

Základy statistiky vědeckých prací v medicíně

Jiří Jarkovský

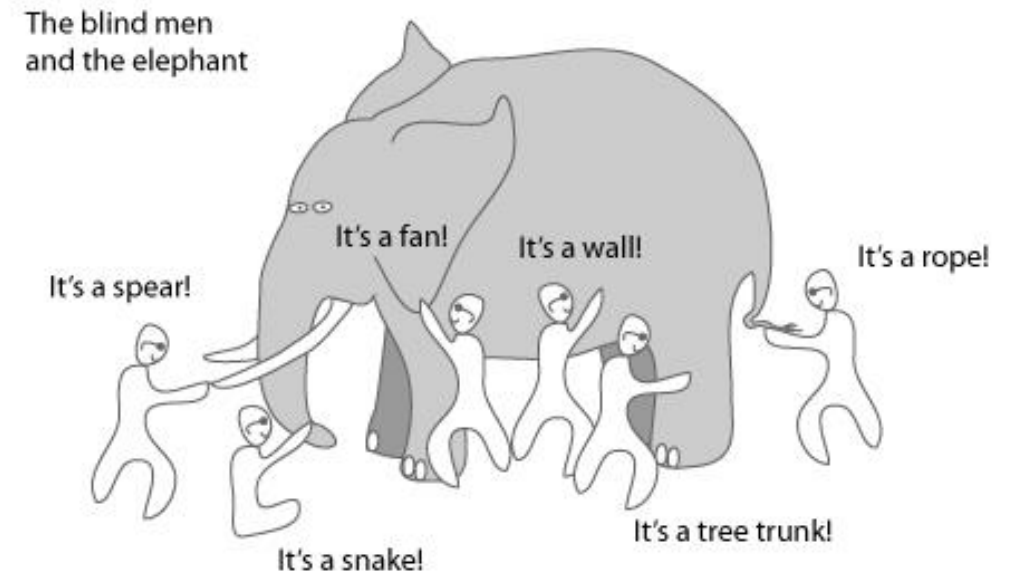
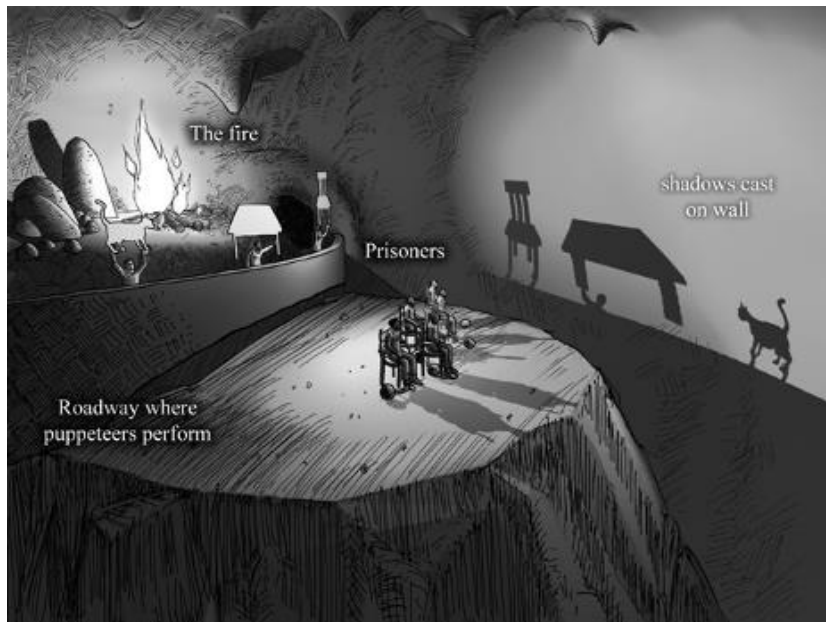
Institut biostatistiky a analýz, Lékařská fakulta, Masarykova univerzita

