proLékaře.cz | 15.7.2025

Analýza dat v neurologii

LIX. Koncept atributivního rizika v analýze populačních studií – VI. Kauzalita vztahů

L. Dušek, T. Pavlík,
J. Jarkovský, J. Koptíková

Institut biostatistiky a analýz
Masarykova univerzita, Brno

doc. RNDr. Ladislav Dušek, Ph.D.
Institut biostatistiky a analýz
MU, Brno
e-mail: dusek@iba.muni.cz

Interpretace vztahu rizikového faktoru a jeho následku velmi často vede k diskuzi o kauzalitě vztahu, což je velmi významné analytické téma. Rozborem tohoto problému zde navážeme na výklad metodiky odhadu populačního atributivního rizika (*PAR*, díl 53.–58. seriálu), které vyjadřuje podíl rizikových událostí vzniklých v daném čase v důsledku působení jednoho nebo více rizikových faktorů. Z opačného úhlu pohledu jde o podíl incidence, který by nenastal, pokud by se podařilo daný rizikový faktor eliminovat. Již v samotné definici *PAR* je tedy kauzalita vztahu „rizikový faktor × následek" zabudována.

V předchozích dílech seriálu jsme se zmínili o tom, že výpočty atributivního rizika a *PAR* mají smysl pouze v případě, že mezi rizikovým faktorem a jeho následkem existuje skutečně kauzální vztah. To je logická a nutná podmínka, neboť hodnota atributivního rizika udává incidenční přírůstek v důsledku působení jednoho nebo více příčinných rizikových faktorů („*excess risk*"). Tuto úvahu lze ovšem zobecnit na všechny statistické ukazatele vztahů různých faktorů, tedy na analýzu kontingenčních tabulek, korelace a další. Ačkoli kauzalita vztahu podmiňuje smysl asociačních statistických analýz, statistické analýzy samotné ji nepochybně prokazují jen výjimečně. Statistická významnost počítaného ukazatele nemůže být jediným a konečným důkazem kauzality vztahu dvou a více veličin. A to nejen v důsledku možného vlivu zavádějících faktorů a možnosti výběrových zkreslení. Otázka existence příčinného vztahu představuje velmi komplexní hypotézu, od jejíhož potvrzení se odvíjí i další poznávání mechanizmu účinku či interakce zkoumaných faktorů.

Zkoumáme-li vliv určitého faktoru na výskyt nemoci, není tedy správné automaticky přisuzovat celkový počet nemocných v exponované populaci jeho příčinnému vlivu. V reálném světě je takto kauzálně vyvolán pouze určitý podíl onemocnění, který je ovšem tím větší, čím těsnější je vztah mezi expozicí a následkem. Identifikace kauzálního vlivu škodlivých faktorů je klíčová zejména ve studiích zaměřených na poškození zdraví v pracovním procesu; u nemocí z povolání je odhad míry rizika kauzálně spojeného s danou expozicí nezbytný („*exposure-induced disease*").

Podíl onemocnění způsobených vlivem expozice, odhadovaný jako populační atributivní frakce (*PAF*), se v literatuře někdy označuje také jako tzv. **etiologická frakce** (Etiologic Fraction; *EF*). Hodnota *PAF* neboli *EF* vyjadřuje rozdíl mezi celkovým rizikem nemoci (exponované i neexponované populace) a rizikem nemoci v neexponované populaci, vztaženo k celkovému riziku. Pro kvantifikaci kauzálního vlivu expozice a ve shodě s výkladem v předchozích třech dílech seriálu zde zavádíme tyto pojmy:

- **Populační atributivní frakce celkové populace** (*PAF*) vyjadřující očekávanou redukci (v %) počtu nemocných při eliminaci expozice zkoumaným faktorem v celé populaci; tzv. populační etiologická frakce.
- **Populační atributivní frakce exponované populace** (PAF_E) vyjadřující očekávanou redukci (v %) počtu nemocných při eliminaci expozice zkoumaným faktorem v populaci exponovaných; tzv. etiologická frakce exponované populace.

Budeme-li výpočet zapisovat formálně pomocí pravděpodobnosti, pak dostaneme následující vztahy:

$$PAF = [P(C) - P(C|NE)]/P(C),$$
$$PAF_E = [P(E) - P(C|NE)]/P(E),$$

kde

$P(C)$ je pravděpodobnost vzniku nemoci v celé populaci za daný časový interval,

$P(E)$ je pravděpodobnost vzniku nemoci v exponované populaci za daný časový interval,

$P(C|NE)$ je pravděpodobnost vzniku nemoci za daný časový interval za podmínky nulové expozice zkoumanému rizikovému faktoru.

Praktickou ukázku těchto výpočtů přinášíme v příkladu 1.

