

ÚVOD DO INTERPRETÁCIE KLINICKÝCH ŠTÚDIÍ (3. časť)

Michal Mego¹, Mária Rečková²

¹ II. onkologická klinika LF UK, NOÚ, Bratislava

² POKO, Poprad

Tretia časť série príspevkov je orientovaná na základné štatistické ukazovatele používané pri referovaní výsledkov klinických štúdií. Jej cieľom nie je podať úplné štatistické vysvetlenie týchto pojmov, len poukázať na ich základné použitie pri interpretácii klinických štúdií.

THE INTRODUCTION TO INTERPRETATION OF CLINICAL TRIALS – PART 3

Third part of series of articles is focused at the fundamental statistical parameters used in reporting results from clinical studies. The aim of this part is to point out basic application of these parameters during interpretation of clinical trials, rather than completely statistically explain them.

Onkológia (Bratisl.), 2008, roč. 3 (4): 241–242

V nasledujúcom texte si na príklade vysvetlíme najčastejšie štatistické pojmy, s ktorými sa môžeme stretnúť pri referovaní výsledkov klinických štúdií.

Príklad: Štúdiá zameraná na zistenie efektivity kapecitabínu v adjuvantnej liečbe kolorektálneho karcinómu. Do štúdie bolo zaradených 200 pacientov, 100 pacientov dostávalo chemoterapiu a 100 pacientov bolo na ramene s placebom. Primárny cieľ bol prežívanie bez relapsu. Za udalosť sa považuje relaps choroby a smrť z akejkoľvek príčiny. Pri mediáne sledovania 54 mesiacov malo relaps 35 pacientov liečených kapecitabínom a 40 pacientov liečených placebom.

	kontrolná liečba	experimentálna liečba
udalosť	a	b
bez udalosti	c	d

miera udalostí kontrolnej liečby

$$(CER) = a/(a+c)$$

miera udalostí experimentálnej liečby

$$(EER) = b/(b+d)$$

Vyplníme si tabuľku:

	Placebo	Kapecitabín
udalosť	40	35
bez udalosti	60	65

$$CER = 40/(40+60) = 40/100 = 0,4$$

$$EER = 35/(35+65) = 35/100 = 0,35$$

Absolútna redukcia rizika (ARR) – vypovedá o absolútnom rozdieli medzi kontrolným a experimentálnym ramenom

$$ARR = CER - EER$$

V našom prípade: $ARR = 0,4 - 0,35 = 0,05$; po prepočítaní na percentá = 5 %

Relatívna redukcia rizika (RRR) – vypovedá, o koľko sa zníži (zvýši) riziko bez ohľadu na veľkosť

bazálneho rizika. ARR je klinicky relevantnejšie ako RRR, pretože RRR „ignoruje“ bazálne riziko. Preto malé rozdiely v riziku môžu byť výrazné, keď sa porovnajú s malým bazálnym rizikom.

$$RRR = (CER - EER)/CER$$

V našom prípade:

$$RRR = (0,4 - 0,35)/0,4 = 0,125; \text{ po prepočítaní na percentá} = 12,5 \%$$

Vidíme teda, že opticky vyzerá RRR výraznejšia ako ARR. Skutočný klinický prínos nám lepšie priblíži nasledujúca hodnota.

Počet pacientov potrebných liečiť (number needed to treat, NNT)

NNT udáva absolútny počet pacientov, ktorí musia byť liečení, aby sa zabránilo vzniku jednej udalosti, t. j. aby jeden z nich mal z liečby úžitok. Obdobné spojenie platí i pre vznik nežiaduceho účinku, čo sa niekedy označuje ako NNH (number needed to harm).

NNT = 1/ARR resp. 100/ARR (ak ARR udávame v percentách)

V našom prípade:

$$NNT = 100/ARR$$

$$NNT = 100/5$$

$$NNT = 20 \text{ resp. } 1/ARR = 1/0,05 = 20$$

V praxi to teda znamená, že 20 pacientov je potrebné liečiť adjuvantnou liečbou kapecitabínom, aby jeden pacient mal z liečby úžitok a nedošlo u neho k recidíve primárneho ochorenia.

Testovanie hypotéz a štatistická signifikancia

Cieľom klinickej štúdie je testovanie hypotézy, ktorá má klinickú relevanciu. V praxi väčšinou platí, že je jednoduchšie určité tvrdenie vyvrátiť, ako potvrdiť. Preto rozlišujeme nulovú a alternatívnu hypotézu a cieľom je vyvrátenie nulovej hypotézy. Ak predpokladáme, že existuje rozdiel medzi štandardným a ex-

perimentálnym ramenom, tak nulová hypotéza hovorí, že neexistuje skutočný rozdiel medzi sledovanými ramenami a pozorovaný rozdiel je daný iba náhodou. Alternatívna hypotéza na druhej strane hovorí, že existuje skutočný rozdiel medzi sledovanými ramenami a pozorovaný rozdiel nie je daný iba náhodou.