VÝZNAM STATISTICKÉ ANALÝZY DAT V KLINICKÉM VÝZKUMU

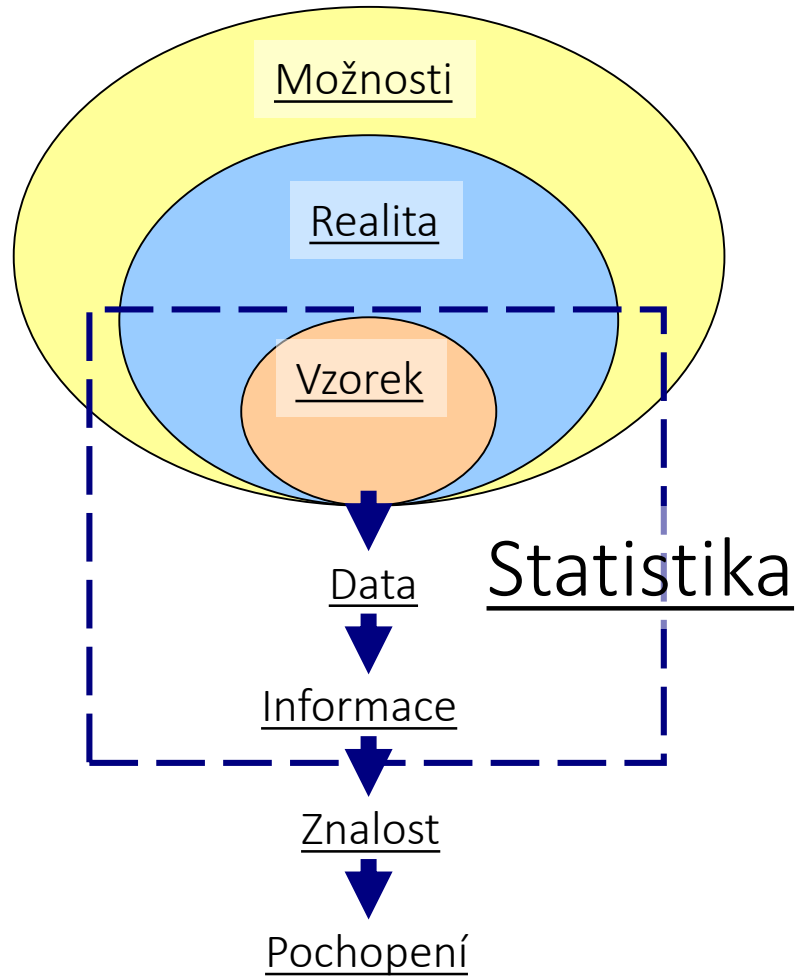
- Výzkum na základě sběru dat je naším způsobem porozumění realitě
- Ale jak přesné a pravdivé je naše porozumění?



Statistika je jedním z nástrojů vnášejících do našich výsledků určitou spolehlivost.
Statistiku můžeme považovat za ekvivalent k mikroskopu či jinému laboratornímu nástroji



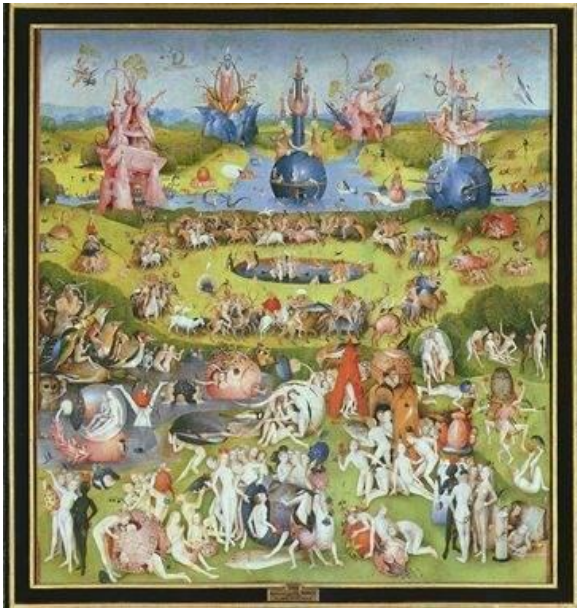
CO MŮŽE STATISTIKA ŘÍCI O NAŠÍ REALITĚ?



- Statistika není schopna činit závěry o jevech neobsažených v našem vzorku.
- Statistika je nasazena v procesu získání informací z vzorkovaných dat a je podporou v získání naší znalosti a pochopení problému.
- Statistika není náhradou naší inteligence !!!

