

ZÁKLADNÍ POJMY V DESIGNU, ANALÝZE DAT A INTERPRETACI VÝSLEDKŮ KLINICKÝCH HODNOCENÍ LÉČIV

DESIGN, DATA ANALYSES AND INTERPRETATION OF CLINICAL TRIALS-BASIC CONCEPTS

SVOBODNÍK A., COUFAL O., DUŠEK L.

CENTRUM BIOSTATISTIKY A ANALÝZ LF A PŘF MASARYKOVY UNIVERZITY V BRNĚ

Souhrn

Metodika designu a analýzy dat randomizovaných klinických studií představuje v současnosti samostatnou vědní disciplínu s vlastní terminologií. Vzhledem k tomu, že nauka o provádění a hodnocení výsledků klinických studií není součástí standardního vzdělávání lékařů, může docházet k desinterpretaci publikovaných dat. V textu jsou diskutovány základní pojmy z experimentálního designu a analýzy klinických studií.

Klíčová slova: klinické studie, randomizace, analýza dat, interim analýza, metaanalýza

Summary

Methodology of the design and analysis of clinical trials currently represents a substantive branch of knowledge with its own terminology. As courses on this subject are not incorporated in the standard educational process in medicine, misinterpretation of published results of clinical trials could occur. In this paper are discussed basic terms used in the design and analysis of randomized clinical trials.

Key words: clinical trials, randomization, data analysis, interim analysis, metaanalysis

ÚVOD

Randomizovaná klinická hodnocení (randomizované klinické studie) představují v současnosti nejlepší dostupnou technologii pro objektivní zhodnocení vlastností (především účinnosti a bezpečnosti) léčivých přípravků a zdravotnických prostředků. Metodika a zásady správného provádění klinických studií jsou dnes již podrobně zpracovány, definovány a standardizovány. Rychlý vývoj této oblasti v posledních desetiletích vedl k tomu, že se problematika designu a analýzy dat klinických studií poněkud vzdálila příbuzným oborům (medicína, farmakologie, matematika, statistika, informatika aj.) a komunikuje dnes vlastní terminologií a jazykem. Vzhledem k tomu, že nauka o designu a analýze dat klinických studií není součástí standardního vzdělávacího procesu uživatelů výsledků těchto experimentů (tedy především lékařů), může snadno docházet k desinterpretaci publikovaných dat.

Cílem tohoto textu je seznámit čtenáře s vybranými základními pojmy designu klinických studií. Tento text neumí nahradit úplné vzdělání v dané problematice a všem zájemcům o podrobnější informace je doporučeno samostudium (existuje velkého množství kvalitní a dostupné literatury) nebo návštěva specializovaných vzdělávacích kurzů u nás nebo v zahraničí.

ZÁKLADNÍ KLASIFIKACE KLINICKÝCH STUDIÍ

Existuje několik přístupů ke klasifikaci klinických studií, nejpoužívanější je třídění z hlediska pozice v procesu vývoje a registrace nových léčivých přípravků. Podle této klasifikace se experimenty dělí na prováděné v období *preklinického* vývoje (studie na zvířecích modelech) a na experimenty prováděné v období *klinického* vývoje (vlastní klinické studie), které se dále dělí na studie předregistrační (Fáze I-III) a poregistrační (Fáze IV). Samostatnou skupinu dále tvoří studie postmarketingové (neintervenční) a specifické léčebné programy.

Studie Fáze I

Hlavním cílem klinických studií Fáze I je stanovení základních humánních farmakokinetických parametrů nových léčivých přípravků a stanovení maximální tolerovatelné dávky (MTD). Vzhledem k tomu, že tyto studie navazují bezprostředně na experimenty prováděné na zvířecích modelech, jsou v těchto projektech podrobně sledovány a analyzovány nežádoucí účinky přípravků, důraz je tedy kladen primárně na bezpečnost léčby. Do klinických hodnocení Fáze I je zařazován relativně malý počet subjektů hodnocení (zpravidla 12-20), většinou se jedná o zdravé dobrovolníky, výjimkou je testování toxických léčivých přípravků (cytostatika apod.), v těchto případech jsou zařazováni pacienti. Tyto experimenty jsou zpravidla uspořádány tak, aby výsledná data umožňovala matematické modelování vztahu „dávka-odpověď“ a odhad efektu různých koncentrací látky na lidský organismus.

