proLékaře.cz | 22.8.2025

Analýza dat v neurologii

XLIV. Grafy usnadňující studium zavádějících faktorů v asociačních studiích – II. Zdroj dat jako zavádějící faktor

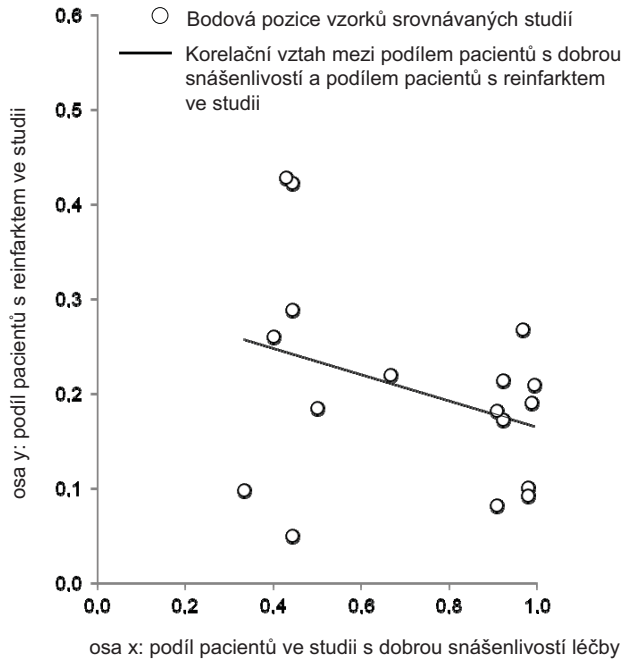
L. Dušek, T. Pavlík,
J. Jarkovský, J. Koptíková
Institut biostatistiky a analýz
Masarykova univerzita, Brno

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

V tomto díle seriálu bychom chtěli představit další zajímavé možnosti grafického znázornění vlivu zavádějících faktorů v asociačních studiích, tedy studiích identifikujících a kvantifikujících vztah „expozice–účinek". Připomeňme, že zavádějící faktor (*„confounding factor"*) definujeme jako proměnnou, která je významně asociována s expozičním faktorem (např. kouřením) a/nebo s následkem expozice (např. onemocněním plic). V tomto jednoduchém příkladu může být zavádějícím faktorem věk, neboť s rostoucím věkem se mění incidence sledovaného onemocnění a zároveň i podíl kuřáků v populaci. Zavádějící faktor může závažně zkreslit výsledky studie, zejména je-li nerovnoměrně distribuován ve srovnávaných ramenech studie. Jsou-li srovnávané skupiny kuřáků a nekuřáků (skupiny tvořené kategoriemi expozičního faktoru) různě staré, pak lze jen těžko zachycené rozdíly v incidenci onemocnění plic interpretovat. V takovém případě musíme výsledky na vliv věku korigovat (adjustovat), což jsme vysvětlovali v dílech XXXIX a XL našeho seriálu.

Minulý díl seriálu představil užitečné grafické nástroje vyvinuté pro snadnou identifikaci vlivu zavádějících faktorů, zejména grafy dle Paika (1985) a dle Bakera a Kramera (2001), tzv. BK graf. Při výkladu těchto technik jsme zdůrazňovali, že jsou připraveny pro analýzu situací,

Analyzujeme souhrnně výběr studií, které hodnotily výskyt reinfarktu u pacientů s dobrou a špatnou snášenlivostí antiagregační léčby. V analýze je třeba vzít v úvahu možný zavádějící vliv heterogenity mezi studiemi co do podílu pacientů s dobrou snášenlivostí léčby. Užitečnou vizualizací je bodový graf analyzující vztah mezi podílem pacientů s dobrou snášenlivostí a podílem pacientů s reinfarktem v jednotlivých studiích.



Interpretace: V našem příkladu pozorujeme, že studie s vyšším podílem pacientů s dobrou snášenlivostí k léčbě vykazují nižší pravděpodobnost výskytu sledované události. Podíl pacientů s dobrou snášenlivostí se však výrazně liší mezi studiemi, a je tedy zdrojem významné heterogenity.