VARIABILITA JAKO ZÁKLADNÍ POJEM VE STATISTICE

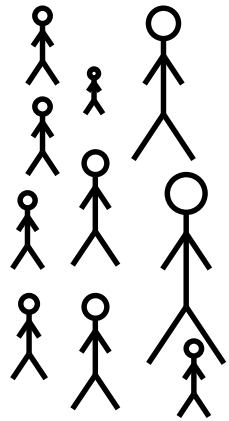
- Naše realita je variabilní a statistika je vědou zabývající se variabilitou
- Korektní analýza variabilita a její pochopení přináší užitečné informace o naší realitě
- V případě deterministického světa by statistická analýza nebyla potřebná



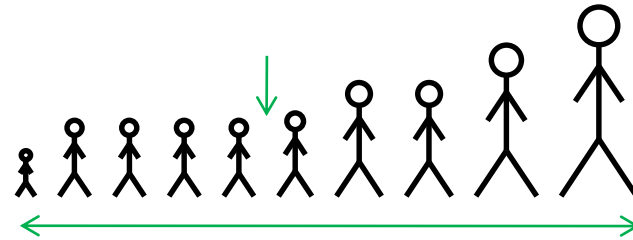
PRÁCE S VARIABILITOU V ANALÝZE DAT

- V analýze dat existují dva hlavní přístupy k práci s variabilitou

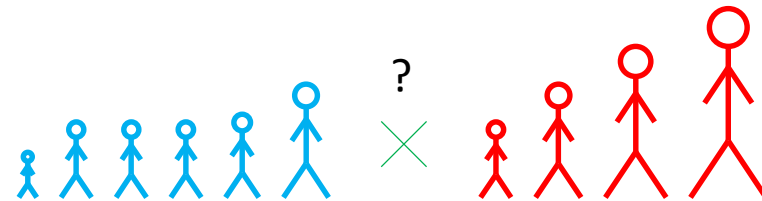
Variabilita dat



Popisná analýza: popis variability



Testování hypotéz: vysvětlení variability



TYPY STUDIÍ

- Dle fáze
 - Preklinické
 - Fáze I: Cílem je zjištění farmakologických charakteristik sledované látky u lidí, zejména zjištění maximální tolerované dávky.
 - Fáze II: Cílem je zjištění účinnosti na omezeném souboru jedinců a bezpečnosti v krátkodobém horizontu.
 - Fáze III: Cílem je srovnání účinnosti a bezpečnosti se standardní léčbou na jasně definované populaci pacientů.
 - Fáze IV: (post marketingová) Observační studie pro sledování morbidity a nežádoucích účinků.
- Dle času
 - Retrospektivní (vybíráme záznamy z databází)
 - Prospektivní (sledujeme budoucí vývoj, spadají sem běžné randomizované studie i observační kohortové studie)
- Dle cíle
 - hodnocení superiority (superiority trial) - Experimentální léčba je lepší než standardní (placebo)
 - hodnocení non-inferiority (non-inferiority trial) - Experimentální léčba není horší než standardní (placebo)
 - hodnocení ekvivalence (equivalence trial) - Experimentální léčba je „stejně dobrá“ jako standardní (placebo)
- Dle randomizace
 - Randomizované (zajišťujeme srovnatelnost skupin randomizací)
 - Observační (pouze sledujeme, srovnatelnost skupin zajišťují statistické metody; často RWE analýzy)
- Dle zaslepení
 - Single blind (neví pacient)
 - Double blind (neví lékař ani pacient)
 - Triple blind (neví lékař ani pacient ani statistik)

CO MUSÍME VĚDĚT PŘED ZAHÁJENÍM STUDIE?

- Cílová populace
 - Skupina pacientů na něž je studie zaměřena
- Primární hypotézy
 - Hlavní otázka položená ve studii – odhad velikosti vzorku a design studie je vypracován vzhledem k primární hypotéze
- Sekundární hypotézy
 - Vedlejší otázky, na něž by studie měla odpovědět
- Výběr adekvátní metodiky
 - Hypotézy jsou zodpovězeny prostřednictvím konkrétních proměnných (endpointů) – jejich typ (binární, kategoriální, spojité proměnné, přežití, mortalita atd.) určuje výběr způsobu statistického zpracování