	Skutočnosť	
Rozhodnutie	Nulová hypotéza pravdivá (rozdiel neexistuje)	Nulová hypotéza nepravdivá (rozdiel existuje)
Nulová hypotéza pravdivá	správne	chyba II. typu
Nulová hypotéza nepravdivá	chyba I. typu	správne

Pri testovaní hypotéz môže dôjsť k dvom typom chýb – k tzv. chybe I. typu (alfa chyba), alebo k chybe II. typu (beta chyba).

Ako **chyba I. typu (chyba α)** je označovaná falošná pozitivita, to znamená, že nulová hypotéza je odmietnutá, napriek tomu, že je pravdivá. Potvrdíme teda, že rozdiel existuje, napriek tomu, že neexistuje.

Ako **chyba II. typu (chyba β)** je označovaná falošná negativita, to znamená, že nulová hypotéza je prijatá napriek tomu, že je nepravdivá. Potvrdíme teda, že rozdiel neexistuje napriek tomu, že existuje.

Príklad: (pozri tiež fázy klinických štúdií, Úvod do interpretácie klinických štúdií, časť I).

Nulová hypotéza pre štúdiu fázy II: objektívna odpoveď na liečbu A **nie je** väčšia ako 20 %.

Alternatívna hypotéza: objektívna odpoveď na liečbu A je väčšia ako 20 %.

Nulová hypotéza pre štúdiu fázy III: Liečba A **nevedie** k viac ako 20 % zvýšeniu počtu remisí v porovnaní s liečbou B.

Alternatívna hypotéza: Liečba A vedie k viac ako 20 % zvýšeniu počtu remisí v porovnaní s liečbou B.

p-hodnota je daná arbitrárne (konvenciou) a je vyjadrením sily dôkazu, že nulová hypotéza, ktorá hovorí, že rozdiel v sledovaných parametroch v populácii je nula, je nepravdivá. To znamená, že p-hodnota je vyjadrením sily dôkazu proti nulovej hypotéze. Čím menšia p-hodnota, tým je silnejší dôkaz, že nulová hypotéza je nepravdivá (teda rozdiel existuje).

$p = 0,05$ znamená 5 % riziko falošnej pozitivity (teda zistenie, že rozdiel existuje (nulová hypotéza je nepravdivá), napriek tomu, že neexistuje).

$p = 0,01$ znamená 1% riziko falošnej pozitivity (teda zistenie, že rozdiel existuje (nulová hypotéza je nepravdivá), napriek tomu, že neexistuje).

Konvenčne dohodnutá možnosť pochybenia I. typu je menej ako 5 %, t. j. **p** nižšie alebo rovné 0,05. Treba si však uvedomiť, že:

- Potenciálne medicínsky dôležité pozorované rozdiely v malých štúdiách, kde p-hodnota je vyššia ako 0,05, sú uzatvárané ako štatisticky nesignifikantné a nie je im venovaná pozornosť. Avšak, nemusí to byť vždy pravda.
- Všetky štatisticky signifikantné nálezy sú považované za výsledok skutočného liečebného efektu, avšak pri p-hodnote 0,05 je v priemere v 1 z 20 prípadov platná nulová hypotéza.
- Všetky štatisticky signifikantné nálezy sú považované za medicínsky dôležité, avšak ak budeme mať dostatočne veľký súbor, bude možné detekovať aj extrémne malý, štatisticky signifikantný rozdiel v sledovanej populácii.

Okrem p-hodnoty je potrebné všimnúť si vždy aj interval spoľahlivosti (*confidence interval*). Malý rozdiel nemusí byť klinicky signifikantný, a tak platí, že štatistická signifikancia neznamená vždy aj klinickú signifikanciu. Tiež platí, že v malých štúdiách, nemusí veľká p-hodnota znamenať, že nulová hypotéza je platná. Takéto štúdie nemajú dostatočnú silu, aby preukázali štatisticky signifikantný rozdiel. Absencia dôkazu však neznamená dôkaz absencie.

Sila testu (power) (sila testu = $1 - \beta$)

Je to schopnosť testu odhaliť rozdiel (napr. medzi štandardnou a experimentálnou liečbou), ak tento rozdiel naozaj existuje. Je to schopnosť správne rozoznať, že nulová hypotéza je nepravdivá. Čím je súbor väčší tým sa sila testu zvyšuje. V klinických štúdiách sa jej hodnota pohybuje v rozpätí 80 – 90 %. Napríklad sila štúdie 85 % znamená,

že existuje 85 % šanca, že štúdia zistí štatisticky signifikantný rozdiel medzi experimentálnou a štandardnou liečbou, ak tento rozdiel skutočne existuje.

