

Szerzői válasz az „Egy nagymintás exploratív vizsgálat érvényessége a limitációi tükrében” című olvasói levélre

Az exploratív adatelemző vizsgálatok helye az orvosi tudás előállításában

Az „Egy nagymintás exploratív vizsgálat érvényessége a limitációi tükrében” (1) című olvasói levélben a „Nagyszámú laboratóriumi érvényességi rangkorrelációval” (2) című közlemény érvényességének határaitól szerepelnek gondolatok.

Kontextus

Az orvosok számára evidens, hogy a D-dimerek emelkedett szintje jellemző a mélyvénás trombózisban és a tüdőembóliában szenvedő betegekben, azonban a teszt pozitivitása más betegségekben is előfordulhat, tehát a pozitív D-dimer-teszt nem bizonyítja a fenti betegségek meglétét. Egyéb tényezőket (*kontextus*) is vizsgálni kell a helyes diagnózis felállításához. A nyelvész számára evidens, hogy hiba keletkezhet abból, ha a szavakat önálló entitásként kezelik, kiragadva azokat a környezetükből (3, 4). *Mark Twain* egy alkalommal azt a csípős megjegyzést tette, hogy az angol humort nehéz megérteni, ha az ember nincs felkészülve rá (5). Az írás az adatelemzés különböző hagyományainak kontextusában mutatja be a kritizált feltáró vizsgálatot.

Történeti perspektíva

Az adatok összegyűjtésének, elemzésének és az eredmények értelmezésének, interpretációjának módszereit érdemes történeti perspektívából nézni, mivel ezen a területen az elmúlt kétszáz év során egymással párhuzamosan és egymásra hatva több, tudománytörténeti szempontból is fontos fejlődési mozzanatot láthatunk (6). Az olvasó így bepillantást nyerhet az adatelemzések dinamikus tudománytörténeti folyamatába anélkül, hogy statikus, esetenként dogmatikusnak tűnő kinyilatkoztatások csapdájába esne.

Érvényesség (validitás)

Vizsgáljuk meg először az olvasói levél címében szereplő *érvényesség* jelentését és kontextusát.

Campbell óta a nemzetközi szakirodalomban elfogadott megközelítés, hogy a validitás arra vonatkozik, hogy vajon azt mérjük-e, amit mérni vélünk vagy mérni szeretnénk (7). Amennyiben a kutatás validitása nagy, akkor az eredmények összhangban vannak a fizikai, kémiai, biológiai vagy szociális világ valódi természetével, jellemző vonásaival és változékonyságával, továbbá a meglévő elméletekkel. Tekintettel arra, hogy az elméletek maguk is változhatnak, a validitás definíciója paradox módon magában hordozza a tudomány fejlődése előmozdításának a lehetőségét éppúgy, mint akadályozását.

A nagy validitás egyik indikátora a *megbízhatóság*, tehát az, hogy megismételt mérés esetén hasonló eredmények keletkeznek. Ha egy módszer nem megbízható, akkor feltehetőleg nem is valid. Az *oksági* kapcsolatok feltárásával kapcsolatos validitás két fontos aspektusa, oldala a *belső* és a *külső validitás*. A *belső validitás* a kísérletek tervezésével, míg a *külső validitás* az eredmények általánosíthatóságával van összefüggésben. *Campbell* és *Stanley* a *belső validitást* a következő kérdésre adandó válaszként definiálta: „A kísérletben alkalmazott inger, stimulus szignifikáns különbséget okozott-e a vizsgált szempont szerint?” A *külső validitást* a szerzők a következő kérdésre adandó válaszként definiálták: „A kísérletben ki-váltott hatás milyen »populációra« (a hagyományos »frekventista« statisztikai gondolkodás alapfogalma a »populáció«, amely valós vagy elképzelt sokaság, amelyre a »mintából« következhetnek), kísérleti elrendezésre és változókra *általánosítható?*” A *Campbell*-cikk címében figyelemre méltó, hogy hangsúlyozottan *kísérletekről*, a később elkészült könyv címében pedig *kísérletekről és félig-kísérletekről* van szó (8). Később a validitás egyes kategóriáit átminősítették és a veszélyeket is átsorolták (9). *Shadish* és munkatársai hangsúlyozták, hogy az *érvényesség* a „populációról” levont következtetések igaz voltát jellemzi (9), másutt viszont a validitást fenyegető tényezőket nem az igazság, hanem a *hatásnagyság* becslésének pontosságával hozták kapcsolatba (10).

