

METODICKÉ PROBLÉMY KLINICKÝCH STUDIÍ V ONKOLOGII

METHODOLOGICAL PROBLEMS OF CLINICAL TRIALS IN ONCOLOGY

NOVOTNÝ J.¹, PETRUŽELKA L.¹, PECEN L.²

¹ONKOLOGICKÁ KLINIKA VFN A 1. LF UK, PRAHA

²ÚSTAV INFORMATIKY AV ČR, PRAHA

Souhrn: Randomizované klinické studie jsou v dnešní době považovány za nejhodnotnější vědecký důkaz, který má ve svém konečném důsledku nalézt nový léčebný postup pro vybranou populaci pacientů, který je ve srovnání s dosavadním postupem účinnější, bezpečnější, levnější, anebo má více těchto pozitivních charakteristik současně. V článku se zabýváme problémy, které se objevují při využívání takových studií jako klíče při rozhodování o léčebném postupu u konkrétního pacienta.

Klíčová slova: randomizovaná klinická studie, metaanalýza, metodické otázky

Summary: Randomized clinical trials are considered to be one of the most important tools, that allow direct comparison of current standard and experimental therapies. Subsequently, the more efficient and/or better tolerated regimen is selected as the new standard. This article focuses on the problems with the interpretation of trial results as well as on the use of these trials in the selection of appropriate therapies for the individual patient.

Key words: randomized clinical trial, metaanalysis, methodological issues

Úvod

Randomizované klinické studie jsou stále častěji považovány za nejhodnotnější vědecký důkaz, který má přinést klinické praxi odpověď na otázku, zda nová experimentální léčba „B“ přináší našim nemocným více než stávající standardní léčba „A“.

Pokusme se ale zamyslet nad tím, zda tuto jasnou odpověď můžeme od klinických studií očekávat.

Problémy s designem studií

V tomto článku nechceme diskutovat otázky spojené s úvodními fázemi klinických zkoušek, ale zaměřujeme se na výsledky studií třetí fáze klinických zkoušek, které mají poskytovat údaje, jež umožňují provádění správné lékařské praxe. K tomuto účelu nejlépe slouží rozsáhlé randomizované studie¹⁾, ve kterých se jeden nebo více postupů experimentálních srovnává s ramenem standardním. V případě příznivého výsledku pak experimentální rameno nahrazuje stávající standard a stává se tak standardem novým. V randomizované studii můžeme testovat nejen nový léčebný postup, ale také i standardní léčebné postupy v kombinaci s potenciálním prognostickým či prediktivním faktorem. Randomizované může být též potenciální prognostický či prediktivní faktor stanovován. Vždy ale v randomizované studii musí vystupovat nějaký léčebný postup být i standardní²⁾.

Některé typy designů randomizovaných studií užívaných v onkologii ukazuje schématicky obrázek 1 (a to v souvislosti s prediktivním faktorem/markerem).

V onkologii se až na výjimky u randomizovaných studií fáze I obvykle nepoužívá crossover design, kde jsou v randomizovaném pořadí aplikovány dva nebo více léčebných postupů. Bohužel, ne vždy jsou randomizované studie připraveny důkladně, a tak se nezřídka setkáme s problematickým návrhem studie, který je krátce po její publikaci důvodem nespokojenosti řady kliniků. Jednou z nejčastějších chyb designu je nesprávně zvolené srovnávací rameno. Z neznámější tak-

