

# Analýza dat v neurologii

## XXXVI. Hodnocení statistické významnosti poměru šancí a relativního rizika

V minulém díle seriálu jsme otevřeli velmi frekventované téma hodnocení těsnosti vztahu (asociace) mezi expozicí osob nějakému faktoru a výskytem studované rizikové události (vznik nemoci, relaps nemoci, úmrtí...). V tomto a dalších pokračováních seriálu se zaměříme na správnou interpretaci výsledku těchto analýz. V příkladech minulého dílu jsme probírali specifika asociačních analýz u různých typů observačních (neintervenčních) studií, které jsme za tímto účelem rozdělili na prospektivní (kohortové), retrospektivní (studie případů a kontrol) a průřezové. Toto členění a odhady poměru šancí („odds ratio“, *OR*) a relativního rizika („relative risk“, *RR*) se k hodnocení vztahu různých jevů standardně využívají i v klinickém výzkumu, přičemž musíme respektovat zásadní pravidla, rovněž vysvětlená v XXXV. díle seriálu:

- *RR* lze využít pouze u prospektivních a průřezových studií, kde je výběr jedinců náhodný z hlediska výskytu sledované události; výskyt události zde sledujeme reprezentativně v čase.
- *OR* má své uplatnění naopak především u retrospektivních studií, i když jej lze odhadnout i u prospektivních nebo průřezových pozorování; jelikož u retrospektivních studií přímo zařazujeme osoby dle výskytu události, nelze výskyt události považovat za náhodný a reprezentativní – tyto studie směřují pouze k hodnocení vztahu mezi expozicí a výskytem události, neumožňují odhad populační incidence a prevalence sledované události.

V nejjednodušším případě je výstupem těchto pozorování 2 × 2 tabulka četností, ve které zaznamenáváme absolutní četnosti toho, kolikrát nastaly možné kombinace „expozice faktorem (ano/ne) vs událost (ano/ne), např. diabetes mellitus vs výskyt centrální mozkové příhody. Výpočet a rozdílnou interpretaci *OR* a *RR* přibli-

žuje obr. 1. Je patrné, že způsob výpočtu relativního rizika předpokládá reprezentativní výskyt sledované události v hodnocené populaci, neboť odhadujeme její relativní výskyt ve skupině exponované, a dělíme jej relativním výskytem události ve skupině bez expozice. Pokud by výskyt událostí v daném vzorku pacientů nebyl náhodný a reprezentativní, byl by odhad *RR* zkreslen. Naopak výpočet poměru šancí odhaduje šanci výskytu události v rámci exponované skupiny a tu dělí šancí výskytu události v rámci skupiny neexponované. Rozdíl mezi odhadem *OR* a *RR* je tedy zásadní a oba statistické ukazatele nelze zaměňovat.

Zde se hodí připomenout, proč jsou asociační studie v medicíně tak často využívány a co je jejich hlavním cílem. Tato pozorování slouží zejména k:

- identifikaci rizikových nebo protektivních faktorů, které ovlivňují výskyt nějaké klinicky významné události (komplikace léčby, recidivy nemoci, smrt); v epidemiologii takto studujeme i faktory ovlivňující samotný vznik, a tedy incidenci nemoci,
- posouzení a kvantifikaci těsnosti těchto vztahů a seřazení faktorů podle významu (síly) jejich vlivu,
- srovnání vlivu různých faktorů ovlivňujících výskyt sledované události mezi různými skupinami (kohortami) pacientů a kontrolních osob.