Studie Fáze II

Klinické studie Fáze II navazují na provedené a vyhodnocené studie Fáze I a jejich hlavním cílem je zjištění skutečné účinnosti hodnoceného přípravku za účelem opodstatnění jeho dalšího testování v experimentech pokročilejší fáze (studie Fáze III). Dalším cílem studií Fáze II je souběžné rozšířené hodnocení tolerance a bezpečnosti sledované léčby. Do těchto experimentů bývá zpravidla zařazováno 20-200 subjektů hodnocení. Jedná se většinou o jednoramenné experimenty, kde zjištěná účinnost a bezpečnost je srovnávána s literárně dostupnými údaji o srovnatelných přípravcích nebo současně nejlepší dostupné a používané léčbě.

Studie Fáze III

Klinické studie Fáze III představují zřejmě nejznámější a nejcharakterističtější typ klinického hodnocení. Jedná se o experimenty, ve kterých je nejčastěji srovnávána účinnost a bezpečnost dvou a více léčebných postupů v rámci jednoho randomizovaného

experimentu. Cílem těchto studií je tedy srovnat vlastnosti nového léčivého přípravku s kontrolou, kterou může představovat placebo, nebo „aktivní“ kontrola, tedy alternativní léčebný postup nebo léčivý přípravek. Výsledky těchto studií jsou základními podklady pro schválení používání přípravku regulačními autoritami (např. FDA, SÚKL apod.). Do těchto experimentů bývá zpravidla zařazováno 100-1000 subjektů, počty se však mohou značně lišit v závislosti na designu studie a epidemiologii hodnoceného onemocnění, limitace mohou být dány praktickými možnostmi náběru subjektů. Studie Fáze III bývají nejčastěji uspořádány podle *paralelního* nebo *cross-over* designu. Při klasickém paralelním uspořádání platí, že při srovnání dvou a více ramen studie je základním prvkem designu randomizace, která přiřadí jednotlivé subjekty hodnocení do srovnávaných ramen studie a v těchto ramenech, reprezentujících vždy jeden jediný léčebný postup zůstávají subjekty hodnocení v průběhu celého experimentu. Naproti tomu při *cross-over* uspořádání, jsou subjekty hodnocení také randomizovány do jednotlivých ramen studie, tato ramena však představují pouze pořadí, ve kterém pacienti podstoupí dva nebo více léčebných schémat srovnávaných v daném experimentu. Vzhledem k tomu, že při tomto uspořádání je každý subjekt „sám sobě kontrolou“, je z hlediska statistické analýzy výhodnější a vyžaduje nižší počet subjektů ve srovnání s paralelním uspořádáním.

Studie Fáze IV

Klinické studie Fáze IV jsou prováděny v období po úspěšné registraci nových léčivých přípravků, tedy v době jejich běžného používání v klinické praxi. Hlavním cílem těchto projektů je potvrzení vlastností hodnocených přípravků známých z předchozích experimentů (tedy ze studií Fáze I-III) na širších populacích pacientů (bez výraznějšího omezení přísnými vstupními a vylučujícími kritérii) a za „reálných“ podmínek klinické praxe (na rozdíl od „skleníkového“ prostředí klinických studií Fáze I-III). Tyto studie vzhledem k vysokému počtu zařazených pacientů bývají často zaměřeny na detailní analýzu nežádoucích účinků léčby, především vzácnějších projevů, které nemusely být zachyceny v předchozích studiích.