Jej význam spočíva v tom, že ak máme malú štúdiu s nízkym počtom pacientov a výsledok tejto štúdie je „nesignifikantný“, t. j., že sa nenašiel štatisticky významný rozdiel medzi experimentálnou a štandardnou liečbou, nemusí to znamenať, že tento rozdiel tam nie je. Sila štúdie bola jednoducho nedostatočná na to, aby existujúci rozdiel odhalila. Zvýšením počtu pacientov sa zvýši sila štúdie, a teda i pravdepodobnosť, že ak existuje rozdiel medzi efektívnosťou skúmaných liečebných postupov, tak štúdia ho odhalí. Tiež z tohto dôvodu rôzne subanalýzy súboru uskutočnené pri vyhodnocovaní štúdie nemajú väčšinou dostatočnú silu na zistenie existujúceho rozdielu. I z tohto dôvodu sú výsledky *post hoc* subanalýz len hypotézy generujúce. Napríklad pri vyhodnocovaní štúdie skúmajúcej efektivity adjuvantnej chemoterapie FAC sa na konci uskutočnil subanalýza u pacientok nad 70 rokov. Keďže týchto pacientok bolo málo zaradených do štúdie, nepozorujeme signifikantný vplyv chemoterapie na ich prežívanie. Vzhľadom na nízky počet pacientov nevieme povedať, či je to prejav nedostatočnej efektivity chemoterapie, alebo je to spôsobené nedostatočnou silou testu.

Interval spoľahlivosti (confidence interval, CI) hovorí s určitou pravdepodobnosťou, že určitý parameter leží medzi dvoma limitmi. 95 % CI hovorí, že s 95 % pravdepodobnosťou sa daná hodnota nachádza v uvedenom intervale. Obdobne 90 % CI hovorí, že s 90 % pravdepodobnosťou sa daná hodnota nachádza v uvedenom intervale.

V praxi sa v publikáciách môžeme stretnúť ešte s nasledujúcimi termínmi:

Hazard function – v praxi je odhadnutý ako proporcia pacientov, u ktorých nastane udalosť (zomrú, majú rekurenciu a pod.) v určitom časovom intervale k počtu pacientov, ktorí prežijú čas t bez vzniku tejto udalosti.

$H(t)$ = počet zomrelých za časovú jednotku v intervale/počet prežívajúcich čas t .

Hazard ratio, HR – meria relatívny efekt liečby, je to vlastne miera zlyhania (failure rate) nejakej liečby.

HR = miera zlyhania novej liečby/miera zlyhania štandardnej liečby

HR < 1 preferuje novú liečbu

HR > 1 preferuje štandardnú liečbu

HR = 1 medzi liečbami nie je rozdiel

Literatúra

- Kirkwood BR, Sterne JAC. Medical Statistics. 2nd Edition. 2003 by Blackwell Science Ltd.
- Badenoch D and Heneghan C. Evidence-based Medicine Toolkit. BMJ Books 2002. BMJ Publishing Group.

Napríklad HR 0,875 znamená, že miera udalostí je v experimentálnom ramene je o 12,5 % nižšia $(1 - HR) * 100 \%$, teda $(1 - 0,875) * 100 \%$ v porovnaní so štandardnou liečbou. Rozdiel $(1 - HR)$ je zároveň i relatívnou redukciou rizika.

Relatívne riziko (relative risk, risk ratio, RR) je pomerom rizík v jednotlivých liečebných ramenech.

Na predchádzajúcom príklade je to teda pomer rizika vzniku kolorektálneho karcinómu v experimentálnom ramene, teda v ramene, kde pacienti dostávali kapecitabín a rizika vzniku kolorektálneho karcinómu v kontrolnom ramene, kde pacienti dostávali placebo. (EER = miera udalostí experimentálnej liečby, CER = miera udalostí kontrolnej liečby). Udalosť v našom prípade je teda vznik kolorektálneho karcinómu.

RR = EER/CER

V našom prípade $RR = 0,35/0,4 = 0,875$

Keďže je číslo menšie ako jedna, to znamená že riziko v experimentálnom ramene je nižšie ako v ramene kontrolnom (avšak v čitateli musí byť samozrejme EER).

Odds ratio, OR – pomer pravdepodobnosti (šance) – je to podiel frekvencie, s ktorou sa daný jav vyskytuje k frekvencii s ktorou sa nevyskytuje.

Na predchádzajúcom príklade je to pomer „šance“ (odds) objavenia sa verus neobjavenia sa kolorektálneho karcinómu v jednom verzus v druhom liečebnom ramene

OR = (b/d)/(a/c)

$OR = (35/65)/(40/60) = 0,54/0,66 = 0,82$

V prípade, že sa OR rovná jednej, tak sú „šance“ v oboch ramenách rovnaké.

OR je vždy viac vzdialený od hodnoty 1, ako RR, teda ak $RR > 1$, potom $OR > RR$

a ak $RR < 1$, potom $OR < RR$

V prípade zriedkavo sa vyskytujúcich udalostí je OR približne rovnaké s RR. V prípade, že sa udalosť vyskytuje pomerne často, môžu byť analýzy založené na výpočte RR postihnuté chybami výpočtu, ale tieto nevznikajú pri výpočtoch OR. V prípade OR sú závery rovnaké, či berieme do úvahy udalosť ako výskyt alebo absenciu udalosti. OR sa udávajú preferenčne v štúdiách „case-control“



MUDr. Michal Mego, PhD.

Národný onkologický ústav
Klenová 1, 833 10 Bratislava
e-mail: misomego@nou.sk