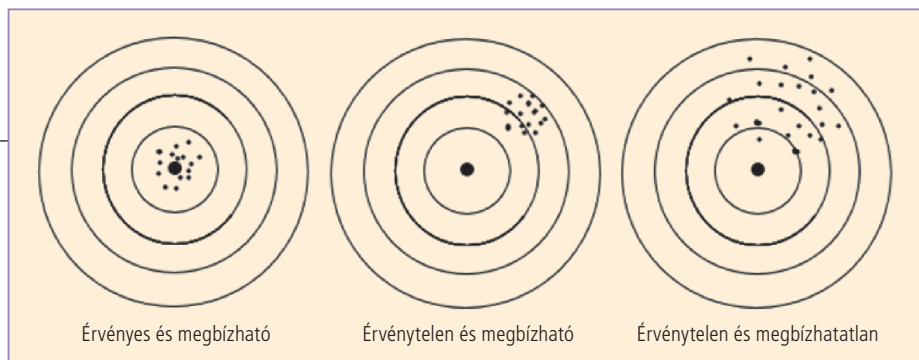
A könnyebb érthetőség és átláthatóság céljából egy gondolat kísérlettel fogjuk áttekinteni a kutatás egyes lépéseit, amely vérnyomáscsökkentő szerek hatékonyságával kapcsolatos, és segítségünkre lesz a különböző hibák szemléltetésében.

### A vizsgálat tervezése

A vizsgálat megtervezésének első lépése a megfelelő szakirodalmi háttér áttekintése után az ötlet, amelyet szükséges konceptualizálni. A gyakorlatban ez nem jelent mást, mint a vizsgálat alapvető kérdésének megfogalmazását, amely a nullhipotézis és az ellenhipotézis formájában írható le. A gondolat kísérletünkben két gyógyszer hatékonysága a vérnyomás csökkentésében a vizsgálat tárgya. Ekkor a nullhipotézis szerint egyik gyógyszernek sincs hatása a vérnyomásra, függetlenül a dózistól, az ellenhipotézis pedig az, hogy a kezelés hatékony. A hipotézisek megfogalmazásának részeként el kell dönteni, hogy melyek legyenek a függő és független változók. Az említett kísérletben a függő változó nem egyéb, mint a vérnyomás, független változók a vérnyomáscsökkentő gyógyszer két típusa, csoportonként három különböző dózisa és a vizsgálati személyek neme.

A vizsgálat tervezés e lépésénél rendkívül fontos – és a leggyakrabban mellőzött – lépés a szükséges mintanagyság (a vizsgálatban részt vevő személyek számának) becslése, amely szoros összefüggésben áll a statisztikai erő elemzésével (a gyakran használt angol szakkifejezés erre a „power analysis”). A statisztikai erő egyszerűen úgy ragadható meg, mint annak mértéke, hogy egy adott hatást a használni kívánt statisztikai próba a minta esetében milyen valószínűséggel mutat ki. A statisztikai erő

1. ábra. Az érvényesség és megbízhatóság kérdése céltáblán ábrázolva, középpontban a mérni kívánt értékkel. A mért adatokat a golyónyomok szemléltetik



vizsgálatának két fajtája ismeretes: az *a priori*, amely megelőzi a statisztikai elemzést és a tervezési folyamat részét, valamint az elemzést követő *post-hoc* forma. A *post-hoc* statisztika erőmeghatározásnak azonban nagyon kis jelentősége van, legfeljebb a tervezés hibájára hívja fel a figyelmet utólag, ez kevésbé ajánlott eljárás. Minden vizsgálatban kiemelkedően fontos a megfelelő méretű, nem túl kicsi és nem túl nagy mintanagyság kiválasztása, mivel ez befolyásolja a kutatás megbízhatóságát és eredményességét, más néven az elsőfajú (a nullhipotézis téves elutasítása, alfa, „ $p < 0,05$ ”) és a másodfajú hiba (a nullhipotézis téves elfogadása, béta) esélyét.