ZASLEPENÍ

O *zaslepení* hovoříme v klinickém hodnocení tehdy, pokud je cílem experimentu srovnání dvou nebo více léčivých přípravků, a kdy v zájmu objektivního zhodnocení jejich účinnosti a bezpečnosti je subjekt hodnocení nebo i další účastníci studie neschopen na základě vlastností podávaného přípravku (způsob aplikace, vzhled, chuť, velikost apod.) rozlišit, který ze srovnávaných přípravků je aplikován. O *jednoduchém zaslepení* (*single blinding*) hovoříme tehdy, jestliže je zaslepena pouze jedna strana a to buď subjekt hodnocení nebo investigátor, *dvojité zaslepení* (*double blinding*) se používá tehdy, jestliže jsou zaslepeny subjekty hodnocení i investigátoři, termín *trojitě zaslepení* (*triple blinding*) se používá v případě, že je kromě subjektů hodnocení, a investigátorů zaslepen rovněž personál spravující data studie a konečně termín *čtyřnásobné zaslepení* (*quadruple blinding*) se používá tehdy, jestliže jsou osoby aplikující studijní medikaci odlišné od osob vyhodnocujících léčebný efekt a bezpečnost a tyto osoby jsou rovněž zaslepeny. Zaslepení považujeme za další prvek v designu klinického hodnocení zvyšující objektivnost a validitu výsledků, v některých případech je však samozřejmě nerealizovatelné (např. srovnání chirurgické léčby s jen medikamentózní).

RANDOMIZACE

Randomizací rozumíme proces náhodného (nebo pseudonáhodného) rozdělování subjektů hodnocení do dvou nebo více léčebných skupin srovnávaných v rámci klinického hodnocení. Použití randomizace je dnes zlatým standardem především při provádění srovnávacích klinických studií Fáze III, tedy v projektech bezprostředně předcházejících registraci nového léčivého přípravku. Hlavním cílem randomizace je zamezit subjektivnímu a selektivnímu zařazování konkrétních subjektů hodnocení do konkrétních léčebných skupin (ramen studie) zkoušejícím lékařem. Moderní mechanismy randomizace umožňují navíc zajistit rovnoměrný

(nebo jiný požadovaný) poměr počtu subjektů hodnocení ve srovnávaných ramenech i v průběhu studie (možnost provedení interim analýz) a zajistit rovnoměrnou distribuci vybraných prognostických faktorů. I přes zjevný přínos randomizace k validitě a objektivitě výsledků srovnávacích experimentů má tato technika i v současnosti některé odpůrce (etické důvody) a první řádně provedené randomizované klinické studie byly provedeny teprve po druhé světové válce, tedy z hlediska celkového vývoje metodiky v relativně nedávné době.

V současnosti nejčastěji používané randomizační techniky můžeme rozdělit podle vhodnosti jejich použití na skupinu technik *nepřípustných*, *méně vhodných* a *doporučených*.

Mezi *nepřípustné* randomizační techniky tak patří v minulosti často používané randomizace subjektů hodnocení na základě jejich *pořadového čísla* vstupu do studie (např. každý „lichý“ pacient zařazen do ramene A a každý „sudý“ pacient do ramene B), dále randomizace podle *iniciál pacienta* (např. pacienti se jmény začínajícími písmenem z první poloviny abecedy zařazení do ramene A a ostatní do ramene B), analogicky nepřijatelná je randomizace na základě *data narození* nebo *data vstupu* do studie. Nevhodnost těchto technik je dána především tím, že v případě otevřeného klinického hodnocení (pacient ani zkoušející lékař nejsou zaslepeni) může experimentátor ovlivnit zařazení konkrétního subjektu do konkrétní léčebné skupiny tím, že ještě před podpisem informovaného souhlasu ví, do které léčebné skupiny by byl pacient zařazen a účast v daném experimentu nemusí na základě této znalosti pacientovi vůbec nabídnout. V případě zaslepených studií je nevhodou těchto postupů nemožnost kontrolovat distribuci prognostických faktorů v ramenech studie a nemožnost zajistit požadovaný poměr počtu subjektů v léčebných skupinách.