A *következtetések érvényessége* azzal kapcsolatos, hogy a kvalitatív vagy kvantitatív jellegű kutatásból vagy kísérletből levont következtetések mennyire *észszerűek*. A következtetések érvényessége azzal a kérdéssel kapcsolatos, hogy vajon „az adatok alapján van-e kapcsolat vagy nincs a vizsgált tényezők között.” A *statisztikai következtetések érvényessége* (SCV) értelemszerűen a számításokon alapuló összefüggések keresésére vonatkozik. Fontos észrevenni, hogy a validitás fogalmát és teljesülési feltételeit elsősorban az *oksági kapcsolatok* megragadását célzó kísérletekkel és félig-kísérletekkel kapcsolatban vizsgálták.

Az érvényességgel kapcsolatos kutatások vezettek ahhoz, hogy az egyes vizsgálati elrendezések között fontossági *hierarchiát* állapítsanak meg. Eszerint az *oksági kapcsolatok feltárásához* a legkevésbé alkalmasak a *megfigyeléses vizsgálatok* [különböző módszerekkel *gyűjtött adatok*, a *survey* típusú vizsgálatok (szociológiai felmérések), „hivatalos” adatok (11), internetes adatok stb.]. Ezeknél jobban alkalmasak az oksági kapcsolatok feltárására a *félig-kísérletek* („csak utána”, „előtte-utána” vizsgálatok, „megszakított idősoros elrendezések” stb.), elsősorban a lélektan, a társadalomtudományok és az orvostudomány területén. Az oksági kapcsolatok feltárására leginkább a *valódi kísérletek* alkalmasak, azok is elsősorban a fizika, kémia és biológia területén, amelyek között számos kísérleti elrendezés (factorial, fractional factorial, split-plot design stb.) található. A klinikai vizsgálatokat is a félig-kísérletek közé sorolják, amelyekkel a *véletlen besorolás* (random allocation) miatt egyfelől okságinak *tűnő* kapcsolatok igazolhatók, azonban a *mintakiválasztás önkényes* jellege (site) miatt több vizsgálat együttes elemzésére lehet szükség az eredmények általánosíthatóságához (metaanalízis). Mindezt azért is érdemes felidézni, mivel az alább tárgyalandó *exploratív* adatelemzések célja *nem* az oksági kapcsolatok, hanem sokkal inkább az adatok megismerése, a *lehetséges* összefüggések kezdeti vizsgálata, ezek vizualizációja, esetleg orvosi hipotézisek generálása. Az *érvényesség* fogalma tehát elsősorban az *oksági viszonyok feltárására irányuló kísérletes és félig-kísérleti vizsgálatok* ellenőrzésére szolgáló eszközrendszer, nem alkalmazható feltáró, exploratív vizsgálatokra.

Exploratív és konfirmatív kutatás

Történetileg a fenti folyamatokkal párhuzamosan, részben azokkal összefüggésben két, egymást kiegészítő megközelítés is kialakult. Az egyik a már említett *exploratív* (feltáró jellegű, megismerési célzatú), a másik a *konfirmatív* (megerősítő, oksági összefüggések bizonyítására törekvő) kutatás és a hozzájuk kapcsolódó statisztikai eljárások.

Az *exploratív* statisztikai elemzés fogalmát *Tukey* (12) vezette be úgy, hogy hangsúlyozta egyebek mellett az *adatok* vizuális megjelenítésének a fontosságát, újfajta, beszédes statisztikai ábrák elkészítését (13, 14), mindezt az *adatok* jobb megértése érdekében. Az exploratív jellegű kutatások elsődleges célja az adatok jobb megismerése, megértése, vizualizálása, a későbbi adatgyűjtések és adatfeldolgozások módjának meghatározásához támogatás nyújtása, a *lehetséges*

összefüggések előzetes felderítése, és esetleg hipotézisek generálása, amelyek általában a kutatás kezdeti szakaszában lehetnek hasznosak. Az exploratív kutatások eredményeképpen, elsődlegesen az orvosszakmai szempontok érvényesülésére tekintettel, *az orvosokkal szoros szakmai együttműködésben*, a *lehetséges* oksági összefüggésekről *előzetes* hipotézisek fogalmazhatók meg, amelyek később megalapozott konfirmatív kutatásban verifikálhatók vagy falszifikálhatóak.