to kritizovaných studií zmiňme např. studie hodnotící účinnost adjuvantní hormonoterapie proti chemoterapii, ve kterých se opakovaně setkáváme s neadekvátním chemoterapeutickým režimem, zařazováním pacientek s negativními hormonálními receptory v primárním nádoru, s nepodáním hormonální léčby pacientkám léčeným chemoterapií apod. Do této kategorie patří také studie s paklitaxelem v adjuvantní léčbě karcinomu prsu, které neodpovídají na otázku, zda je lepší léčit paklitaxelem či jen atracykliny, ale pouze poukazují na nedostatečnost 4 cyklů chemoterapie AC pro pacientky s uzlinovým postižením. Zatímco v některých případech výše uvedených studií musíme být shovívaví, neboť to, co se z dnešního pohledu zdá nesprávné, nebylo v době zahajování studií známo, jindy je volba kontrolního ramene zcela nepochopitelná. Příkladem takovéto prvotně nesprávné volby je „standardní“ režim paliativní chemoterapie karcinomu žaludku cisplatina - 5-fluorouracil, který byl testován proti experimentálnímu ramenu docetaxel - cisplatina - 5-fluorouracil. V literatuře však nenajdeme jedinou studii, která by nám dovolila zmíněný režim cisplatina - 5-fluorouracil považovat za standard paliativní léčby nádorů žaludku. Podobných příkladů bychom mohli uvést celou řadu.

Druhou běžnou chybou randomizovaných studií je nedostatečná velikost souboru nemocných. Ta bývá způsobena nesprávnou statistickou konzultací, nebo, což je mnohem častější, nedostatečným nábořem pacientů do studie. Bylo zjištěno, že třetina studií prováděných v USA se účastní méně než 75% plánovaného množství pacientů³⁾. Proto řada srovnání, která nejsou statisticky významná, může být způsobena právě nedostatečnou možností tyto rozdíly postihnout⁴⁾. Typickým příkladem tohoto jevu je celá plejáda studií, která zkoumala experimentální dvojkombinace léčiv proti monoterapii gemcitabinem u nemocných s inoperabilním nebo lokálně pokročilým karcinomem pankreatu, a která až na dvě výjimky neprokázala žádný přínos kombinované léčby na celkové přežívání nemocných. Minimální efekt, který dvojkombinace mohou u tohoto c-

hemorezistentního onemocnění přinést, by totiž znamenal nutnost extrémně velkých studií a pak, musíme také zvážit, zda je statisticky prokázán efekt také klinicky relevantní (např. prodloužení přežívání nákladnou léčbou o 1-2 týdny).

Problémy s výběrem nemocných

Velkou chybou mnoha studií je, že z nich jsou často vyřazováni starší nemocní nebo nemocní s alterací jaterních či ledvinových funkcí. Tento zvyk pochází z klinických testů léků na méně závažná onemocnění, ale tento postup není v případě onkologických onemocnění vždy optimální, protože právě podíl těchto prognosticky méně příznivých pacientů není v rutinní klinické praxi zanedbatelný. U nich se pak drobné zlepšení léčebných výsledků dosažené experimentálním postupem „B“ nemusí vůbec projevit. V případě kaskády předregistračních studií sice musí studie zaměřená na starší nemocné či na nemocné s alterací jaterních či ledvinových funkcí nutně existovat, může být ale rozsahem mnohem menší, ačkoliv v pozdější rutinní aplikaci nového léku je podíl takových pacientů značný.

Problémy s interpretací výsledků

Při čtení závěrů studie je kliník zavalen velkým objemem dat, jejichž interpretace bez dostatečné znalosti statistiky může být zavádějící.