Je zřejmé, že odhady poměru šancí a relativního rizika směřují ke klinicky velmi závažným interpretacím. V dalším výkladu se zaměříme především na opatření, která pojmou dosažení věcně správného, nezkráceného výsledku těchto analýz. To je v první řadě schopnost rozpoznat statisticky významnou hodnotu poměru šancí, resp. relativního rizika, a tyto hodnoty mezi sebou relevantně srovnat. U srovnávacích analýz potom musíme dbát i na věcnou srovnatel-

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

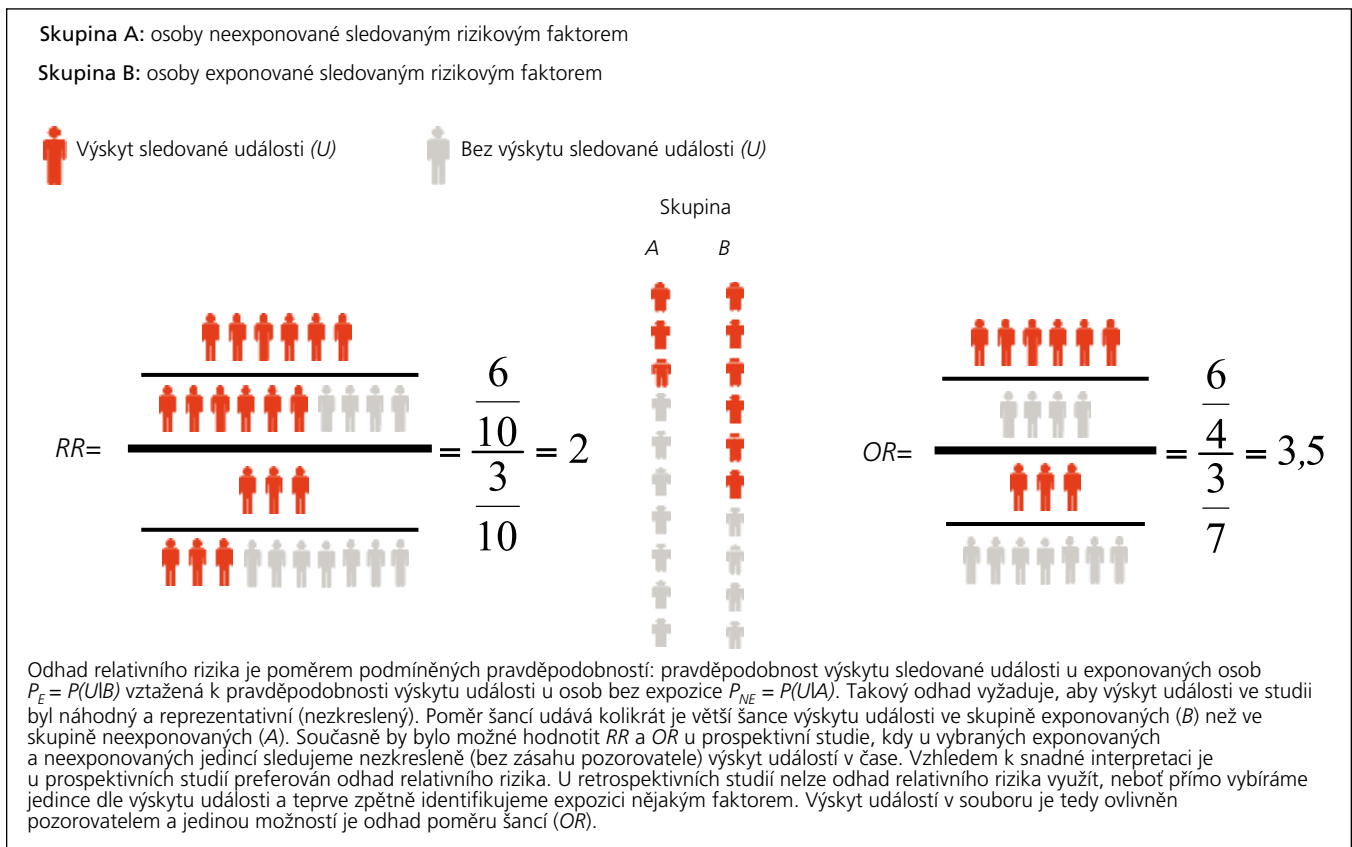
Institut biostatistiky a analýz  
MU, Brno



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

nost různých skupin pacientů a kontrol, jinak i dobře provedený výpočet povede k chybnému závěru.