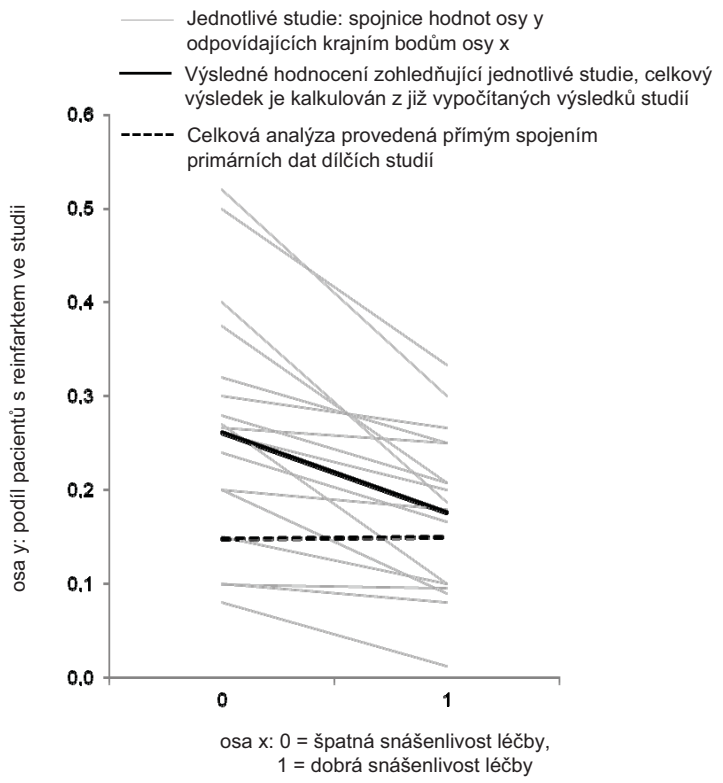
Výše popsaný vztah jako důsledek nevyvážené struktury pacientů mezi studiemi je v grafu dobře patrný.

- Negativní korelace popisující vztah mezi podílem pacientů s dobrou snášenlivostí a podílem pacientů s reinfarktem v jednotlivých studiích vede k závěru, že výskyt události (reinfarkt) klesne, pokud větší podíl pacientů bude mít dobrou snášenlivost léčby.
- Ačkoliv z praktického hlediska jde o pozitivní závěr, jde zároveň o situaci, která v případě sloučení dat ze studií s takto nevyváženou strukturou pacientů může vést k Simpsonovu paradoxu (viz příklad 2).

Příklad 1. Vizualizace výstupů studií v bodovém grafu navrženém v práci Rücker a Schumacher (2008).

proLékaře.cz | 22.8.2025

Srovnáváme více studií, které hodnotily výskyt reinfarktu u pacientů s dobrou a špatnou snášenlivostí antiagregační léčby. Příklad rozvíjí grafický nástroj prezentovaný v příkladu 1 a představuje čárový graf srovnávající relativní výskyt sledované události (osa y) mezi pacienty s dobrou snášenlivostí (x = 1) a špatnou snášenlivostí (x = 0). Sklon spojnice hodnot na ose y pro hodnoty osy x 0 a 1 ukazuje relativní vliv léčby na výskyt reinfarktu v rámci jednotlivých studií. Graf je dále doplněn čárami, vzniklými na základě analýzy spojených dat, kterou vysvětlujeme v legendě..



Interpretace: V našem příkladu pozorujeme, že ve všech dílčích studiích mají pacienti s dobrou snášenlivostí léčby nižší pravděpodobnost výskytu reinfarktu ve srovnání s pacienty se špatnou snášenlivostí.

Pokud provedeme souhrnnou analýzu na přímo sloučených datech studií bez jakéhokoli vážení jejich vlivu (černá čárkovaná čára, přímé spojení primárních dat), tento efekt dobré snášenlivosti se ztrácí a jako skupina s nižším rizikem vychází pacienti se špatnou snášenlivostí. Jde o důsledek Simpsonova paradoxu způsobeného nevyváženou strukturou vzorků ve studiích (viz příklad 1).