Ha a megfelelő méretű minta nagyságának becslése megtörtént, meg kell tervezni a véletlenszerű mintavételt; ha ez elmarad, vagy ha a megvalósítása nem megfelelő, akkor a vizsgálati csoportok már a beavatkozás előtt nagymértékben különbözhetnek egymástól, ami torzíthatja az eredményeket. A jól végrehajtott véletlenszerű mintavétel a teljes populációra nézve reprezentatív mintát eredményez. A véletlenszerű mintavétel helyes használata esetén nem mindig szükséges a csoportok kezdeti állapotának összehasonlítása statisztikai próbákkal, hiszen az esetleg felfedett különbségek hasonlatosak lehetnek az elsőfajú és másodfajú hibákhoz.

A kutatásban használt mérőeszközök kiválasztásának folyamatában figyelembe kell venni azok megbízhatóságát (reliabilitását) és érvényességét (validitását) a mérési hibák minimali-

zálása céljából. A reliabilitás nem más, mint az eszközzel kapott eredmények stabilitása az ismételt mérések alkalmával, míg a validitás lényege, hogy a kiválasztott eszköz valóban a számunkra fontos változókat mérje (1. ábra). Bizonyos, a mérést befolyásoló tényezők kiküszöbölésére megfelelő lehet az ún. vak vizsgálatok alkalmazása, míg egyes mérési hibák esetében megoldást jelenthetnek az olyan statisztikai eljárások, amelyekbe a mérőeszköz hibája beépíthető, mint pl. a strukturális vagy látens változós modellek (structural equation modeling, SEM).

Az eddigiek folyamán a gondolat kísérlet hipotéziseinek, az alanyok számának és kiválasztásának tervezéséről és a használandó mérőeszközök kiválasztásáról volt szó. A tervezés lépésével kapcsolatban előfordulhat, hogy a kéziratban nem teljesen egyértelmű a változók, illetve a hipotézisek megfogalmazása, illetve elmarad a mintanagyság becslése és a mérőeszközök, valamint a mintavétel pontos leírása. Példaként említhető Shere Hite könyve, a *Women in Love*, amely a nők szerelemről, szexualitásról és párkapcsolatokról alkotott véleményéről szól, s amelynek eredményei egy alapvetően hibás felmérésre támaszkodnak. A felmérés alapján az elvált nők 91%-a kezdeményező fél a válásban, és a legalább 5 éve házas nők 70%-a házasságtör-

rő. A könyv adatai százezer kiküldött kérdőívből származnak, melyek közül azonban csak 4500-at juttattak vissza a szerzőnek, tehát a minta véletlenszerűsége és ennek megfelelően eredményei erősen megkérdőjelezhetők.

## Az eredmények feldolgozása, elemzése

A kutatás következő lépésében az eddig kapott eredmények feldolgozása és elemzése a legfőbb feladat. Ennek a lépésnek a pontos leírása kiemelt fontosságú a publikálni kívánt kéziratban, törekedni kell a helyes elnevezésre („unpaired vs. impaired”/páratlan vs. gyengített) annak meghatározására, hogy a különböző próbákat pontosan mely adatoknál használták. Az elemzést meghatározza az adatok folytonossága, illetve kategorikus jellege és eloszlása. A kiválasztott statisztikai eljárások esetében alapvető hiba lehet alkalmazhatóságuk feltételeinek figyelmen kívül hagyása, pl. normális eloszlástól eltérő esetekben paraméteres próbák használata, mint a varianciaanalízis. Az alkalmazhatóság feltételére további példa lehet a gondolat kísérletben szereplő férfiak és nők arányának összehasonlítása a vizsgálati csoportok között. Ha khi-négyszet próbával vizsgáljuk a nemek arányát, figyelembe kell venni: a próba feltétele, hogy az egyes nemek minden cso-

**1. táblázat.** Az elfogadható közlemények aránya az 1966 és 1996 között megjelent átfogó tanulmányok alapján

Vizsgált közlemények száma	Vizsgált folyóiratok száma	Elfogadható közlemények (%)
295	10	28
77	1	48
139	1	55
79	2	39
74	1	34
114	1	28
86	4	10
243	2	15
120	4	<40
28	1	61
103	1	34
164	1	60
145	1	30

portban várhatóan öt főnél többször forduljanak elő (a teljes minta nemi eloszlásából és a csoportok nagyságából számítható ki).