Za méně vhodné randomizační techniky je považována např. tzv. *kompletní randomizace*, tedy technika, kterou si lze zjednodušeně představit tak, že po podpisu informovaného souhlasu pacienta je o jeho přiřazení do konkrétního ramene studie rozhodnuto pouze na základě hodů mincí (nebo tahu losu, použití náhodného generátoru čísel apod.). Tato technika je považována za méně vhodnou, protože sice znemožňuje „předvídat“ experimentátorem přiřazení konkrétního subjektu konkrétní léčebné skupině, na druhé straně však opět není možno kontrolovat dosažení požadovaného poměru počtu subjektů v ramenech studie v jejím průběhu a už vůbec není kontrolována distribuce prognostických faktorů. Mezi *doporučované* randomizační techniky dnes patří především *stratifikovaná permutační blokovaná* randomizace a *adaptivní* randomizace. Stratifikovaná randomizace vychází z principu, kdy jsou prospektivně definovány nejdůležitější prognostické faktory, které mohou u konkrétního subjektu zásadně ovlivnit účinnost a bezpečnost hodnoceného léčebného režimu (např. stadium onemocnění, věk, pohlaví aj.) a jsou popsány všechny teoreticky možné kombinace těchto faktorů u konkrétních jedinců, tzv. *strata* (tedy jedno stratum mohou např. představovat muži mladší 45 let s klinickým stadiem I). Randomizace je následně provedena samostatně v rámci všech těchto strat a je tak zajištěno, že ve všech ramenech jsou rovnoměrně zastoupeni pacienti s časným a pokročilým stadiem, že bude mezi rameny stejný podíl mužů a žen a že mezi rameny budou srovnatelně zastoupeny všechny věkové skupiny. Pro zajištění požadovaného poměru počtu subjektů v ramenech studie v jejím průběhu je v případě permutační blokované randomizace použito náhodně opakované sekvence pořadí jednotlivých ramen studie, v důsledku čehož je vždy po daném počtu subjektů (např. při velikosti bloku $n=4$) po každých čtyřech subjektech hodnocení zajištěn vyrovnaný počet pacientů ve všech ramenech. V případě adaptivních randomizačních technik se na rozdíl od výše uvedených postupů nevytváří prospektivní randomizační plán, ale randomizace každého nově zařazeného subjektu hodnocení se provádí až po vyhodnocení dosavadního průběhu experimentu a to buď z hlediska průběžně vyhodnocené účinnosti léčby, nebo distribuce prognostických faktorů.

Požadavek nutnosti zajištění kontroly distribuce prognostických faktorů v ramenech studie dokládají příklady následujících dvou onkologických studií. Ve studii autorů Twelves et al. 2005, ve kte-

ré byla srovnávána účinnost a bezpečnost léčby fluorouracilem a leucovorinem versus orálně podávaného fluoropyrimidin capcitabinu u pacientů s metastatickým kolorektálním karcinomem bylo prokázáno, že z pohledu DFS (disease free survival) a OS (overall survival) je léčba capcitabinem minimálně stejně účinná jako léčba fluorouracilem a leucovorinem. V této studii byly nicméně zjištěny rozdíly v hodnotách významných prognostických faktorů (hladina karcinoembryonálního antigenu a stupeň zasažení lymfatických uzlin) v ramenech studie a po adjustaci na tyto rozdíly byla prokázána vyšší účinnost capcitabinu z hlediska DFS i OS. Naproti tomu např. ve studii autorů Hurwitz et al. 2004, srovnávající účinnost a bezpečnost léčby irinotecanem, fluorouracilem, leucovorinem (IFL) a bevacizumabem ve srovnání s IFL a placebem byla randomizace pacientů stratifikována podle centra, vstupních hodnot ECOG performance status, lokalizace primárního nádoru a počtu metastaticky postižených míst. Díky tomuto postupu bylo dosaženo srovnatelného zastoupení pacientů s různou prognózou v ramenech studie a byla validně prokázána vyšší účinnost léčby při použití bevacizumabu.