Jednotlivé studie se totiž podstatně liší zastoupením pacientů s dobrou snášenlivostí. Pro korektní spojení dat takto heterogenních dílčích studií je třeba spojení provést průměrováním jejich již vyhodnocených dílčích výstupů. Dílčím studiím lze přitom přidělit stejnou váhu anebo je vážit podle velikosti vzorku, variability, významu, apod. Tento postup vede k výsledku (černá plná čára) dokladujícímu vliv dobré snášenlivosti léčby na nižší pravděpodobnost výskytu rizikové události.

Příklad 2. Vizualizace výstupů více nezávislých studií v čárovém grafu navrženém v práci Rücker a Schumacher (2008).

kdy je potenciální zavádějící faktor binární (1/0) a snadno lze vyjádřit relativní četnost jeho kategorií v různých kohortách, resp. podsouborech většího souboru. Druhé omezení, které jsme dosud ve všech příkladech uplatňovali, je zaměření pouze na stratifikační faktory dělící celkový soubor dat na menší podsoubory. Příkladem může být dělení souboru pacientů a kontrol dle potenciálně zavádějících faktorů, jako jsou pohlaví, věk či anamnéza pacientů a kontrolních osob. V těchto situacích zkoumáme „interní" zavádějící faktory, kterými jsou charakteristiky osob či nemoci zkoumané v souboru.

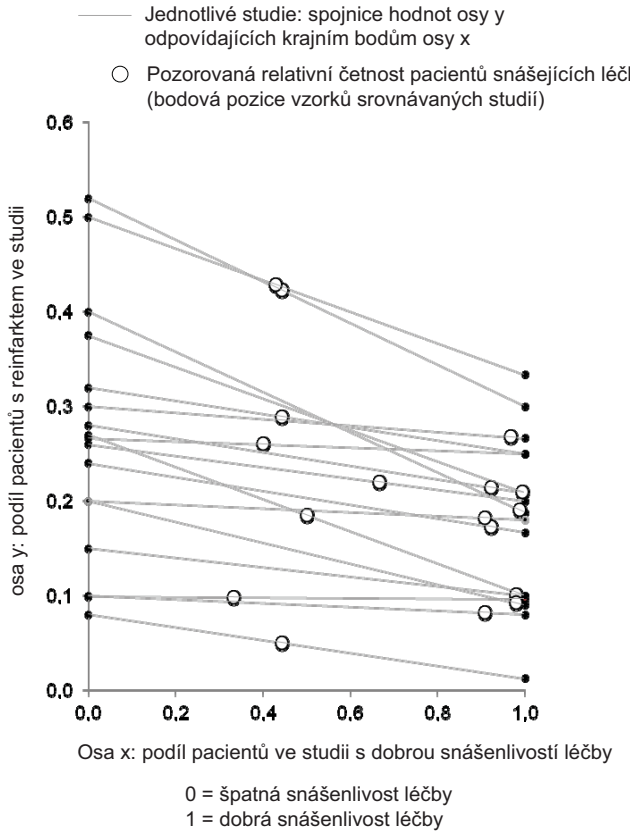
Avšak metody vědeckého zkoumání často vedou k jinému modelu spojování větších souborů z dílčích podsouborů. Jde o tzv. metaanalýzy, tedy rozbory již publikovaných, vzájemně nezávisle realizovaných studií, které zkoumaly stejný problém. Hlavním cílem metaanalýzy je zjistit, do jaké míry jsou výstupy dílčích prací konzistentní, a zda je tedy lze spojit do většího souboru, jenž by sledovaný fenomén potvrdil s větší reprezentativností a statistickou silou. V těchto úkolech rovněž spojujeme větší soubor z menších, ale kategorizační proměnnou jsou přímo dílčí zdroje dat, tedy jednotlivé publikované práce. Jde o „externí" třídicí faktor, na který v určitém smyslu můžeme rovněž pohlížet jako na potenciálně zavádějící. Dílčí studie mohou být zdrojem výrazné heterogenity, např. ve velikosti souborů, vyváženosti designu, době sledování, incidenci zachycených událostí apod. V důsledku toho můžeme při jejich spojování narazit na řadu problémů včetně extrémního jevu, kdy dílčí výsledky většinově prokazují statisticky významný vliv zkoumané expozice (např. léčby), ale spojený soubor významný efekt neprokáže (tzv. *„effect reversal"*).