Pro správnou interpretaci hodnot *PAF* je důležité vždy uvádět, pro jaké období sledování je účinek dané expozice hodnocen. Správná volba doby sledování („*follow-up*") je další nutnou podmínkou úspěchu studií směřujících k průkazu kauzality vztahů. Doba sledování reflektuje nutnou časovou následnost mezi expozicí a následkem, která může trvat i desítky let (např. v případě expozice vedoucí ke zhoubným nádorům). Pozorování tedy nesmí být příliš krátké, nicméně problém je i nesmyslně dlouhodobé sledování. Představme si jako extrémní příklad studii, která kvantifikuje riziko kouření na celkovou mortalitu u kohorty mužů narozených v nějakém období a jako dobu, přes kterou bude sbírat data, nastaví 110 let. Potom je ovšem hodnota *PAF* nevyhnutelně nula, protože na konci sledování budou mrtví všichni kuřáci i nekuřáci a prokazování selektivního rizika kouření není možné. Tento příklad je jistě nerealistický, avšak dobře ilustruje riziko příliš dlouhé doby sledování, která již neprospívá srovnatelnosti exponované a neexponované populace a která naopak populaci neexponovanou přibližuje ve výskytu nemocných populaci exponované.

Je zřejmé, že statistické analýzy hrají při správné interpretaci vztahů expozice–ná-

proLékaře.cz | 15.7.2025

Zdrojová data příkladu představuje tabulka četností 2 × 2 z populační studie zkoumající vliv expozice faktorem X na následný výskyt určité nemoci.

Populace dle expozice faktorem X	Výskyt nemoci ano	Výskyt nemoci ne	Celkem	Odhad rizika nemoci *(incidence rate)*
exponovaní	500	9 500	10 000	0,050
neexponovaní	900	89 100	90 000	0,010
celkem	1 400	98 600	100 000	0,014

Z dat je patrné, že řešíme problém, kde kauzalita účinků expozice není 100%, neboť i neexponovaná populace vykazuje kvantifikovatelnou míru rizika vzniku dané nemoci. V takovém případě má odhad *PAF* (etiologické frakce) význam nejen pro celkovou populaci, ale i pro populaci exponovaných osob.

Nejprve odhadneme hodnotu *PAF* pro exponovanou populaci (PAF_E):

PAF_E = (0,050 – 0,010)/0,050 = 0,800.

Z primárních dat je patrné, že z celkového počtu nemocných 1 400 se pouze 500 vyskytuje v exponované populaci, a tedy pouze 35,7 % nemocných bylo exponováno zkoumaným faktorem (500/1 400 = 0,357). V žádném případě však nelze prohlásit, že těchto 35,7 % je přímým důsledkem samotné expozice. Jelikož hodnota PAF_E je 80 %, pak podíl nemocných přiřaditelný dané expozici kalkulujeme takto: 0,357 × 0,80 = 0,286. Výpočet implikuje fakt, že z 500 nemocných v exponované populaci pouze 400 (80 %) má vztah k dané expozici.

PAF jsme zde kvantifikovali podle vztahu:

PAF = podíl exponovaných osob × populační atributivní frakce v exponované populaci.

Ke stejnému výsledku dospějeme, pokud odhadneme populační hodnotu atributivní frakce pro celou populaci (*PAF*) dle vztahu uvedeného v tomto díle seriálu:

$PAF = PAF = [P(C) - P(C|NE)]/P(C) = (0{,}014 - 0{,}010)/0{,}014 = 0{,}286.$

Příklad 1. Ukázka výpočtu populační atributivní frakce (etiologické frakce) na celkové populaci a na populaci exponovaných jedinců.

sledek velkou roli. Nicméně průkaz kauzality vztahu nekončí odhadem atributivního rizika a jeho statistické významnosti, vždy je třeba posoudit více tzv. **kritérií kauzality**. V následujícím výčtu uvádíme nejčastěji používané postupy, z nichž každý nabízí poněkud jinou interpretaci prokazovaného vztahu a také vede na jiný typ hodnocení experimentálních dat:

1. **Statistická významnost asociace zkoumaných faktorů a síla jejich asociace.** Jsou-li zkoumané faktory kauzálně asociovány, pak by jejich vztah měl být prokazatelný ve frekvenčních tabulkách jako statisticky významná asociace (např. test dobré shody, statistika χ^2) nebo jinou statistickou analýzou. Rovněž následně odhadované ukazatele síly vztahu (poměr šancí; *OR*, relativní riziko; *RR*) by měly přesvědčivě dokládat existenci silné závislosti, a to i po vyloučení všech potenciálních zdrojů zkreslení a zavádějících faktorů.
2. **Prokázaný vztah dávky a účinku** je velmi silný důkaz možné kauzality vztahu, i když i zde mohou negativně působit zkreslující faktory. U binárních nebo kategoriálních faktorů je typickým příkladem vztah mezi zvyšující se hladinou expozice a rostoucí incidencí nemoci, u kvantitativních faktorů lze provádět korelační analýzu a přímo vykreslovat křivky „dávka–účinek".
3. **Následnost expozice a účinku.** Je samozřejmé, že u kauzálního vztahu musí expozice příčinného faktoru předcházet následek, a proto se průkazu kauzality musí přizpůsobit i vlastní sledovací či experimentální plán studií a jejich časové nastavení. Z tohoto pohledu jsou nejlepší volbou kohortové studie, které mají prospektivně pod kontrolou celou dobu sledování. Naopak studie případů a kontrol a zejména studie průřezové příliš ideální nejsou, neboť studují příčinu i potenciální následek v podstatě ve stejném čase.
4. **Konzistence závěrů různých studií.** Pokud se různé vzájemně nezávislé studie shodují v závěrech o síle vztahu zkoumaných faktorů, roste významně pravděpodobnost, že daný vztah je skutečně kauzální.
5. **Koherence informačních zdrojů.** Průkaz kauzality vztahu mezi škodlivým faktorem a nemocí by měl být v souladu i s jinými pozorováními, např. epidemiologickými. Prokážeme-li ve studii vztah mezi látkou X a nemocí Y, pak by např. pracovníci v průmyslu exponovaní zvýšenými dávkami této látky měli vykazovat vyšší incidenční zátěž danou nemocí apod. Informační zdroje by si neměly zásadně odporovat.
6. **Reverzibilita vlivu.** Průkazu kauzality velmi napomáhá, pokud jsme schopni na relevantních studiích doložit, že omezení expozice zkoumaným faktorem vedlo ke snížení jeho následků.

Výše uvedené postupy nepřinášejí do hodnocení kauzality vztahu žádné komplikované výpočty, spíše sledují komplexnost dostupných informací a jejich konzistenci. Nástroje zmíněné v bodech 1–3 má pod kontrolou přímo tvůrce studie rozhodující o jejím designu, velikosti vzorku, sběru dat a statistickém hodnocení. Body 4 a 5 spíše odkazují na práci s literaturou s cílem potvrdit popisovaný vztah z více zdrojů, např. metodami metaanalýzy.

Zcela novou kvalitou analýz je potom průkaz vztahu „dávka–účinek". Jsme-li schopni získat data kvantifikující dosažený účinek v různých hodnotách expozice, dostává asociace zkoumaných faktorů nový rozměr, a to směr. Nemusí přitom jít vždy o klasický lineární vztah, kdy s rostoucí dávkou proporčně roste i účinek. Reálný svět mnohem častěji nabízí komplikovanější profily. V každém případě jde o analýzy, které umožňují prostý průkaz existence vztahu dvou a více faktorů povýšit na studium možných dávkově dependentních synergických či antagonistických interakcí, kvantifikaci citlivosti subjektů k různým dávkám expozice, hledání oblasti největšího účinku apod. Těmto analýzám se budeme věnovat v příštím díle seriálu a na příkladech doložíme možnosti

proLékaře.cz | 15.7.2025

trendových analýz v tabulkách četností „expozice × účinek".

Ovšem ani sebelepší analýza nezabrání chybnému závěru, je-li studie od počátku zatížena zavádějícími faktory a špatným plánem sběru dat. Zde hraje dominantní roli znalost zkoumaného problému a zkušenost experimentátora, neboť vliv zavádějících faktorů může být velmi komplikovaný. Například nemoc sama může zpětně ovlivňovat expozici. Pacienti trpící závažnou chorobou mohou zásadně změnit svůj životní styl v pozitivním i negativním smyslu (omezení kouření, zlepšení stravy, nedostatek pohybu...), a tak ovlivňovat profil i účinek různých faktorů a rovněž nepřímo riziko vzniku a progrese dalších nemocí.

Na závěr tohoto dílu nabízíme čtenářům citaci čtyř prací, které lze k danému tématu doporučit pro jejich metodickou hodnotu.

Literatura

Greenland S, Robins J. Conceptual problems in the definition and interpretation of attributable fractions. Am J Epidemiol 1988;128(6):1185–97.

Walter SD. The estimation and interpretation of attributable risk in health research. Biometrics 1976;32(4):829–49.

Walker AM. Proportion of disease attributable to the combined effect of two factors. Int J Epidemiol 1981;10(1):81–5.

Whittemore AS. Statistical methods for estimating attributable risk from retrospective data. Stat Med 1982;1(3):229–43.