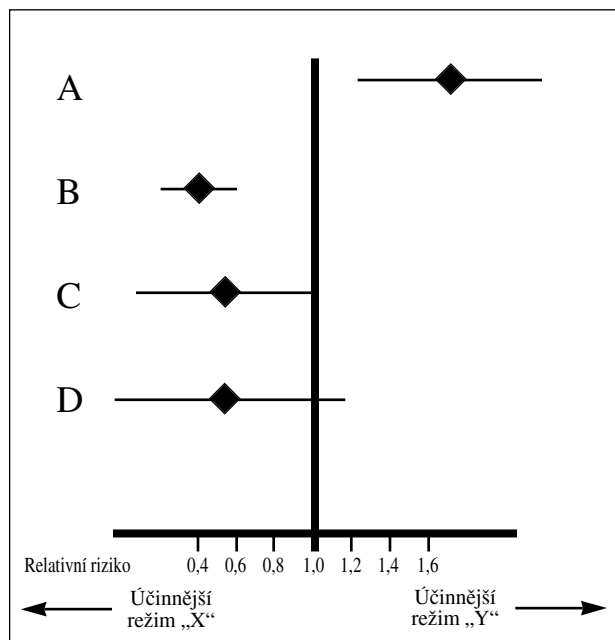
Máme-li k dispozici publikaci plánované průběžné analýzy, jde zcela určitě o závěr hodný zřetele. Průběžné analýzy se totiž publikují především tehdy, vyskytnou-li se nějaké neočekávané nežádoucí účinky léčby, o kterých by měl být kliník včas informován, anebo dochází k významnému zlepšení celkového přežívání nebo kvality života nemocných. V poslední době je však předčasné ukončování studií na základě plánovaných analýz poměrně časté i z jiných důvodů. To je však značně kontraproduktivní⁵⁾. Z nedávno takto ukončených studií zmiňme SWOG 9701 (udržovací léčba paklitaxelem 12 vs 3 měsíce) a MA-17 (5-leté užívání letrozolu vs placebo po 5 letech tamoxifenu). O tom, zda má být studie předčasně ukončena, zda má být studie odslepena a pacientům s méně účinnou nebo rizikovější léčbou má být nabídnuta alternativa, bdí nezávislá „Data and Safety Monitoring Board“. Důvodem k ukončení studie by rozhodně nemělo být zlepšení času dogrese do onemocnění, neboť v mnoha dříve provedených studiích bylo prokázáno prodloužení času dogrese nemoci (čas bez známek onemocnění), aniž by bylo jakkoliv ovlivněno celkové přežívání (např. ve studii GOG 132 a dalších). Tento rozpor je dán především existencí účinné léčby druhé a další linie, která umožňuje u relabujících pacientů opět získat kontrolu nad nádorovou nemocí. A tak v době, kdy se snažíme praktikovat medicínu založenou na důkazech, by neměla být standardem pouhá víra ve skutečnost, že prodloužení času dogrese nebo času bez známek onemocnění bude jistě doprovázeno i prodloužením přežívání. Z takového stavu nemůže mít prospěch ani pacient s lékařem, kteří nevědí, jak postupovat co nejeefektivněji, tak ani výrobce léčiva, který s neúplnými daty obtížně přesvědčí váhající lékaře o správnosti nasazení svého produktu.

Čteme-li později kompletní sdělení o randomizované studii, musíme velmi opatrně interpretovat závěry studií, abychom zjistili, zda je výsledek statisticky významný či nikoliv. V případě, že je studie postavena jako tzv. non-inferiority studie (např. kapecitabin vs FUFA Mayo v adjuvantní léčbě karcinomu tračníku), pak nesignifikantní p-value naznačují horší účinnost experimentálního postupu, zatímco signifikantní p-value (tj. pod 0,05 při tradičně užívané hladině významnosti či 0,025 jak to doporučují FDA guidelines v případě jednostranných testů) ukazují, že experimentální postup je „srovnatelný či lepší“ než standardní. Z takto postavené studie nejsme nikdy schopni prokázat přednost ramene experimentálního před standardním. Naproti tomu při designu studujícím superioritu režimu A před B nemůžeme nikdy prokázat, že experimentální postup je horší než standardní.

Poměrně často se také setkáváme s pojmem nesignifikantní tr-

end. Za tímto pojmem se skrývá zcela nesignifikantní difference, tj. difference která nebyla zatím nijak statisticky prokázána a nemá tedy žádný statistický význam. Ale zvláště v prezentacích farmaceutických firem se tento trend používá k přesvědčení lékařů o vhodnosti právě jejich produktu.