Jistá komplikace je, že zavádějící faktor je zde proměnná „studie" (nebo také „zdroj dat" či „klinické centrum"), a ta je typicky kategoriální, nikoli binární (běžně provádíme metaanalýzu více než dvou zdrojů dat). Pro studium takových zavádějících faktorů byly navrženy nové grafické metody, které rozšiřují techniky představené v předchozím díle seriálu. Věříme, že jejich vysvětlením obohatíme metodickou výbavu čtenářů, neboť jde o grafy, které nebývají běžně vysvětlovány v učebnicích. V použitých příkladech jsme se nechali inspirovat prací Rückera a Schumachera (2008), jež velmi invenčně posunula možnosti srovnání výsledků z různých zdrojů dat. Rok jejího zveřejnění také dokládá, že jde o téma stále aktuální a zajímavé i pro recentní vědeckou literaturu.

V ideálním světě by různé zdroje dat, které mají být sloučeny za účelem společné analýzy, měly být zcela homogenní ve všech klíčových parametrech. Tato situace ovšem nastává jen zcela výjimečně, a proto je velmi podstatné posoudit možný zavádějící vliv rozdílů mezi studiemi (*„between trials"*). Pokud totiž pozorujeme v rámci individuálních stu-

proLékaře.cz | 22.8.2025

Srovnáváme více studií, které hodnotily výskyt reinfarktu u pacientů s dobrou a špatnou snášenlivostí antiagregační léčby. Tento příklad představuje smíšený graf kombinující výsledky bodového a čárového grafu z příkladů 1 a 2.



Interpretace: Komplexní vizualizace spojuje bodový a čárový graf, které představují odděleně příklady 1 a 2.

Z grafu je patrná nevyváženost jednotlivých studií v podílu pacientů s dobrou snášenlivostí léčby i vztah tohoto podílu k relativní četnosti sledovaného jevu. Snadno tak vyjadřujeme rozdíl ve výskytu události mezi pacienty s dobrou a špatnou snášenlivostí k aplikované léčbě v rámci jednotlivých studií.

Graf je dále možné interpretovat tak, že hodnoty na ose x představují všechny možné podíly pacientů s dobrou snášenlivostí léčby ve vzorcích studií a hodnoty na ose y jsou očekávanou pravděpodobností výskytu události v dané studii pro jednotlivé hodnoty x (pro x = 0 hodnota y odpovídá relativní četnosti události ve skupině se špatnou snášenlivostí, pro x = 1 četnosti události ve skupině s dobrou snášenlivostí). Při této interpretaci body v grafu popisují pozorovanou pozici jednotlivých studií a klesající úsečky očekávanou relativní četnost události y v závislosti na hodnotě osy x.

Příklad 3. Komplexní vizualizace výstupů více nezávislých studií dle práce Rücker a Schumacher (2008).