Az alkalmazni kívánt próba kiválasztásával közvetlenül befolyásolható az elsőfajú hiba mérete. Ha több csoport között végzünk páronkénti összehasonlítást t-próbákkal, akkor minden egyes összehasonlítás növeli a téves eredmények valószínűségét, illetve ha a mérési eredmények két különböző csoportból származnak, akkor nem páros, hanem csak független mintás t-próba alkalmazható, amely egyébként teljesen más eredményre vezethet. A példában különböző személyek csoportjai kapnak különböző dózisu vérnyomáscsökkentőt, ezért az adatok normális eloszlása esetén megfelelő lehet az eredmények összehasonlítására a varianciaanalízis.

Többszörös összehasonlítás esetén gyakran előfordul az a helyzet, hogy a kutató több változó egyúttjárására kíváncsi, ebben az esetben azonban

különösen körültekintően érdemes eljárni. Ha 12 teljesen független folytonos változó kapcsolatát szeretné megvizsgálni, 66 korreláció kiszámítására van szükség, de ilyen módon nagyjából három nem valós kapcsolatot fogunk találni. Az elsőfajú hiba mérséklésére ebben a helyzetben megoldás lehet az alfa korrekciója, ám a szükséges szignifikanciaküszöb ( $p < 0,00076$ ) olyan alacsony lesz, hogy a statisztika ereje nulla közelébe csökken, azaz szinte semmi esélye a valós kapcsolatok megtalálásának. Megfelelő források esetén elvégezhető a teljes vizsgálat még egyszer, vagyis a mintaszámot növelhetjük, ezzel megtartva a megfelelő statisztikai erőt. Az ilyen jellegű vizsgálatokat a halászat és horgászat közti különbség analógiájával lehetne megragadni. Ha a horgász egy különleges fajtájú halat szeretne kifogni a Tiszából, akkor először tájékozik annak szokásairól – milyen csalival,

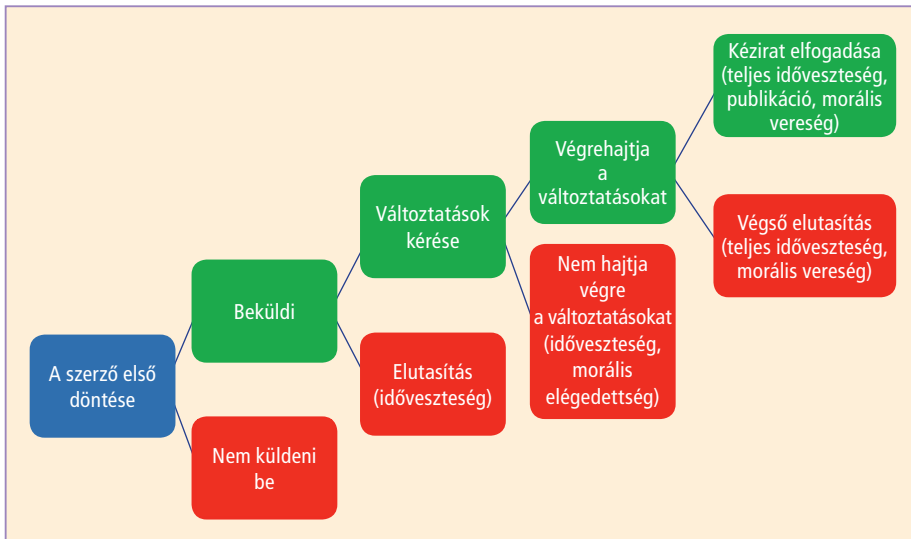
mely napszakban, pontosan hol van esély a megtalálására –, és ennek megfelelően cselekszik. A halász azonban kiveti hálóját valahol, és kifog rengeteg halat, amelyek nagy része teljesen érdektelen számára, esetleg a szemébe is kerülnek; szerencsés esetben a keresett ritka példányt is elkapja, ha felismeri a sokaságban. A két megközelítés között jelentős különbség, hogy a horgász jóval több ismerettel rendelkezik a keresett fajjal kapcsolatban, előre tudja, mit keres. Az adathalászat (rengeteg összefüggés egyidejű vizsgálata) egyedül abban az esetben tekinthető indokoltnak, ha a mintanagyság megfelelően nagy ahhoz, hogy az alfa korrekciójával is megfelelően nagy maradjon a vizsgálat eredményességének esélye.