V případě použití konfidenčních intervalů pro poměr hazardů mezi rameny A a B je třeba upozornit na skutečnost, že pokud konfidenční interval obsahuje hodnotu/bod 1,00 je to vždy známkou nesignifikantního výsledku, byť je střed uvedeného konfidenčního intervalu pro poměr hazardů sebevíce vychýlen na jednu či druhou stranu.



Obrázek 2.

Pozornost si jistě zaslouží také referování relativních rizik a poměru hazardů. Relativní snížení rizika recidiv o polovinu je sice zajímavý údaj, ale kolik nemocných se opravdu vyvaruje lepším postupem recidivy? Relativní riziko 0,5 v sobě může skrývat jak vysoce efektivní postup - např. redukcí recidiv 30% vs 60%, tak i něco velmi nepatrného (např. rozdíl 0,003% vs 0,006%). Proto je třeba vždy zkontrolovat absolutní rozdíly sledovaných parametrů mezi jednotlivými rameny. Aby problémů nebylo málo, hned narážíme na další, který je ale bohužel moderními statistickými postupy již vyřešen. Je-li primárním cílem např. čas bez známek recidivy, čas do lokálního selhání nebo čas bez známekgrese onemocnění, je třeba disponovat statistickým nástrojem, který eliminuje chybu vzniklou tím, že nemocný např. podlehně jině souběžné nemoci nebo recidivuje ve vzdálených lokalizacích, zatímco my hodnotíme lokální kontrolu (např. u radioterapie). Ne vždy je ale optimální statistický postup skutečně použit a čtenář-lékař to ani nemusí poznat.

Za určitých okolností bývají publikovány také údaje o předem neplánovaných analýzách. Ty by podle zásad správné klinické praxe neměly být vůbec zveřejňovány, neboť právě proto, že nejsou předem plánovány, mohou obsahovat některé potenciálně zavádějící údaje - např. heterogenitu ve věku, pohlaví, charakteristikách nemoci, výkonnostním stavu apod. Výsledky těchto analýz je tedy třeba považovat jen za orientační.

Často řeší jeden problém velké množství klinických studií - např. chemoterapii kolorektálního karcinomu, radioterapii karcinomu prostaty apod. Proto vznikají systematická review, tzv. metaanalýzy, která se snaží zmenšit heterogenitu mezi jednotlivými studiemi a učinit závěry lépe využitelné v klinické praxi. V zásadě se setkáváme se dvěma typy metaana-

lýz. Metaanalýza typu „literature search“ využívá k získání závěru definic studovaných populací (včetně přesných inclusion/exclusion kritérií) hodnot p-value, konfidenčních intervalů a dalších proměnných, které jsou publikovány ve sbornících abstrakt, na přednáškách či v odborných časopisech. Ovšem hodnotnější a zcela nejvýše na pomyslném žebříčku kvality poskytovaných údajů stojí metaanalýza prováděná ze zdrojových dat. Při práci na této analýze oslovuje statistický tým autory jednotlivých studií, sjednocuje metodiku statistického zpracování a získává aktualizovaná data o přežívání, počtu léčebných odpovědí apod. Ne vždy jsou ale autoři jednotlivých studií ochotni spolupracovat na této metaanalýze prováděné ze zdrojových dat.

Pokud se tedy rozhodujeme o volbě léčebného postupu na základě medicíny založené na důkazech, pak nejprve zohledníme výsledky metanalýz, poté rozsáhlých randomizovaných studií fáze III, a až nakonec studií fáze II. Pokud neexistuje standardní postup, zvažujeme zařazení nemocných do klinických studií, jsou-li k dispozici.

Problémy s publikací výsledků

Při výčtu problémů, které se vyskytují při práci s klinickými studiemi, nelze opomenout ani tak zdánlivě jednoduché téma, kterým je publikace výsledků.