A *konfirmatív* kutatások során általában célzott adatgyűjtések, félig-kísérletek vagy kísérletek alapján a pontosan megfogalmazott elgondolások statisztikai hipotézisvizsgálatoknak vetik alá. A *Neyman* és *Pearson* által kidolgozott konfirmatív statisztikai hipotézisvizsgálatok elméletét (15) eleinte rendkívül hevesen támadták. A kezdeti óriási ellenállás után a nullhipotézisre alapozott szignifikanciavizsgálat elmélete (NHST) gyakorlatilag dogmává merevedett és az eljárás alkalmazása rutinszerűvé vált. A NHST és a p-érték kiszámítási módja körül azonban időközben számos problémát azonosítottak (16), amelyeket nagyrészt (17), de nem teljes mértékben sikerült megoldani. Részben ezzel összefüggésben manapság egyre népszerűbbek a *bayesi* statisztikai megoldások, amelyek a hagyományos frekventista megközelítéstől alapvetően különböző filozófiai alapon nyugszanak, a kapott eredmények azonban sokszor számszerűleg rendkívül hasonlóak lehetnek a hagyományos statisztikai módszerekkel kapott eredményekhez (18).

Adatbányászat és gépi tanulás

Az adatbányászat a konfirmatív statisztikai módszerektől eltérő szemléletet használ: az adatbányászok *hipotézisek megtalálásának módjaira* fókuszálnak és azokat tesztelik a hipotézisalkotásba nem bevont adatrészen (19). A hagyományos adatbányászathoz tartozó eszközök a klaszterezés, az osztályozás, a nem klasszikus idősoros elemzések stb. Matematikai és számítástudományi ismeretek felhasználásával rengeteg új és hasznos adatfeldolgozási technikát fejlesztettek ki, ezeket összefoglaló néven gépi tanulásnak nevezhetjük. Három nagy csoportját különböztethetjük meg: *felügyelt tanulás*, *nem felügyelt tanulás* és a *megerősítéses tanulás*. A felügyelt tanulás egyik alcsoportja a mesterséges mély neuronhálózati modellek – amelyek egy részét egyebek mellett a prediktív modellezéseknél is rutinszerűen alkalmazzák, például SAS szoftver környezetben (20–22). A régi és új adatfeldolgozási, adatbányászati és gépi tanulási technikák bonyolult csalásfelderítési (23), kép-

feldolgozási, diagnosztikai, lingvisztikai (24) és számos egyéb problémák megoldására is hasznosnak bizonyultak (25, 26).

„Big data”

Időközben megjelentek a többféle értelemben vett nagyobb adatállományok is, amelyeket „big data” néven emlegetnek. Az eredeti közleményben idézőjelben szereplő „big data” általában az angolul három V betűvel jellemzett adatokat (Volume: méret, Velocity: sebesség, Variety: változatosság) jelenti (27, 28). A három V utalást tartalmaz, egyebek mellett a gyorsan nagy adattömegeket létrehozó *high throughput, next generation sequencing* (29–33) és más hasonló eljárásokkal keletkezett adatokra. A három V mellett újabb V-ket is javasoltak, ilyen például a V: veracity (igazságnak megfelelés). Mindazonáltal a „big data” kifejezés a mai napig nem pontosan definiált fogalom.