FLOW-CHART – KLÍČOVÝ PRO UCHOPENÍ DESIGNU STUDIE

TRIAL POPULATION

Patients who were at least 18 years of age were enrolled if they could undergo randomization within 12 hours after having an acute ischemic stroke with a score of 3 or less on the National Institutes of Health Stroke Scale (NIHSS) (scores range from 0 to 42, with higher scores indicating greater stroke severity) or a high-risk TIA with a score of 4 or more on the ABCD² scale¹⁴ (which estimates the risk of recurrent stroke after a TIA on the basis of age, blood pressure, clinical features, duration of symptoms, and presence of diabetes; scores ranges from 0 to 7, with higher scores indicating a greater risk of stroke). They were also required to undergo computed tomography or magnetic resonance imaging to rule out intracranial bleeding or other conditions that could explain the neurologic symptoms or detect any contraindications to a trial treatment. Patients with TIA and minor,

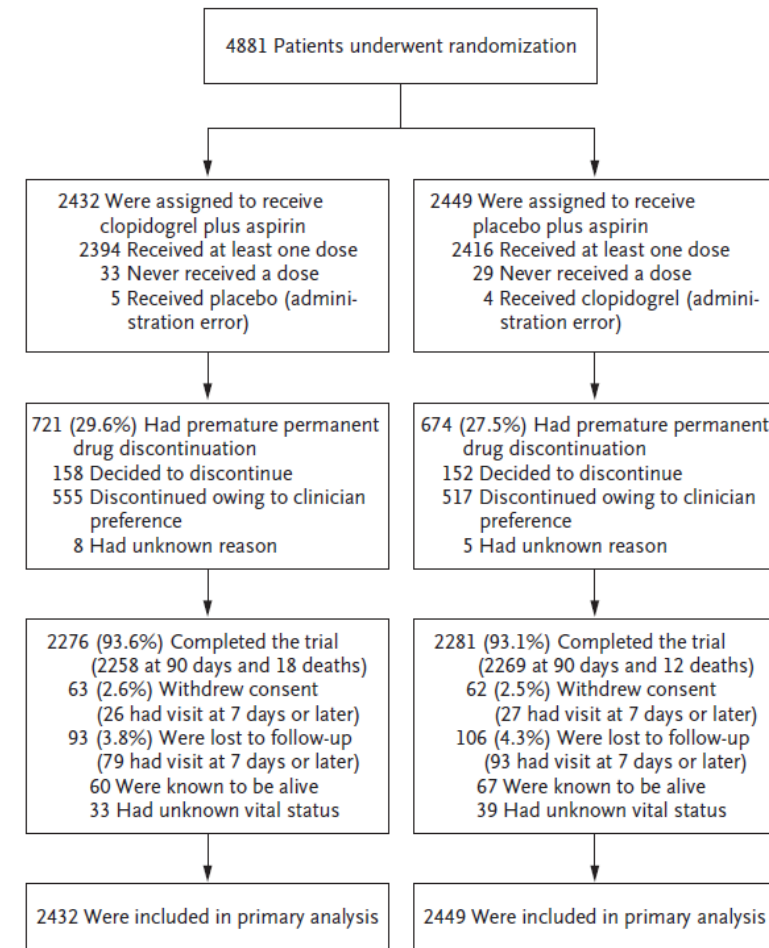
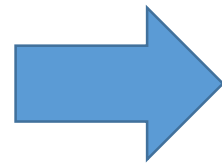


Figure 1. Enrollment and Outcomes (Intention-to-Treat Analysis).

Patients who discontinued a trial drug were included in the intention-to-treat analysis, as were patients who withdrew consent or were lost to follow-up. In the as-treated analysis, data for patients who received a trial drug were censored at the time of discontinuation.

TESTOVÁNÍ DODATEČNÝCH ENDPOINTŮ A SUBGROUP ANALÝZA

Testování dodatečných endpointů studie

- **Hodnocení statistické významnosti dalších rozdílů mezi léčebnými skupinami je lákavé, ale zavádějící!**
- Vede to k problému násobného testování hypotéz.
- Testování dalších proměnných je nutné brát s velkou rezervou – mělo by sloužit pouze jako podklad pro další hypotézy.

Subgroup analýza

- Vedle násobného testování hypotéz lze u analýz podskupin předpokládat ještě fenomén **nedostatečné síly testu**.
- Neprokázání statistické významnosti rozdílů v účinnosti mezi podskupinami neznamena, že ten rozdíl neexistuje! Pro korektní ověření by byla potřeba další studie.