Az eddigi problémákat összefoglalva Altman tanulmányának<sup>3</sup> eredményét érdemes áttekinteni, amely hasonló témájú átfogó munkákat foglal össze (1. táblázat). Altman alapján nem

**2. táblázat.** A *British Medical Journal* száz közleményével kapcsolatos hibák és hiányosságok 1991 és 1993 között

	Igen	Bizonytalan	Nem
A célok tisztáztak?	83	6	11
Megfelelő a vizsgálati elrendezés?	72	25	3
Megfelelő a minta?	83	6	10
Volt-e mintanagyságbecslés?	0	0	63
Megfelelő volt-e a válaszdók aránya?	49	23	2
A módszerek leírása megfelelő?	47	–	53
Helyes-e a statisztikai elemzés?	41	37	22
A statisztika leírása megfelelő?	14	–	86
Feltüntették a megbízhatósági tartományokat?	51(+8*)	–	41
A konklúzió megalapozott?	40	49	11
A közlemény statisztikai szempontból elfogadható?	4	–	96
A hibák kijavíthatók?	89	6	1

\*A megbízhatósági tartományokat nem megfelelően írták le



2. ábra. A publikációs folyamat

sok javulás tapasztalható módszertani szempontból 1966 és 1996 között, meg kell azonban jegyezni, hogy az idő előrehaladtával egyre több és bonyolultabb statisztikai eljárás állt a kutatók rendelkezésére, amelyek elsajátítása időt igényel, tehát ez az eredmény arra utalhat, hogy az új eljárásokat viszonylag lassan, nehézkesen sajátítják el. A *British Medical Journal* esetében 1991 és 1993 között a problémákat nagyobb részletességgel vizsgálta Altman, ebben az időszakban száz, egymást követő publikációt áttekintve meglepő eredményre jutott (2. táblázat). A vizsgált közleményekből 96 bizonyult valamilyen szempontból hiányosnak vagy hibásnak, leggyakrabban a minta nagyságának becslése hiányzott, vagy a statisztika bemutatása nem volt megfelelő.

### Statisztikai próbák, grafikus ábrázolás

A korábban felsorolt szempontok betartásával és hibák elkerülésével elérkeztünk a statisztikai próbák eredményeinek értelmezéséhez és bemutatásához. A példának vett kísérletben megtörténhet, hogy mind-

két gyógyszer hatásosnak bizonyult, mégpedig szignifikánsan különböző mértékben, a két gyógyszer hatásának különbsége azonban átlagosan csak 2 Hgmm volt. A hasonló eredmények értelmezésekor különbséget kell tenni a statisztikai és a praktikus értelemben vett szignifikancia között. Hiába mutatható ki eltérés két érték között, ha klinikai szempontból nem releváns a különbség mértéke. Az adatok leírásakor pedig megfelelő precizitással érdemes eljárni, nincs értelme négyöt tizedesjegy pontossággal közölni eredményeinket, ha a rendkívül kis értékek jelentősége elhanyagolható. Az értelmezéssel kapcsolatban gyakori hiba az ok-okozat kérdésének hibás felfogása, a korreláció legtöbbször csak együttjárást jelent, az összefüggés meglétére nincs bizonyíték.