Hodnocení statistické významnosti výsledku u jednoduchých asociačních studií není nijak složité. Nejjednodušší způsob, jak vyhodnotit význam odhadu poměru šancí i relativního rizika, je posouzení jejich intervalu spolehlivosti, přičemž podobně jako u jiných statistik se i zde standardně využívá 95% interval. Každý odhad *OR* nebo *RR* by měl být publikován se svým intervalem spolehlivosti, neboť bez něj nemůže být interpretace těchto ukazatelů plnohodnotná. Výpočet intervalu spolehlivosti pro odhad *OR* a *RR* dokumentují příklady 1 a 2, které pracují s různými soubory dat a nabízejí rozdílné výsledky. Interpretace je nicméně velmi jednoduchá. Pokud interval spolehlivosti zahrnuje hodnotu 1, nemůžeme daný odhad *OR* či *RR* považovat za statisticky významný. Hodnota *OR* = 1 nebo *RR* = 1 totiž označuje situaci, kdy sledovaný faktor (expozice) neovlivňuje výskyt dané události; jde tedy o referenční hodnotu („no effect“). Pokud tedy interval spolehlivosti zahrnuje hodnotu 1, pak nelze prokázat statisticky významný rozdíl ve výskytu události mezi exponovanými a neexponovanými jedinci. Za těchto okolností jednoduše nemůžeme vyloučit, že rozdíl odhadnuté hodnoty *OR* (*RR*) od 1 je důsledkem náhodných výkyvlů a variability v měření.



Obr. 1. Grafické srovnání výpočtu relativního rizika ( $RR$ ) a poměru šancí ( $OR$ ).

Pokud je bodový odhad  $OR > 1$  nebo  $RR > 1$  a korespondující interval spolehlivosti nezahrnuje hodnotu 1, je takto identifikován statisticky významný rizikový faktor (při 95% intervalu spolehlivosti hovoříme o 5% hladině významnosti). Obdobně hodnoty  $OR < 1$  nebo  $RR < 1$  s horní hranicí intervalu spolehlivosti pod 1 identifikují statisticky významný protektivní faktor.

Jistou zvláštností intervalů spolehlivosti pro  $OR$  nebo  $RR$ , například ve srovnání s intervalovými odhady pro průměr jako odhad střední hodnoty, je jejich asymetrie. Výpočty doložené v příkladech 1 a 2 pracují s aproximací pomocí normálního rozdělení pro hodnoty  $\ln(OR)$  a  $\ln(RR)$ . Po odlogaritmování (exponenciální funkcí) tak vždy nutně získáváme asymetrické hranice intervalů. Výpočty s pomocí aproximace na normální rozdělení jsou zjednodušením, které nám umožňuje snadný výpočet intervalů spolehlivosti. Konkrétně tento postup předpokládá, že logaritmus  $OR$  (tzv. Woolfova metoda) nebo logaritmus  $RR$  (tzv. Katzova metoda) mají přibližně normální rozdělení. Tento před-

poklad platí ale jen u dostatečně velkých souborů dat, proto by se tyto postupy neměly využívat u malých souborů. Jejich nesprávná aplikace může paradoxně vést k velmi úzkým intervalům spolehlivosti, které neodpovídají reálné skutečnosti. Neplatí totiž, že užší interval spolehlivosti znamená vždy lepší výsledek a lepší odhad  $OR$  nebo  $RR$ . Rovněž by byla velká chyba vybírat metodu odhadu intervalu spolehlivosti pokusně podle jeho výsledné šířky. U malých souborů dat jsou obecně doporučeny metody exaktních odhadů nebo iterační postupy, pro které je nutné využít specializované počítačové programy z důvodů výpočetní náročnosti. U odhadů  $OR$  lze v menu statistických programů nalézt tzv. Cornfieldovy meze spolehlivosti (podle Cornfieldovy metody výpočtu intervalu spolehlivosti), které jsou poměrně přesnou aproximací i u malých souborů dat. Jelikož tedy aproximace na normální rozdělení představená v příkladech 1 a 2 není univerzálně využitelná, doporučujeme uživatelům statistických programů konzultovat výběr vhodné metody s odborníkem na statistiku, a to ze-

jména u souborů malého rozsahu, kde přesnost odhadu intervalu spolehlivosti závisí i na tom, jak moc je hodnota  $OR$  ( $RR$ ) vzdálena od hodnoty 1.