Statisztika, biostatisztika

A világot először számokká alakítjuk át. A számok nem beszélnek önmagukért. Mi beszélünk a számokról. Mi látjuk el a számokat jelentéssel (34), amelynek egyik eszköze a statisztika. Az egyik definíció szerint a biostatisztika olyan tudomány, amely lehetővé teszi a bemutatott tények alapján következtetések levonását a bennünket körülvevő világ jobb megértése, megismerése érdekében (35). *McElreath* szerint a statisztikai módszerek, eljárások a gólemekhez, robotszerű agyagszobrokhoz, kis robotokhoz hasonlíthatók (36). *McElreath* szerint a statisztika *nem* matematika. Sőt, szerinte a statisztika *nem* is tudomány. Szerinte a statisztika művelői leginkább a gépészmérnökökhöz hasonlíthatók, akik utasítások végrehajtására képes absztrakt gépeket, algoritmusokat terveznek, hoznak létre és működtetnek. Ebben a megközelítésben a statisztikai módszerek különböző, ember által konstruált és folyton változó eszközök gyűjteménye. Ismét mások szerint a statisztika az emberi megismerést (tudás-előállítást) és kommunikációt elősegítő eszközrendszer.

Adatelemzés, adattudomány

A történetileg alig néhány évtizedes múlta tehető *data science* lényeges eleme a tudás, jelen esetben az *orvosi tudás* és az *adatok* értelmezése és kommunikációja között a kapcsolat erősítése, és amelynek fő lépései az orvosokkal szoros együttműködésben az adatok importja, rendezé-

se, integrálása (*data integration*), transzformációja, megjelenítése, modellezése és kommunikációja. Nem statikus, hanem iteratív, körkörös visszacsatolásokra épülő elemzési folyamatról van szó.

Az adatelemzés, adattudomány és a matematikai statisztika összehasonlítása, szembeállításának kapcsán *Hayashi* a következőket írta: „...a matematikai statisztika művelői hajlamosak arra, hogy elszakadjanak a valóságtól. Ezzel szemben az adatelemzés módszereinek kialakításakor nem fordítottak túl sok figyelmet a matematikai statisztikára, mégis hasznos eszközöket hoztak létre bonyolult problémák megoldásához, amelyek a hagyományos értelemben nem mindig statisztikai következtetések, hanem gyakran leíró jellegűek” (37). A döntési fák, a véletlen erdő (random forest) és egyéb módszerek kidolgozásához köthető *Breiman* neve, aki a következőket írta: „Az adatok alapján a következtetések levonásához két kultúra létezik a statisztikai modellezésben. Az egyik szerint feltételezik, hogy az adatokat valamilyen ismert sztochasztikus, véletlen folyamat generálta. A másik kultúra képviselői algoritmikus modelleket használnak és az adatok keletkezési mechanizmusát ismeretlennek tekintik. A statisztikusok közössége többnyire az első kultúrához sorolja magát. Ez a fajta elköteleződés irreleváns elméletekhez, megkérdőjelezhető következtetésekhez vezetett és távol tartotta őket egy sor érdekes és fontos jelenleg meglévő probléma megoldásától” (38).

Következtetés

A validitás fogalmának ismeretében az olvasói levélben szereplő *érvényesség* alig alkalmazható olyan vizsgálatra, amely bevallottan és egyértelműen *exploratív, leíró, tapogatózó* jellegű volt, hiszen *nem kísérleti, nem félig-kísérleti* adatokkal foglalkozott, nem volt szó semmilyen kísérleti elrendezésről, a statisztikusok által használt értelemben nem szerepelt benne semmilyen „populáció”, nem szerepelt „minta”, az adatok keletkezési mechanizmusát ismeretlennek tekintettük, nem szerepelt benne semmilyen inger vagy stimulus. Kizárólag leírni, ábrázolni, megismerni kívánt adatok voltak, és a vizsgálat alapján a szerzők nem kívántak semmilyen oksági kapcsolatot bizonyítani, sőt, a kapott eredményeket nem kívánták általánosítani sem. A vizsgálat során nem történt semmilyen, a hagyományos *konfirmatív statisztikai értelemben vett hipotézisvizsgálat*, és a kapott p-értékek kiértékelése kizárólag *exploratív* jellegű volt. Az iparban széles körben elfogadott és alkalmazott „*exploratív p-érték-számítás*” azt jelenti, hogy a kapott p-értékek csupán előzetes, tájékozási célra használható iránymutatásnak tekinthetők, és ezekre p-érték-korrekciónak (Bonferroni-

korrekció, Holm-eljárás stb.) elvégzése sem szükséges (a gyógyszeriparban gyakran *secondary endpoints*). Ugyanez vonatkozik a mindössze érzékeltetésképpen elvégzett erőelemzésre is. A *konfirmatív* statisztikai elemzésekhez szokott statisztikusok szemszögéből nézve az alkalmazott módszerek és a kapott eredmények szokatlanok tűnhetnek, azonban az *exploratív* adatelemzések és az orvosi tudás szemszögéből nézve a kapott tájékoztató jellegű eredmények ismeretében több lehetőség kínálkozik később a célzott vizsgálatok megtervezéséhez.