INTERIM ANALÝZA

Interim analýzou v klinickém hodnocení léčiv rozumíme jednorázové nebo opakované testování primární hypotézy (tedy např. zda je experimentální léčba účinnější než kontrolní) před protokolem plánovaným řádným ukončením experimentu. Cílem interim analýzy je zvážit možnost předčasného ukončení studie v případě, že existuje vážný předpoklad, že další pokračování experimentu již nepřinese žádnou novou informaci. Hlavním cílem co nejčasnějšího ukončení klinického hodnocení je urychlit přístup pacientů k nové léčbě nebo naopak minimalizace počtu pacientů vystavených léčbě, která se neprokázala jako bezpečná. Nezanedbatelná je rovněž úspora finančních nákladů a úspora lidské práce související s pokračováním studie.

Z výše uvedených důvodů se zdá být logické provádět interim analýzy co nejčastěji a neomezovat se v počtu testování výsledků studie v jejím průběhu. Na druhé straně je však důležité si uvědomit, že opakovaným testováním primární hypotézy se zvyšuje pravděpodobnost průkazu falešně pozitivního výsledku, tedy potvrzení určité konkrétní léčby jako účinnější a bezpečnější i v případě, že tato léčba ve skutečnosti účinnější a bezpečnější není. Abychom si dokázali lépe představit tento problém, demonstroe ho jednoduchou analogií. Představme si, že budeme v rámci experimentu srovnávat účinnost dvou ve skutečnosti naprosto stejně účinných léčivých přípravků a naším primárním parametrem účinnosti bude podíl pacientů s dosaženou pozitivní léčebnou odpovědí. Tato situace je podobná příkladu, kdy si budeme házet vyváženou mincí, a vždy po určitém počtu hodů vyhodnotíme počet padnutí „lícové“ strany mince (reprezentující počet pacientů s dosaženou léčebnou odpovědí v jednom rameni) a počet padnutí „rubové“ strany mince (reprezentující počet pacientů s dosaženou léčebnou odpovědí ve druhém rameni studie). Představme si, že máme jako vlastníci patentu na syntézu určité molekuly zájem na tom, aby jako účinnější léčba bylo prokázáno rameno, které u nás představuje „lícovou“ stranu mince. Pokud bychom mohli kdykoliv v průběhu studie (pro nás analogicky po jakémkoliv počtu hodů mincí) ukončit experiment, házeli bychom mincí tak dlouho, dokud by se situace nevyvíjela v náš prospěch (tedy dokud by nepřevážil počet hodů, kdy padnul „líc“ nad počtem hodů, kdy padnul „rub“) a v tento okamžik bychom experiment ukončili a prohlásili, že naše léčba byla prokázána jako účinnější a bezpečnější. O tom, že v průběhu testování (házení mincí) nastanou opakovaně a čistě náhodou situace, kdy jednoznačně převažuje počet padnutí „lícové“ nebo „rubové“ strany mince se můžeme snadno přesvědčit (při dostatečné trpělivosti) opakováním dostatečného počtu hodů (sofistikovaněji je možné ověřit tento fakt počítačovou simulací třeba i na datech přežití).

Z důvodu výše uvedených protichůdných požadavků na průběžné hodnocení výsledků klinických studií (jednoznačné *pro* z důvodu možnosti ukončit studii pokud další průběh je již zbytečný a jednoznačné *proti* z důvodu možnosti manipulace

s výsledky) jsou v současnosti interim analýzy vysoce doporučovány, ale pouze při respektování určitých jednoznačně definovaných pravidel. Mezi tato pravidla patří např. povinnost plánovat každou interim analýzu ještě před vlastním zahájením experimentu, tedy již v protokolu studie. V tomto dokumentu musí být jednoznačně uvedeno kdy (tedy např. po jakém přesném počtu zařazených pacientů) bude interim analýza provedena a kolik je jich v průběhu celé studie plánováno. Zároveň je nutné modifikovat hladinu významnosti a právě podle počtu plánovaných interim analýz a nastavit si tak „přísnější“ kritéria pro průkaz lepší účinnosti či bezpečnosti hodnoceného přípravku při opakovaném testování primární hypotézy.