Po přečtení předchozího odstavce čtenáře jistě napadne otázka, jaký je tedy skutečně správný postup při posuzování významnosti asociace náhodných veličin v tabulce četností. A může vůbec uživatel-nematematik posoudit statistickou významnost poměru šancí nebo relativního rizika? Odpověď je naštěstí pozitivní a správný postup lze stručně popsat v následujících krocích:

1. Zvážíme uspořádání (design) a typ hodnocené studie (retrospektivní, průřezová, prospektivní) a rozhodneme se pro odhad odpovídajícího ukazatele asociace (poměr šancí nebo relativní riziko).
2. Provedeme statistický test nulové hypotézy o nezávislosti veličin v kontingenční tabulce (Pearsonův chí-kvadrát test nebo Fisherův exaktní test) jako obecný test potvrzující nebo vyvracející hypotézu o nezávislosti zkoumaných veličin.

| Výskyt nádoru varlete (událost) | Podprůměrná porodní váha (expozice) |               |               |
|---------------------------------|-------------------------------------|---------------|---------------|
|                                 | Ano                                 | Ne            | Celkem        |
| ano                             | a = 17                              | b = 75        | a + b = 92    |
| ne                              | c = 200                             | d = 3 000     | c + d = 3 200 |
| Celkem                          | a + c = 217                         | b + d = 3 075 | 3 292         |

Vztah mezi expozicí a výskytem události můžeme obecně hodnotit pomocí Pearsonova chí-kvadrát testu a následně pomocí relativního rizika (RR) či poměru šancí (OR), které kvantifikují vliv expozice rizikovým faktorem na výskyt události. Níže uvedené výpočty všechna tato hodnocení dokumentují.  
IS: interval spolehlivosti  
SE: standardní chyba odhadu

$$RR = \frac{\frac{a}{a+c}}{\frac{b}{b+d}} = \frac{\frac{17}{217}}{\frac{75}{3075}} = 3,21 \quad SE(\ln(RR)) = \sqrt{\frac{1}{a} - \frac{1}{a+c} + \frac{1}{b} - \frac{1}{b+d}} = \sqrt{\frac{1}{17} - \frac{1}{217} + \frac{1}{75} - \frac{1}{3075}} = 0,259$$

=> RR (95% IS) = 3,21 (1,93; 5,34)

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{\frac{17}{75}}{\frac{200}{3000}} = 3,40 \quad SE(\ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} = \sqrt{\frac{1}{17} + \frac{1}{75} + \frac{1}{200} + \frac{1}{3000}} = 0,278$$

=> OR (95% IS) = 3,40 (1,97; 5,87)

$$\chi^2_{(v-1)} = \sum \frac{(\text{pozorovane} - \text{ocekavane})^2}{\text{ocekavane}} = \frac{(17 - 6,1)^2}{6,1} + \frac{(75 - 85,9)^2}{85,9} + \frac{(200 - 210,9)^2}{210,9} + \frac{(3000 - 2989,1)^2}{2989,1} = 21,72$$

=> statistická významnost testu: p < 0,001

V této variantě příkladu všechny tři způsoby hodnocení vztahu mezi expozicí a událostí poskytují konzistentní výsledek: intervaly spolehlivosti pro RR resp. OR nezahrnují hodnotu 1 a Pearsonův chí-kvadrát test vedl k jednoznačnému zamítnutí hypotézy o nezávislosti sledované expozice a události.