Konstruktív kritikák

Megemlíthető, hogy az eredeti kéziratra összesen négy anonim lektori vélemény érkezett, ezekből kettő biostatistikai jellegű kéréseket tartalmazott. Az anonim lektorok úgy találták, hogy exploratív adatelemzési szempontból elfogadható ez az előzetes vizsgálat.

A már megjelent közleményre konstruktív kritikák is érkeztek orvosoktól. Az egyik klinikus azt javasolta, hogy az érzékeltetésképpen leírt erőelemzés helyett szimulálhatóak változónkként hasonló eloszlású véletlen adatok és az ott számított korrelációk felhasználhatóak összehasonlításra. A másik orvos pedig azt jelezte, hogy

néhány helyen tévedésből minta és mintanagyság szerepel a cikk szövegében – exploratív vizsgálat esetében mintáról és érvényességéről általában nem lehet beszélni, csak adatról és adatvizualizációról. Orvoskonferencián pedig felmerült, hogy a vérképadatokban az összefüggések exploratív feltárására használhatóak lennének a gépi tanulás egyes elemei (support vector machine, mesterséges neuralis hálók stb.), illetve a vizualizációra pedig az *Epskamp* által leírt módszer.

Az oksági viszony megállapítására *nem* törekvő *feltáró* vizsgálat (1) igazán meglepő eredménye az, hogy az ismétlődés, az adathalmazba való bekerülés valószínűségeinek ismeretlensége és további (széles értelemben vett) „zajok” ellenére az alkalmazott nagyon egyszerű, robusztus távolságmétri (adatgeometriai szempontból egy sokdimenziós térben vektorok között bezárt szögek értékeinek számítása) az orvosi tudás egy részét megragadta, leírta. Ez akár azt is jelentheti, hogy a nem kutatási célból gyűjtött zajos vérképadatok jobban hasznosíthatóak, mint eddig gondoltuk.

dr. Molnár D. László,

a vizsgálat orvos-biostatistikus konzultánisa

Brys Zoltán,

a vizsgálat adatelemzési részének vezetője

Irodalom

1. Singer J, Ferenczi I. Egy nagymintás exploratív vizsgálat érvényessége a limitáció tükrében *Lege Artis Medicinae* 2020;30(4-5):XXX-YYY.
2. Brys Z, Nagy E, Magyar G, Molnár DL, Kís JT. Nagyszámú laboratóriumi érvizsgálati eredmény exploratív jellegű vizsgálata rangkorrelációval. *Lege Artis Medicinae* 2019;29(12):609-14.
<https://doi.org/10.33616/lam.29.057>
3. Goodwin C, Duranti A. Rethinking context: An introduction. in: rethinking context: language as an interactive phenomenon. Cambridge: Cambridge University Press; 1992.
4. Whitehead AN. Philosophers do not think in a vacuum. dialogues of Alfred North Whitehead. Recorded by Lucien Price. David R. Godine. 2001.
5. Twain M. English know a joke, says Mark Twain. *The Evening World*; 1907. p. 2.
6. Stigler SM. Statistics on the table. The history of statistical concepts and methods. Harvard University Press; 1999.
7. Campbell DT. Factors relevant to validity of experiments in field settings. *Psychological Bulletin* 1957;54:297-312.
<https://doi.org/10.1037/h0040950>
8. Campbell DT, Stanley J. Experimental and quasi-experimental designs for research. Cengage Learning, 1963.
9. Shadish WR. Revisiting field experiments: Field notes for the future. *Psychological Methods* 2002;7:3-18.
<https://doi.org/10.1037/1082-989X.7.1.3>
10. Reichardt CS. Quasi-Experimentation. A guide to design and analysis. The Guilford Press, 2019.
11. Slattery M. Official Statistic. Tavistock. London, New York; 1986.
12. Tukey JW. Exploratory data analysis. Addison-Wesley, 1977.
13. Chambers JM, Cleveland WS, Kleiner B, Tukey P. Graphical methods for data analysis. Woodsworth & Brooks/Cole, 1983.
14. Tufte ER. Envisioning information. narratives of space and time. Graphics Press, 1990.
15. Neyman J, Pearson ES. On the problem of the most efficient tests of statistical hypotheses. *Phil Trans R Soc Ser A* 1933;231:289-337. <https://doi.org/10.1098/rsta.1933.0009>
16. Wagenmakers EJ. A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review* 2007;14(5):779-804. <https://doi.org/10.3758/BF03194105>
17. Mayo DG. Statistical inference as severe testing. How to get beyond the statistics wars. Cambridge, UK: University Printing House; 2018.
<https://doi.org/10.1017/9781107286184>
18. Dinya E, Molnár DL, Mészáros J, Solymosi N. Biometria a klinikumban. Feladatok bayesi megoldása. Budapest: Medicina Kiadó; 2019.
19. Bodon F. Adatbányászati algoritmusok. Budapest, 2010. <http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/adatbanyaszat.pdf> (Letöltés ideje: 2020. január 2.)
20. Sarma KS. Predictive modeling with SAS enterprise miner. Practical solutions for business applications. 2nd ed. SAS Institute Inc. 2013.
21. Lewis ND. Applied predictive modeling techniques in R. With step by step instructions on how to build them FAST! AusCov, 2015.
22. Kuhn M, Johnson K. Applied predictive modeling. Springer, 2016.