Bylo zjištěno, že větší množství studií s pozitivními výsledky je otiskováno v anglicky psané literatuře, a naproti tomu řada hodnotných prací z evropské či východoasijské oblasti je otištěna pouze v mateřském jazyce v případě, že experimentální postup není účinnější než postup standardní.

Také skutečnost, zda studie byla publikována pouze formou abstraktu nebo v plném znění je zavádějícím faktorem. Nedávno zveřejněná analýza randomizovaných studií v pediatrii prokázala, že pouze 59,1% abstraktů bylo následně publikováno *in extenso*. Častěji byly kompletně publikovány studie s větším počtem

nemocných, s kvalitnějším designem a s pozitivními výsledky⁶⁾. Takto může nepublikování některých podstatných zjištění z mnoha klinických studií vést ke snížení kvality léčebně-preventivních rozhodnutí pro dané skupiny nemocných, což ve svém konečném důsledku může ohrozit život konkrétního pacienta.

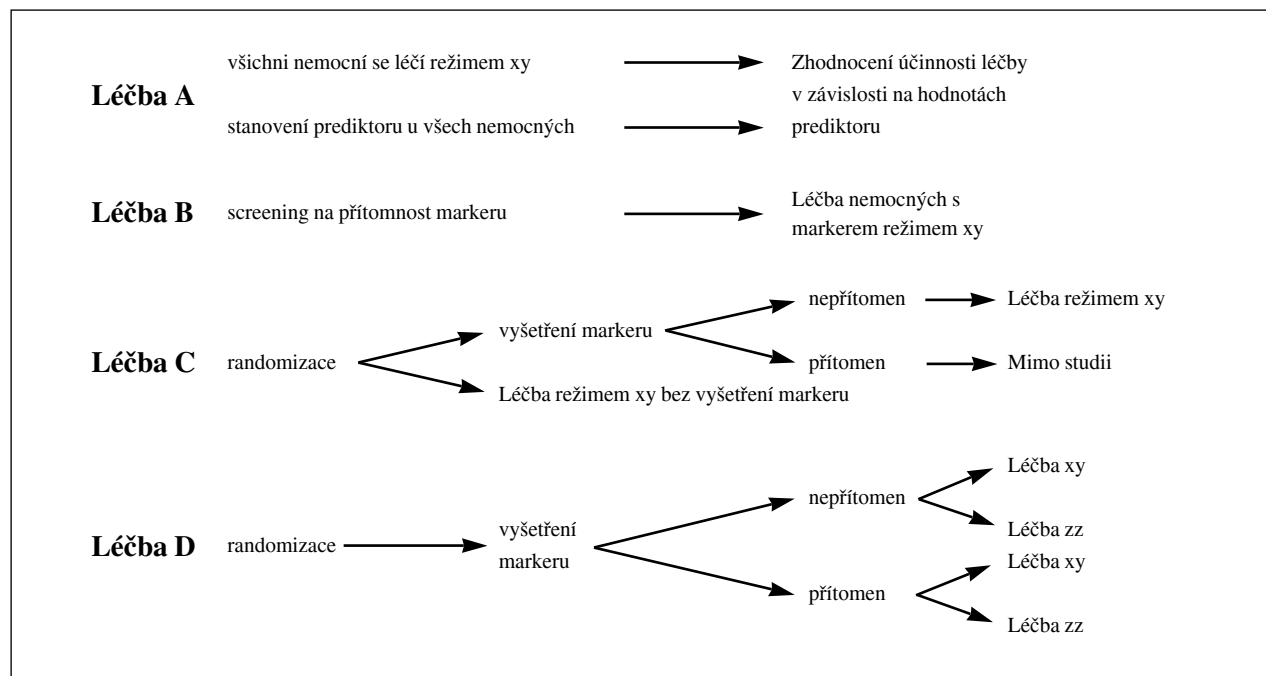
Větší pozornost by také měla být věnována přesnosti citací při přípravě publikací. Jejich kontrola provedená ve čtyřech velkých otorhynolaryngologických časopisech ukázala na přítomnost chyb u 37,5% všech citací; chyby hodnocené jako závažné byly odhaleny u 11,9% všech citací⁷⁾. S nesprávnými, a někdy i neexistujícími citacemi, se často setkáváme v propagačních materiálech farmaceutických firem.