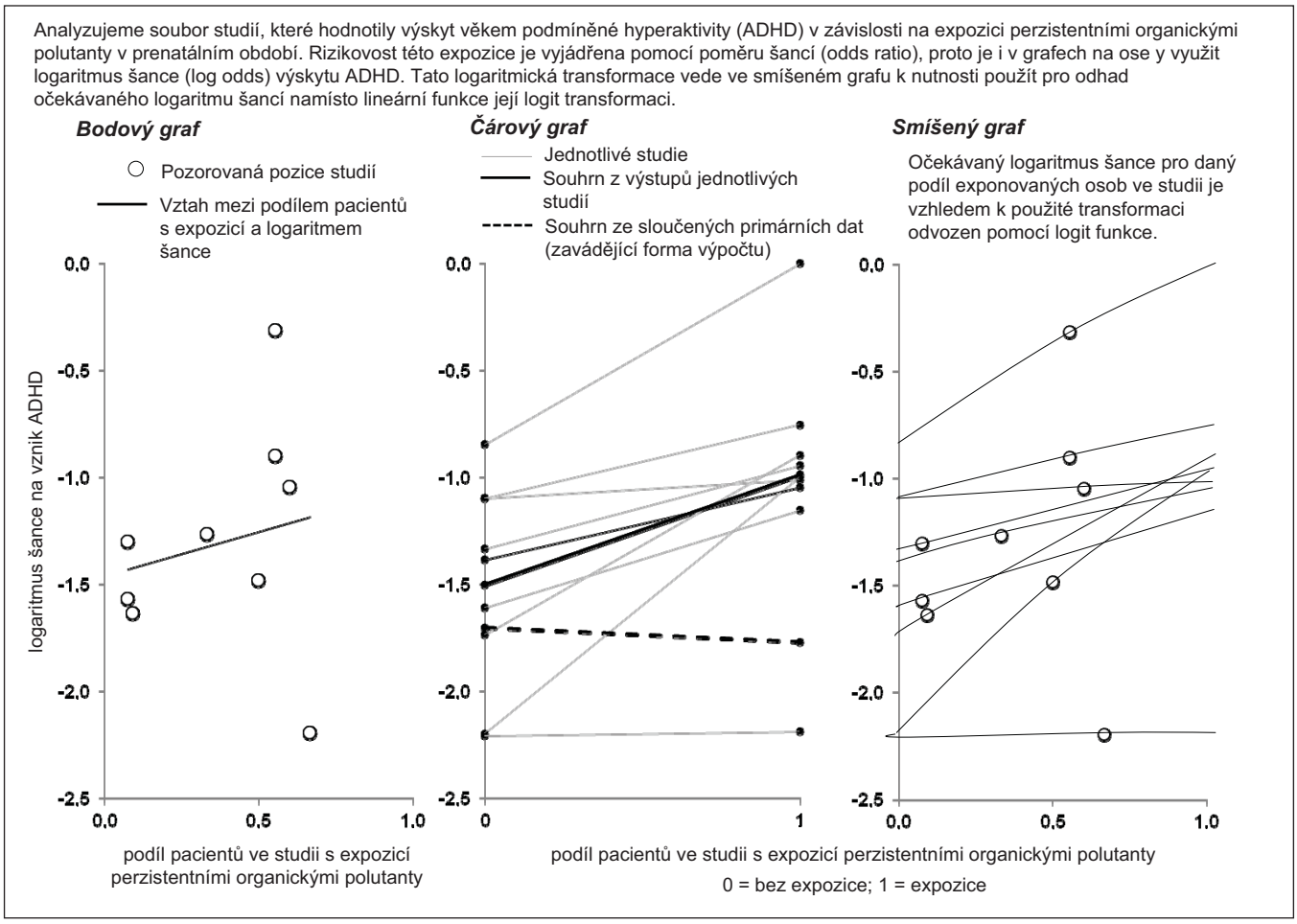
dií (*„within trials“*) nějaký konzistentní jev a společná analýza jej nepotvrdí nebo dokonce vyvrátí, je kritický rozbor srovnatelnosti studií zcela na místě. Rozdíly přitom velmi často nalézáme již v samotném experimentálním plánu, způsobu náběru probandů a také ve vyváženosti ramen. Příklady 1 a 2 ukazují, co vše může způsobit heterogenita dílčích studií spočívající v relativní velikosti experimentálního ramene.

Uvažujme studie se dvěma skupinami pacientů srovnávanými dle snášenlivosti primární léčby onemocnění (tento faktor, který zde reprezentuje expozici, označme X). Cílový parametr je výskyt určitého rizikového jevu, např. následného relapsu onemocnění (Y). Studujeme tedy pravděpodobnost (riziko) nastání relapsu ve vztahu k snášenlivosti léčby: $P(Y = 1|X)$. Rücker a Schumacher (2008) navrhli vykreslit pozici jednotlivých studií do jednoduchého grafu, kde na ose x bude relativní podíl pacientů v experimentálním rameni (tedy těch s $X = 1$) nebo obecně v jedné ze skupin (vztaženo k celkovému počtu probandů ve studii). U vyváženého designu je poměr obou ramen (skupin) 1 : 1, a x tudíž nabývá hodnoty 0,5.