23. *Brown ILJ.* Developing credit risk models using SAS enterprise miner and SAS/STAT. Theory and applications. *SAS Institute Inc.*, 2014.
24. *Géron A.* Hands-on machine learning with scikit-learn, Keras & TensorFlow. Concepts, tools, and techniques to build intelligent systems. *O'Reilly*, 2019.
25. *Fernandez G.* Statistical data mining using SAS applications, 2nd ed. *CRC Press*, 2010.
<https://doi.org/10.1201/EBK1439810750>
26. *Williams G.* Data mining with rattle and R. The art of excavating data for knowledge discovery. *Springer*, 2011.
https://doi.org/10.1007/978-1-4419-9890-3_17
27. *Laney D.* 3D data management: Controlling data volume, velocity and variety. *META Group Research Note 2001;6:70*.
28. *Baumer BS, Kaplan DT, Horton NH.* Modern data science with R. *CRC Press*, 2017.
29. *Buffalo V.* Bioinformatics Data Skills. Reproducible and robust research with open source tools. *O'Reilly*, 2015.
30. *Datta S, Newlton D.* Statistical analysis of next generation sequencing data. *Springer*, 2014.
<https://doi.org/10.1007/978-3-319-07212-8>
31. *Pevsner J.* Bioinformatics and functional genomics. 3rd ed. *Wiley Blackwell*, 2015.
32. *Bhadhadhara K.* Statistical analysis of genomic data using R and bioconductor packages. *LAMBERT Academic Publishing*, 2017.
33. *Hofmann A, Clokie S.* Wilson and Walker's principles and techniques of biochemistry and molecular biology. *Cambridge University Press*, 2018.
<https://doi.org/10.1017/9781316677056>
34. *Silver N.* The Signal and the noise: Why so many predictions fail-but some don't. *Penguin Books*; 2015.
35. *Oliveira AG.* Biostatistics decoded. *Wiley*, 2013.
<https://doi.org/10.1002/9781118670767>
36. *McElreath R.* Statistical Rethinking. A Bayesian course with examples in R and Stan. *CRC Press*, 2016.
37. *Hayashi C.* What is data science? Fundamental concepts and a heuristic example. In: *Hayashi K, Yajima HH, Bock N, Ohsumi Y, Tanaka, Baba Y* (eds.). Data science, classification, and related methods. studies in classification, data analysis, and knowledge organization. *Tokyo: Springer*; 1998.
https://doi.org/10.1007/978-4-431-65950-1_3
38. *Breiman L.* Statistical modeling: The two cultures. *Statistical Science 2001;16(3):199-231*.