METAANALÝZA

Zřejmě každý, kdo sleduje v medicíně podrobněji určitou specifickou oblast, se již setkal s tím, že se stejným léčivým přípravkem bylo provedeno několik nezávislých, nicméně z hlediska designu, indikace a populace pacientů velmi podobných klinických studií, které však měly odlišný a někdy dokonce zcela protichůdný výsledek. Metaanalýzou klinických studií rozumíme metodiku sofistikovaného zobecnění výsledků dvou a více nezávisle provedených experimentů. Tato metodika je založena na definování vstupních a vylučujících kritérií pro jednotlivé klinické studie a na aplikaci speciálních statistických technik, jejichž cílem je analýza heterogenity výsledků jednotlivých experimentů, v případě průkazu heterogenity je tato interpretována a v případě potvrzení homogenity jsou výsledky zobecněny. Cílem metaanalýzy je přinést ucelenou a klinicky relevantní informaci na základě výsledků jednotlivých klinických studií, které mohou mít nejednotné závěry. Protože technologie metaanalýz je relativně složitá, není v tomto textu prostor zabývat se jí detailněji a uvedme si pouze několik málo příkladů, kdy může dojít k nežádoucím zkreslení výsledků metaanalýz. Jedním ze základních problémů, se kterými se při metaanalýzách musíme vyrovnat je problém tzv. „publication bias“. Jedná se o to, že pokud pro metaanalýzu použijeme pouze výsledky klinických studií publikovaných v odborném tisku, můžeme snadno dojít k z hlediska pravdivé skutečnosti zkresleným závěrům, protože nezahrneme provedené, ale nepublikované studie. Je prokázáno, že studie, kterých se zúčastnil relativně nízký počet subjektů hodnocení, a které měly negativní výsledek (např. neprokázaly jednoznačný benefit nové léčby), mají menší šanci být publikovány v odborném tisku a to i přesto, že pokud byly provedeny metodicky správně, reprezentují pravdivou informaci.

Dalším, praktickým problémem při provádění metaanalýz, je problém vícenásobné publikace výsledků z experimentu provedeného na identické populaci subjektů hodnocení, což je důsledek faktu, kdy autoři po provedení náročné studie mají tendenci připravit z výsledků co nejvíce publikací často nazvaných a koncipovaných tak, aby nebylo jednoduché zjistit, že se jedná o tentýž projekt a tutéž populaci pacientů.

Zřejmě nejvýraznějším praktickým problémem metaanalýz je nejednotnost hodnocení účinnosti a bezpečnosti léčby v jednotlivých studiích, kdy např. přežití (celkové, doba do progresu apod.) může být v jedné studii měřeno od data diagnózy, v jiné od data randomizace nebo zahájení léčby apod., což představuje naprosto odlišné a obtížně srovnatelné údaje.

VELIKOST VZORKU (POČET SUBJEKTŮ HODNOCENÍ)

Počet subjektů hodnocení zařazených do řádně designovaného klinického hodnocení není náhodný a je přesně stanoven statistikem na základě tzv. „power analýzy“. Pokud je vzorek podhodnocen (do studie je tedy zařazeno méně subjektů než je třeba), je vysoce pravděpodobné, že daná studie nebude mít dostatečnou „sílu“ prokázat např. vyšší účinnost nebo bezpečnost nově hodnoceného přípravku i přes to, že tento přípravek skutečně účinnější a bezpečnější je. Naopak, pokud je do studie zařazeno více subjektů hodnocení než bylo potřeba, může se stát, že studie prokáže statisticky významně vyšší účinnost nebo bezpečnost nového přípravku, tento benefit však bude z hlediska klinického napros-

to bezvýznamný. Pro pochopení významu optimalizace velikosti vzorku v klinickém hodnocení je nutné správně interpretovat pojmy „klinicky významný rozdíl“ a „statisticky významný rozdíl“, které se někdy zaměňují ale ve skutečnosti znamenají diametrálně odlišné skutečnosti.