Při změně protokolu v průběhu studie, dodatečných subgroup analýzách atd. nemají výsledky stejnou validitu jako původně plánovaná studie

KLÍČOVÉ PRINCIPY BIOSTATISTIKY

Jsme schopni odlišit výsledky studie od pouhé náhody?



*Co skutečně stojí za výsledkem studie?
„Jsou výsledky diktátu ovlivněny věkem,
výškou, hmotností nebo délkou školní
docházky dětí?“*

*Popisuje studie
reprezentativně populaci?
„Vypovídají batolata o
dětech jako celku?“*

*Jak moc se dá na
výsledky studie
spolehnout?*

*Dostaneme v případě
opakování (~ v praxi) s
dostatečnou spolehlivostí
obdobné výsledky?*

*Srovnáváme srovnatelné?
„Hodnotíme vliv počtu knih v domácnosti na
výsledky diktátu mezi skupinami dětí ve školce a v
9. třídě ZŠ – je to smysluplné srovnání?“*

BASELINE CHARACTERISTICS: KDO JSOU PACIENTI VE STUDII, SROVNATELNOST

Table 1. Baseline Characteristics of the Patients, According to Treatment Group.*

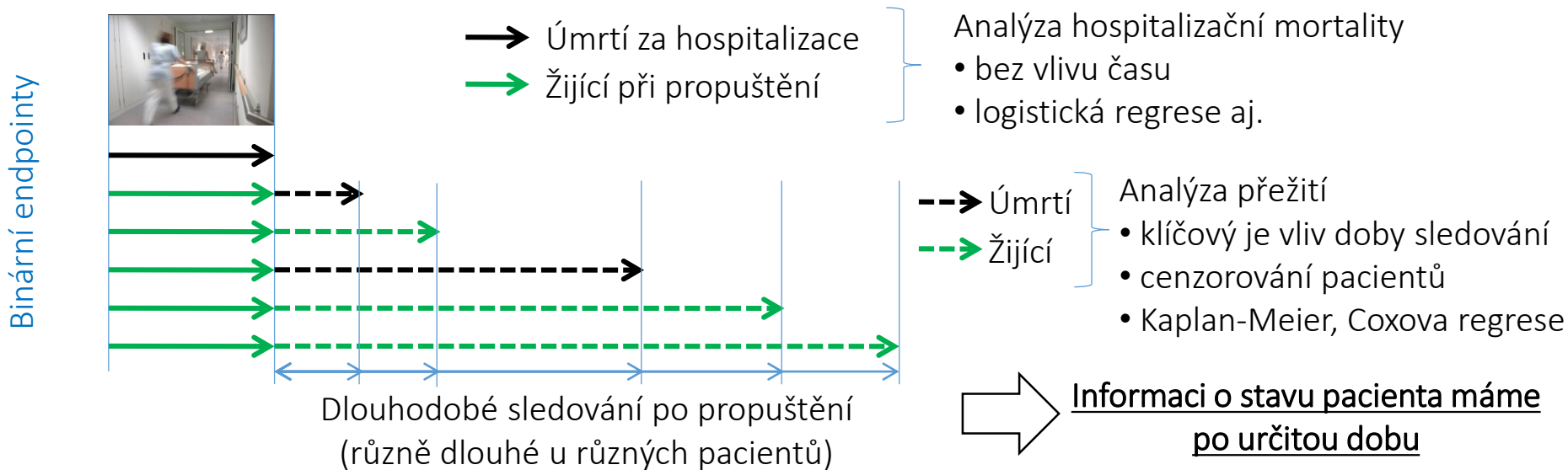
Characteristic	Ticagrelor Group	Clopidogrel Group
Median age — yr	62.0	62.0
Age ≥75 yr — no./total no. (%)	1396/9333 (15.0)	1482/9291 (16.0)
Female sex — no./total no. (%)	2655/9333 (28.4)	2633/9291 (28.3)
Median body weight — kg (range)	80.0 (28–174)	80.0 (29–180)
Body weight <60 kg — no./total no. (%)	652/9333 (7.0)	660/9291 (7.1)
BMI — median (range)†	27 (13–68)	27 (13–70)
Race — no./total no. (%)‡		
White	8566/9332 (91.8)	8511/9291 (91.6)
Black	115/9332 (1.2)	114/9291 (1.2)
Asian	542/9332 (5.8)	554/9291 (6.0)
Other	109/9332 (1.2)	112/9291 (1.2)
Cardiovascular risk factor — no./total no. (%)		
Habitual smoker	3360/9333 (36.0)	3318/9291 (35.7)
Hypertension	6139/9333 (65.8)	6044/9291 (65.1)
Dyslipidemia	4347/9333 (46.6)	4342/9291 (46.7)
Diabetes mellitus	2326/9333 (24.9)	2336/9291 (25.1)
Other medical history — no./total no. (%)		
MI	1900/9333 (20.4)	1924/9291 (20.7)
Percutaneous coronary intervention	1272/9333 (13.6)	1220/9291 (13.1)
Coronary-artery bypass grafting	532/9333 (5.7)	574/9291 (6.2)
Congestive heart failure	513/9333 (5.5)	537/9291 (5.8)
Nonhemorrhagic stroke	353/9333 (3.8)	369/9291 (4.0)
Peripheral arterial disease	566/9333 (6.1)	578/9291 (6.2)
Chronic renal disease	379/9333 (4.1)	406/9291 (4.4)
History of dyspnea	1412/9333 (15.1)	1358/9291 (14.6)
Chronic obstructive pulmonary disease	555/9333 (5.9)	530/9291 (5.7)
Asthma	267/9333 (2.9)	265/9291 (2.9)