A klinikai kutatások története során többször előfordult, hogy különböző vizsgálatokban statisztikai okokból nem volt kimutatható egyes kezelések hatása, és emiatt a valóságban meglévő különbségek nem, vagy csak nagyon lassan váltak bizonyítottá („a bizonyíték hiánya nem bizonyíték a hiányra”<sup>3</sup>). A módszertani hibák destruktív hatására szemlé-

letes példa, hogy a miokardiális infarktus kezelésében évekig vizsgálták a sztreptokináz alkalmazhatóságát, és csak jelentős késéssel sikerült bizonyítani hatékonyságát, mivel 24 kutatásból csak ötben sikerült kimutatni jelentős hatást.<sup>5</sup> Egy, a *Lancet*-ben megjelent közleményben a varixvérzés gyógyszeres (octreotid) és szklerotizáló kezelésének hatását hasonlították össze.<sup>6</sup> A végső konklúzió szerint a két kezelés egyformán hatásos. A közleményt jobban megvizsgálva kiderül azonban, hogy a nem megfelelő mintanagyság miatt mindössze 5% körüli esély volt jelentős különbség kimutatására.

Az adatok grafikus ábrázolása során is több hiba elkövetésére adódik lehetőség, némely esetben akár szándékosság is feltételezhető a kontextusukból kiragadott eredmények megjelenítése során.<sup>7,8</sup> Másrészt az ábrázolt alapvető jellemzők hibás értelmezése a mai napig sok kutatónak problémát okozhat.<sup>9</sup>

### Publikálás

Amennyiben a kutatásnak és a kézirat megírásának minden lépése lehetőség szerint hibáktól mentes, felmerül annak kérdése, hogy a kiválasztott tudományos lap munkatársai és lektorai mennyire tájékozottak módszertani kérdésekben. Frey a 2003-ban kiadott tanulmányában<sup>10</sup> éppen a tudományos közlés folyamatának buktatóira hívja fel a figyelmet, szerinte a kézirat elfogadásának folyamata jelentős problémákat rejt magában. A kutató számára elsődleges fontosságú kéziratának publikálása akadémiai pályá-

futása szempontjából, az esetleg hibás lektori véleményekre való válaszadás azonban morális kérdéseket vethet fel és a komoly idővesztés lehetőségét rejti magában (2. ábra). Természetesen a lektoroknak speciális helyzetükből adódóan (kevés felelősség) lehetőségük van az általuk leginkább helyesnek tartott változtatások kérésére, de sajnos nagyon kevés átfogó módszertani útmutató jelent meg, amely segítené munkájukat.

A különböző hibák korántsem teljes áttekintése után az a tanács juthat eszünkbe: a kutatói pályán mindannyian egymás szövetségesei és nem ellenségei vagyunk. Ha a kutatás és publikálás (vagy lektorálás) folyamán problémák vagy kérdések merülnek fel, ne habozzunk megkérdezni kollégáink véleményét és segítséget kérni tőlük.

Levelezési cím: [drotos.gergely@med.u-szeged.hu](mailto:drotos.gergely@med.u-szeged.hu)

#### **Irodalom:**

1. Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society. *J Appl Physiol* 2004;97:457–459
2. Strasak MA, et al. Statistical errors in medical research – a review of common pitfalls. *Swiss Med Weekly* 2007;137(3-4):44–49
3. Altman GD, Bland MJ. Absence of evidence is not evidence of absence. *Br Med J* 1995;311(7003):485
4. Altman GD. Statistical reviewing for medical journals. *Statist Med* 1998;17:2661–2674
5. Yusuf S, et al. Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results on mortality, reinfarction and side-effects from 33 randomized controlled trials. *Eur Hearth J* 1985;6:556–585
6. Sung JY, et al. Octreotide infusion or emergency sclerotherapy for variceal haemorrhage. *Lancet* 1993;342:637–641
7. Tufte ER. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983
8. Huff D. *How to Lie with Statistics*. W.W. Norton & Company, New York, 1954
9. Belia S, Fidler F, Williams J, Cumming G. Researchers misunderstand confidence intervals and standard error bars. *Psychol Methods* 2005;10(4):389–396
10. Frey BS. Publishing as prostitution? Choosing between one's own ideas and academic failure. *Public Choice* 2003;116:205–223