Příklad 1 ukazuje srovnání více studií a prezentuje až extrémní heterogenitu mezi nimi, kdy hodnoty osy x mají rozsah od 0,3 do téměř 1. Pokud v bodovém grafu sledujeme závislost relativní četnosti sledovaného jevu Y (na ose y) na relativní četnosti pacientů s dobrou snášenlivostí léčby (osa x), můžeme snadno odhalit, jaký vliv mají rozdíly mezi dílčími studiemi na jejich výstup. Body v grafu z příkladu 1 lze nadto různě barevně nebo graficky diferencovat podle typu studie, roku zveřejnění či jiného hlediska (např. podle hodnot jiného zavádějícího faktoru). Jednoduchý graf nám tak dává plastický a vícerozměrný vhled do rozdílů mezi srovnávanými zdroji dat. Je ovšem nutné zdůraznit, že příklad 1 zobrazuje skutečně extrémní heterogenitu mezi dílčími studiemi a v takovém případě je na místě zvážit, zda jde vůbec o studie srovnatelné a zda všechny vstupy jsou pro společnou analýzu akceptovatelné.

Grafickou ukázku příkladu 1 rozvíjí dále příklad 2, a to pomocí čárového grafu. Zde na stejných datech zobrazujeme výsledky dílčích studií jako úsečku spojující hodnotu osy y (vlastní výstup studie, *„outcome“* – v našem případě relativní četnost sledované události) mezi dvěma krajními hodnotami osy x ($x = 0$ označuje skupinu pacientů se špatnou snášenlivostí terapie a $x = 1$ označuje skupinu pacientů s dobrou snášenlivostí terapie). Na sklonu jednotlivých úseček tak vidíme vliv (*„treatment effect“*) snášenlivosti nebo jiné studované expozice (X) na sledovaný jev. Rovněž můžeme snadno identifikovat, zda se studie v tomto efektu shodují či nikoli, anebo určit excentrickou studii s výsledky odlehlými od ostatních nebo nekonzistentními s ostatními.

proLékaře.cz | 22.8.2025



Příklad 4. Komplexní vizualizace výstupů více nezávislých studií, kde je efekt expozice vyjádřen poměrem šancí.

Příklad 2 doplňuje výsledky dílčích studií zobrazením výstupu analýzy spojených dat. Analýza dokládá, že při spojování studií, které jsou velmi heterogenní ve struktuře vzorku (studie se tu významně liší v podílu pacientů s dobrou snášenlivostí léčby), není správné přímo spojit jejich primární data. Tím totiž spojujeme velmi nevyvážené vzorky a snadno dopějeme k celkovému výsledku, který odporuje reálným výstupům většiny dílčích studií. Správný postup je kalkulovat celkový (průměrný) výsledek z výstupů jednotlivých studií a případně ještě tento výsledek vážit podle jejich velikosti či jiného kritéria.

V příkladech 1 a 2 je zobrazeno velmi důležité poselství, které se týká v medicíně velmi běžných srovnávacích analýz. Heterogenita dílčích klinických studií, např. v relativní velikosti experimentální větve vůči větvi kontrolní (u studií případů a kontrol používáme termín *„case-control ratio"*), může zásadně ovlivnit výsledek analýzy spojených dat. V takovém případě je skutečně lépe se vyhnout spojování primárních dat a raději spojovat až kalkulované výsledky ramen dílčích studií. V podstatě tím klinický výstup stratifikujeme podle zdroje dat, resp. studie. Oba grafické přístupy z příkladů 1 a 2 lze nadto spojit v jednom grafu, který přehledně dokladuje zavádějící vliv rozdílů jednotlivých studií. Toto spojení bodového a čárového grafu dokumentuje příklad 3. Opět vidíme značnou heterogenitu dílčích studií a zejména nejzávažnější zdroj zkreslení: studie s nejvyšším až excentrickým podílem pacientů s dobrou snášenlivostí léčby mají současně nejnižší výskyt sledované události.