Klinicky významný rozdíl můžeme jednoduše definovat jako rozdíl v účinnosti a bezpečnosti nového přípravku oproti dosud standardně používanému, který je pro lékaře po zvážení všech ostatních okolností (např. bezpečnost, cena, náročnost aplikace apod.) dostatečný pro to, aby tento nový přípravek začal používat namísto dosud standardně používaného (pro zjednodušení zde neuvažujeme možnost individuálního přístupu k léčbě a možné rozdíly v benefitu léčby standardním a „novým“ přípravkem u různých pacientů). Klinicky významný rozdíl je tedy např. rozdíl v účinnosti (např. v podílu dosažení kompletních remisí nebo v době celkového přežití apod.) který definuje lékař na základě své expertní zkušenosti, bez ohledu na názor statistika. Klinicky významný rozdíl je tedy zcela nezávislý na provedení vlastního experimentu.

Naproti tomu **statisticky významný rozdíl** v účinnosti nebo bezpečnosti může být definován pouze na základě provedení experimentu, a pokud tvrdíme, že např. nová léčba je statisticky významně účinnější, znamená to, že daný experiment prokázal, že existuje pouze velmi malá pravděpodobnost, že tento nový preparát je ve skutečnosti stejně nebo méně účinný než kontrola, avšak bez ohledu na to, „o kolik“ tento nový přípravek účinnější je. Obecně platí, že pro průkaz většího rozdílu v účinnosti potřebujeme v experimentu méně subjektů (menší vzorek) než pro průkaz menšího rozdílu. Při zařazení extrémně vysokého počtu subjektů do teoretické studie srovnávající celkové přežití pacientů užívajících odlišný typ udržovací léčby se nám tedy může stát (příklady byly publikovány), že prokážeme statisticky významný rozdíl v účinnosti léčby, který bude představovat rozdíl v přežití pouze jeden den, tento rozdíl v účinnosti bude zcela jistě považován za klinicky nevýznamný.

Cílem optimalizace velikosti vzorku v klinickém hodnocení je zařadit do projektu takový počet subjektů hodnocení, který bude dostatečný k tomu, aby při zjištění klinicky významného rozdílu v účinnosti a/nebo bezpečnosti mezi rameny studie byl tento rozdíl prokázán rovněž jako statisticky významný, tedy aby bylo možné říci, že pravděpodobnost, že tento výsledek byl s ohledem na počet zařazených subjektů hodnocení dosažen pouze náhodou, je minimální.

Důvody, proč je nutné optimalizovat velikost vzorku v klinických studiích je možné shrnout do následujících: *etické aspekty* (je neetické vystavovat působení experimentálního přípravku více subjektů než je nezbytně nutné), *ekonomické aspekty* (každý další subjekt zařazený do klinického hodnocení zvyšuje jeho náklady) a *požadavek legislativních autorit* (před zahájením klinického hodnocení musí být regulačním autoritám doložena metodika výpočtu velikosti vzorku, která je součástí protokolu každé studie).

ITT A PP POPULACE

V publikovaných výsledcích klinických studií se často setkáváme s konstatováním, že prezentované výsledky vznikly analýzou dat