**Závěr:** existuje statisticky významný vztah mezi expozicí a výskytem události.

**Příklad 1. Hodnocení výstupů průřezové studie sledující vztah mezi podprůměrnou porodní váhou chlapců a výskytem nádoru varlat do 25 let věku – příklad s jednoznačným výsledkem.**

- Vypočítáme odhad OR nebo RR a doplníme jej intervalem spolehlivosti (standardem je 95% interval); pokud je analyzovaný soubor dostatečně velký a v každém políčku tabulky je absolutní četnost minimálně 5, pak lze aplikovat i jednoduchý postup vycházející z aproximace na normální rozdělení (viz příklady 1 a 2).
- Výstup testu a výpočet intervalu spolehlivosti provedené v bodech 2 a 3 se v naprosté většině případů vzájemně potvrzují. Tedy statisticky významný výstup testu v bodě 2 potvrzuje obecně existenci vztahu zkoumaných veličin a potom intervaly spolehlivosti pro OR nebo RR nezahrnují hodnotu 1. A naopak, pokud obecnými testy prokážeme nezávislost veličin (bod 2), pak ani OR, resp. RR, neposkytnou průkazné výstupy. Rozdíly mezi těmito postupy sice nastat mohou, ale pouze v hraničních situacích, kdy se může projevit např. chyba aproximativních výpočtů (příklad 2).

Doporučený postup hodnocení se tedy skládá ze dvou vzájemně komplementárních výpočtů. Statistické testy nezávislosti dvou veličin vedou ke standardnímu výstupu ve formě statistické p-hodnoty („p value“). Čím nižší hodnotu p obdržíme, tím menší je pravděpodobnost pozorovaného výsledku při platnosti nulové hypotézy (zde  $H_0$  odpovídá nezávislosti obou veličin), a tím pravděpodobnější je hypotéza alternativní, tedy že veličiny jsou vzájemně závislé. Například při výsledku Fisherova exaktního testu  $p = 0,013$  zamítáme nulovou hypotézu a obě veličiny označujeme za vzájemně závislé. Avšak tento test nevyjadřuje sílu hodnoceného vztahu a rovněž neurčuje věcnou, zde např. klinickou významnost prokázaného vztahu. Proto odpovídající odhad OR nebo RR (bod 3 výše uvedeného postupu) doplní statisticky prokázanou závislost veličin jako kvantitativní míra síly tohoto vztahu. Lze očekávat, že interval spolehlivosti OR (RR) za této situace nezahrnuje hodnotu 1.

Učebnicová interpretace 95% intervalu spolehlivosti říká, že budeme-li 100krát nezávisle opakovat odhad OR nebo RR na výběrovém souboru stejné velikosti, minimálně 95 % výsledků bude ležet v hranicích vymezených intervalem. Je tak vymezen interval, v němž leží s 95% pravděpodobností skutečná populační hodnota OR nebo RR. Výše jsme doložili, že interval spolehlivosti pro OR nebo RR lze rovněž využít jako indikátor vztahu dvou veličin na určité hladině významnosti. Jeho šířka však rovněž ukazuje na přesnost provedeného odhadu a souvisí i s velikostí analyzovaného souboru. Shoda závěrů vyplývajících z intervalů spolehlivosti a ze statistických testů nezávislosti dvou veličin (body 2–3 výše uvedeného postupu) tak nemusí vždy nastat. Například je možné, že obecným testem prokážeme závislost, nicméně interval spolehlivosti pro OR nebo RR bude velmi široký a může dokonce i zahrnout hodnotu 1. V takovém případě je nutná velká opatrnost při interpretaci zjištěné závislosti.

| Výskyt nádoru varle (událost) | Podprůměrná porodní váha (expozice) |               |               |
|-------------------------------|-------------------------------------|---------------|---------------|
|                               | Ano                                 | Ne            | Celkem        |
| ano                           | a = 17                              | b = 75        | a + b = 92    |
| ne                            | c = 200                             | d = 1 515     | c + d = 1 715 |
| Celkem                        | a + c = 217                         | b + d = 1 590 | 1 807         |