V poslední době se lze setkat v publikační praxi nejprestižnějších časopisů s tzv. „zrychleným uveřejněním klíčových sdělení“. V podstatě se jedná o přednostní uveřejnění studie, kterou autor, editor nebo recenzenti považují za velmi významnou a mající dopad na každodenní klinickou praxi. Problém této politiky spočívá v tom, jak vlastně hodnotit, která práce si zrychlené vydání zaslouží, a která nikoliv⁸⁾. Ghalli publikoval v roce 2002 výsledky průzkumu mezi lékaři, který ukázal, že do určité míry opravdu mají přednostně vydávané články vyšší váhu, na druhou stranu 5 z 12 kontrolních článků vydaných standardním způsobem mělo v očích lékařů vyšší hodnocení než články „urychleně“ vydané.

Závěr

Klinické studie prováděné v posledním desetiletí umožnily ustanovit nové, efektivnější standardy protinádorové léčby, které vedou k vyššímu počtu vyléčených pacientů. Současně ale také vyvolávají nové otázky, na které dnes neznáme odpověď, a s ohledem na biologii nádorových procesů ani dlouho znát nebudeme.

Nezbývá nám tedy, než při léčbě našich pacientů zohlednit kromě studií také vlastní zkušenost a intuici.



Obrázek 1.: Ukázky některých designu studií, které hodnotí klinickou využitelnost prediktivního markeru. Schéma A dokumentuje léčbu všech pacientů jedním režimem se simultánním stanovením prediktoru a zhodnocením výsledků léčby v závislosti na hodnotách prediktoru. Pokud známe z historických souborů účinnost nepredikované léčby „xy“, pak při designu B můžeme nalézt vyšší účinnost léčby „xy“, pokud ji podáváme pouze pacientům s potenciálně pozitivním prediktivním faktorem. Samozřejmě, že je zde silná možnost zkreslení (bias) rozdíly mezi zkoumanou populací a historickými soubory (ve stádiu onemocnění, věku, zastoupení pohlaví a mnoha a mnoha dalších faktorech včetně různých vedlejších onemocnění). Design C umožňuje v randomizovaném porovnání zjistit, zda prediktivní test zlepšuje ve srovnání s neselektovanou populací léčebné výsledky postupem „xy“. Ve studii s designem D hledáme účinnější režim pro nemocné se studovaným znakem, resp. bez něj.

Literatura

1. Altman D.G., Schulz K.F., Moher D., et al.: The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Int Med* 134, 2001, 663- 694
2. Sullivan R.: Clinical trial design in oncology: protocol design. *Lancet Oncology* 5, 2004, 759-763
3. Charlson M.E., Horwitz R.I.: Applying results of randomized trials to clinical practice: impact of losses before randomisation *BMJ* 289, 1984, 1281-1284
4. Costa L.J., Xavier A.C., de Giglio A.: Negative results in cancer clinical trials - equivalence or poor accrual? *Controlled Clinical Trials* 25, 2004, 525-533.
5. *JCO* 2004, 22, 1542-1545
6. Klassen TP, Wiebe N, Russell K, et al.: Abstracts of randomized controlled trials presented at the society for pediatric research meeting: an example of publication bias. *Arch Pediatr Adol Med* 156, 2002, 474-479
7. Fenton J.E., Brazier H., De Souza A., et al.: The accuracy of citation and quotation in otolaryngology/head and neck journals. *Clin Otolaryngol Allied Sci* 25, 2000, 40-44
8. Ghali W.A., Cornuz J., McAlister F.A., et al.: Accelerated publication versus usual publication in 2 leading medical journals. *CMAJ* 166, 2002, 1137-1143