Intention to treat nebo *Per protocol* populace pacientů. Pro vysvětlení těchto termínů použijme příkladu jednoduché hypotetické studie, ve které budeme srovnávat účinnost dvou léčebných postupů, kdy jeden z nich bude představovat např. radikální chirurgický výkon zatěžující významně pacienta a druhá léčebná modalita bude pouze medikamentózní léčba s minimální zátěží pacientů. Představme si, že do studie bude zařazeno celkem 200 pacientů, kteří podepíší informovaný souhlas a budou rovnoměrně v poměru 1:1 zařazeni do obou ramen studie. Kriteříem pro zhodnocení účinnosti léčby bude v obou ramenech podíl pacientů s dosaženým určitým stupněm léčebné odpovědi. V rameni s medikamentózní léčbou bude tedy celkem 100 pacientů, z nichž všichni dodrželi předepsané dávkování a podstoupil všechny úkony dané protokolem studie, při závěrečném zhodnocení bude u 50 (50 %) zjištěno dosažení požadovaného stupně léčebné odpovědi. V rameni s chirurgickou léčbou však celkem 20 (20 %) pacientů poté, co se dozvědí, že byli randomizováni do této skupiny odmítne podstoupit operaci (např. z obavy z rizik výkonu jako takového), operovaných bude tedy pouze 80 (80 %) pacientů a z nich u 40 bude konstatováno dosažení požadovaného stupně léčebné odpovědi. Pro srovnání účinnosti léčby v obou ramenech studie je tedy zásadní, zda budeme úspěšnost léčby v rameni s chirurgickým zákrokem hodnotit jako 40 úspěchů z 80 operovaných (úspěšnost 50 % a tedy srovnatelná s druhým ramenem) a nebo jako 40 úspěchů ze 100 plánovaných operací (tedy 40 % a tedy nižší účinnost než v kontrolním rameni). Řešení tohoto problému není jednoduché, protože oba přístupy nám podávají rozdílnou a vždy relevantní informaci. Jeden přístup svědčí o úspěšnosti dané metody jako takové a druhý o skutečné účinnosti metody při jejím zavedení do klinické praxe (pokud 20 % pacientů odmítne podstoupit operaci v průběhu studie, stejný, nebo spíše vyšší podíl pacientů odmítne podstoupit tento výkon i v průběhu klinické praxe). V případě, že pro hodnocení účinnosti a bezpečnosti léčby použijeme pouze soubor pacientů, kteří dodrželi léčebný protokol (tedy např. podstoupili daný chirurgický výkon), hovoříme o *Per protocol* analýze, přístupu, nebo populaci pacientů. V případě, že analýzu provedeme na souboru pacientů plánovaných pro použití daného léčebného schématu bez ohledu na to, zda bylo aplikováno nebo ne a v jaké míře, hovoříme o *Intention to treat* analýze, přístupu, nebo populaci pacientů. Při analýze klinických studií je nutné uvádět vždy výsledky zjištěné pro *Per protocol* i *Intention to treat* populaci a v ideálním případě se tyto výsledky zásadně neliší. Ve výše uvedeném příkladu bylo v důsledku zjednodušení relativně snadné odlišit, který konkrétní subjekt by byl zařazen do *Intention to treat* nebo jen do *Per protocol* populace (podle toho, zda podstoupil nebo nepodstoupil operační výkon). V praxi může být toto rozhodování komplikovanější, jako příklad uveďme odchylku od plánované dávky chemoterapie např. z důvodu snížení dávky při projevech toxicity, kdy je nutné stanovit určitou toleranci pro celkovou dávku přípravku, kdy až po jejím překročení je subjekt považován vyloučen z *Per protocol* populace.

LITERATURA

1. Bahit, M., C., Cannon, C., P., Antman, E., M. et al.: Direct comparison of characteristics, treatment, and outcomes of patients enrolled versus patients not enrolled in clinical trial at centers participating in the TIMI 9 Trial and TIMI 9 registry. *Am Heart J*, 145(1):109-117; 2003.
2. Hurwitz, H., Fehrenbacher, L., Novotny, W., et al.: Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for the Treatment of Metastatic Colorectal Cancer. *N Engl J Med* 2004; 350:357-367.
3. Meinert, C., L.: *Clinical Trials: Design, Conduct, and Analysis*. Oxford University press, New York 1986.

4. Petitti, D., B.: *Meta-Analysis, Decision Analysis, and Cost- Effectiveness analysis*. Oxford Press, New York 2000.
5. Pocock, S., J.: *Clinical trials: A practical Approach*. John Wiley and Sons, New York 2000.
6. Rosenberger, W., F., Lachin, J., M.: *Randomization in clinical trials: Theory and practice*. John Wiley and Sons, New York 2002.
7. Spilker, B.: *Guide to Clinical Trials*. Raven Press, New York 1991.
8. Twelves, Ch., Wong, A., Nowacki, M., et al.: Capecitabine as Adjuvant Treatment for Stage III Colon Cancer. *N Engl J Med* 2005; 352:2696-704.
9. Weijer, Ch.: Are clinical trial participants comparable to non-participants? *Control Clin Trials*. 16:48S; 1995.