Kromě popisu zařazovacích a vylučovacích kritérií a flow chartu studie je pro korektní interpretaci výsledků nezbytná i tabulka baseline charakteristik.

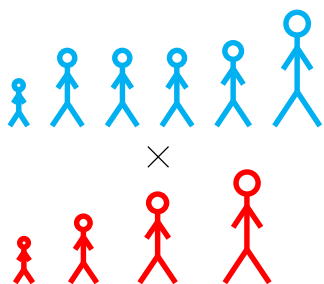
Umožňuje posoudit na jakou populaci je studie zaměřena a srovnatelnost ramen studie.

KLINICKÉ ENDPOINTY: ČÍM SE LIŠÍ JEJICH RŮZNÉ TYPY?

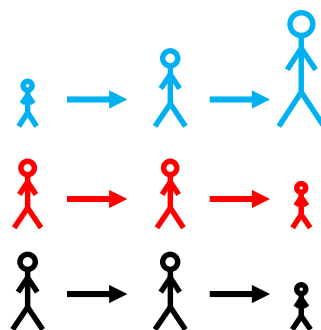
- Základní endpointy v studiích na klinických datech: mortalita, čas do události, kvantitativní ukazatele
- Rozdílný význam i metodika analýzy



Kvantitativní endpointy

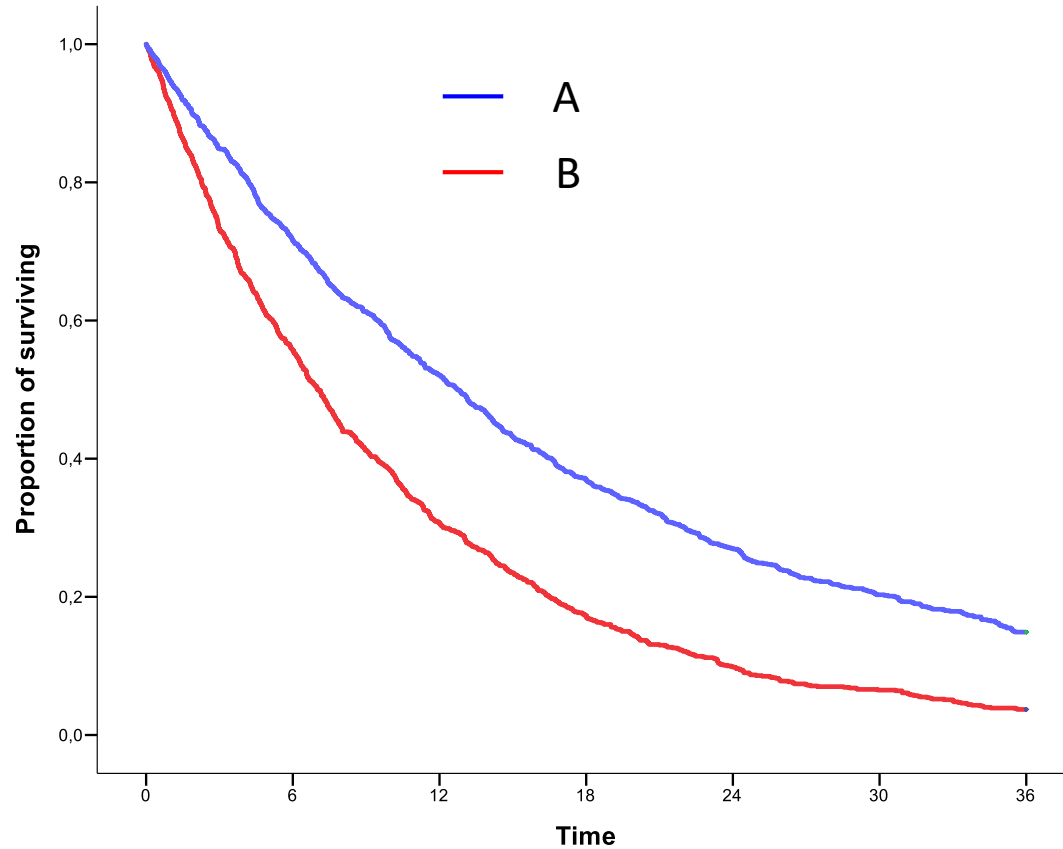


Testy srovnávající kvantitativní údaje nepárově mezi skupinami pacientů



Testy srovnávající kvantitativní údaje párově v rámci jednotlivých pacientů

ANALÝZA PŘEŽITÍ JAKO PRIMÁRNÍ CÍL KLINICKÝCH STUDIÍ



Survival at given time	A	B
6M	0.716 (0.688; 0.744)	0.557 (0.526; 0.588)