Všechny grafy prezentované v příkladech 1 až 3 lze využít pro prostorově úspornou publikaci velkých objemů dat z více zdrojů. V roli studie jako stratifikačního a potenciálně zavádějícího faktoru může být také klinické centrum či nemocnice při srovnávání zdravotnických zařízení nebo obecně jakýkoli zdroj dat. Ukázkou takového publikačního využití je příklad 4, který srovnává výsledky devíti expozomových studií zkoumajících vztah mezi expozicí perzistentními organickými látkami v prenatálním období a následným vývinem nemoci, zde věkem podmíněné hyperaktivity u dětí (ADHD). Příklad zároveň svědčí o tom, že prezentovaná technika grafů se na ose *y* nemusí omezovat pouze na relativní četnost sledovaného jevu, jako jsou míra rizika či účinek expozice. V příkladu 4 pracujeme s logaritmem šance vzniku rizikové události, obdobně lze ovšem vykreslit i jiné kvantitativní výstupy studií.

Čtenáře při studiu příkladů jistě napadla otázka, čím je taková heterogenita srovnávaných studií způsobena a zda je v praxi vůbec možná. Odpověď na druhou část otázky je bohužel kladná. Podstatné rozdíly ve vyváženosti vzorku mezi studiemi existují a mohou vést až k publikování nesprávných závěrů, neboť ne vždy

proLékaře.cz | 22.8.2025

musí být odhaleny běžnými postupy meta-analýzy. Velké rozdíly ve struktuře vzorku mohou nastat i u randomizovaných prospektivních studií. V příkladu 1 až 3 jsme srovnávali studie, které se zásadně lišily ve výskytu rizikové události (reinfarkt) podle toho, jak velký podíl pacientů snášel aplikovanou léčbu. Jednotlivé studie smysluplně prokazovaly pozitivní terapeutický účinek (*„treatment effect"*), tedy dobrá snášenlivost léčby snižovala pravděpodobnost reinfarktu. Avšak silná nevyváženost vzorku studií mohla vést k souhrnnému opačnému závěru. Důvodů takové situace může být mnoho, zmíníme se především o dvou relativně častých:

- Způsob náběru pacientů do jednotlivých studií se lišil. Studie publikující velký podíl pacientů snášejících léčbu primárně „vychytávaly" tyto pacienty z klinické praxe a kontrolní soubory potom zpětně doplňovaly. Jakýkoli nereprezentativní nábor pacientů vnáší zkreslení i do výskytu sledovaných událostí.
- Sledování reinfarktu nebo obecně relapsu onemocnění je silně závislé na době sledování pacientů (tzv. *„follow-up time"*). Představme si situaci, kdy studie v příkladech 1–3 s největším počtem rizikových událostí měly také nejdelší dobu sledování, a tudíž větší šanci, že událost zachytí. Heterogenita v dosažitelné době sledování patří k nejzávažnějším a bohužel často skrytým příčinám zkreslení takových srovnání.

Je zřejmé, že při spojování různých zdrojů dat není opatrnosti nikdy dost. Potenciální zdroje zkreslení mohou být však skryté a statistické techniky je nemusí identifikovat, pokud na ně nemyslí sám experimentátor. Představené grafy umožňují pohodlný a pouhým zrakem kontrolovatelný průzkum velkých objemů dat stratifikovaných dle jejich původu a struktury vzorku probandů.

Literatura

Baker SG, Kramer BS. Good for women, good for men, bad for people: Simpson's paradox and the importance of sex-specific analysis in observational studies. J Womens Health Gend Based Med 2001; 10: 867–872.

Paik M. A graphic representation of a three-way contingency table: Simpson's paradox and correlation. Am Stat 1985; 39: 53–54.

Rücker G, Schumacher M. Simpson's paradox visualized: The example of the Rosiglitazone meta-analysis. BMC Med Res Methodol 2008; 8: 34–42.

Simpson EH. The interpretation of interaction in contingency tables. J Roy Stat Soc B 1951; 13: 238–241.