Vztah mezi expozicí a výskytem události můžeme obecně hodnotit pomocí Pearsonova chí-kvadrát testu a následně pomocí relativního rizika (RR) či poměru šancí (OR), které kvantifikují vliv expozice rizikovým faktorem na výskyt události. Níže uvedené výpočty všechna tato hodnocení dokumentují.  
IS: interval spolehlivosti  
SE: standardní chyba odhadu

$$RR = \frac{\frac{a}{a+c}}{\frac{b}{b+d}} = \frac{\frac{17}{217}}{\frac{75}{1590}} = 1,66 \quad SE(\ln(RR)) = \sqrt{\frac{1}{a} - \frac{1}{a+c} + \frac{1}{b} - \frac{1}{b+d}} = \sqrt{\frac{1}{17} - \frac{1}{217} + \frac{1}{75} - \frac{1}{1590}} = 0,259$$

$$\Rightarrow RR (95\% IS) = 1,66 (1,00; 2,76)$$

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{\frac{17}{75}}{\frac{200}{1515}} = 1,72 \quad SE(\ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} = \sqrt{\frac{1}{17} + \frac{1}{75} + \frac{1}{200} + \frac{1}{1515}} = 0,279$$

$$\Rightarrow OR (95\% IS) = 1,72 (0,99; 2,97)$$

$$\chi^2_{(v=1)} = \sum \frac{(\text{pozorovane} - \text{ocekavane})^2}{\text{ocekavane}} = \frac{(17 - 11,0)^2}{11,0} + \frac{(75 - 80,9)^2}{80,9} + \frac{(200 - 205,9)^2}{205,9} + \frac{(1515 - 1509,1)^2}{1509,1} = 3,83$$

$$\Rightarrow \text{statistická významnost testu: } p = 0,0501$$

V této variantě výpočtu různé způsoby hodnocení vztahu mezi expozicí a událostí neposkytují konzistentní výsledek:

Pearsonův chí-kvadrát test nezamítá jednoznačně nulovou hypotézu o neexistenci vztahu mezi expozicí a účinkem.

Interval spolehlivosti pro OR obsahuje hodnotu 1,00, a tedy nejsme odhadovanou hodnotu OR schopni odlišit od situace, kdy OR = 1.

Interval spolehlivosti pro RR leží hraničně nad hodnotou 1.

**Závěr:** Vzhledem k nejednoznačné statistické průkaznosti testu a intervalů spolehlivosti je nezbytná opatrnost při interpretaci výsledků.

## Příklad 2. Hodnocení výstupů průřezové studie sledující vztah mezi podprůměrnou porodní váhou chlapců a výskytem nádoru varlat do 25 let věku – příklad s nejednoznačným výsledkem.

Doufejme, že se nám příloženými příklady povedlo popsat prakticky využitelné postupy pro hodnocení významnosti odhadů poměru šancí nebo relativního rizika. Aproximativní postupy ukázané v příkladech 1 a 2 mají jedinou podmínku, a sice vyžadují dostatečnou velikost vzorku. Jinak jsou jednoduché a lze je provést i bez specializovaného SW. Avšak pomocí dostupných statistických programů lze snadno aplikovat jinou metodu pro vý-

počet intervalů spolehlivosti pro OR nebo RR. Z tohoto pohledu je dnešní doba velmi vstřícná, rozvoj počítačové techniky umožňuje provádět sofistikované analýzy i laickým uživatelům zmíněných SW nástrojů. Nikdy však nesmíme zapomenout, že ani sebelepší SW a metoda výpočtu z nás nesníma odpovědnost za rozhodnutí, zda v dané situaci využít OR nebo RR, a také odpovědnost za interpretaci výsledku. Statisticky významný výsledek testu prokazu-

jící závislost dvou jevů ve vzorku  $N$  osob neznamená, že ta na 100 % existuje i v cílové populaci, a rovněž nic nevypovídá o jejím klinickém významu. Z jednoduchých tabulek četností  $2 \times 2$  také nevyčteme a nezdůvodníme kauzální vztahy mezi veličinami a násilná interpretace výsledků může vést k zcela chybným, až paradoxním závěrům. Jak rozpoznávat rizika zkruslení a jak jim předcházet, se pokusíme přiblížit v dalším díle seriálu.

[www.urologickelisty.cz](http://www.urologickelisty.cz)