12M	0.521 (0.490; 0.552)	0.307 (0.278; 0.336)
18M	0.369 (0.339; 0.399)	0.173 (0.150; 0.196)
24M	0.270 (0.242; 0.298)	0.099 (0.080; 0.118)
30M	0.203 (0.178; 0.228)	0.065 (0.050; 0.080)
36M	0.149 (0.127; 0.171)	0.037 (0.025; 0.049)
Median survival	12.7 (11.6; 13.8)	7.0 (6.5; 7.6)

Coxův model proporciálních rizik: HR (95% CI)		
A vs. B	0.568 (0.518; 0.624)	p<0.001

*Různé charakteristiky
křivky přežití s různou
interpretací !!!*

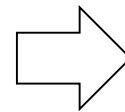
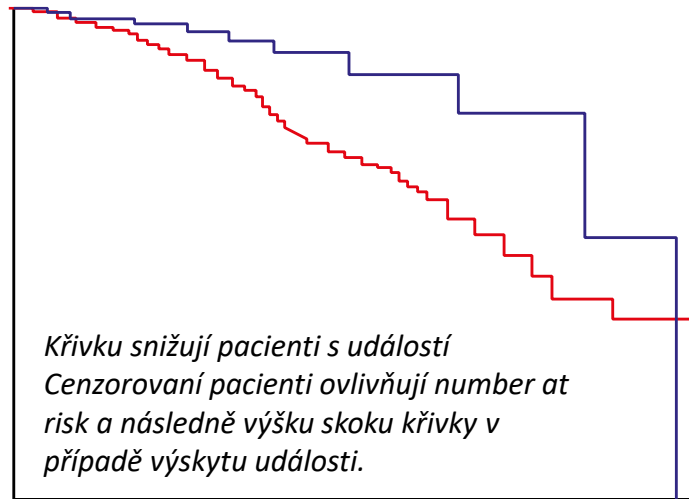
No at risk	0M	6M	12M	18M	24M	30M	36M
A	1000	716	521	369	270	203	149
B	1000	557	307	173	99	65	37

ANALÝZA PŘEŽITÍ

- Studujeme čas předcházející události, která nás zajímá ... T
- Na rozdíl od mortality nechceme pouze počty událostí
- **Událost** - smrt, progrese nemoci, relaps, ...
- **Počátek sledování**
 - Diagnóza, operace, začátek studie, narození, expozice faktoru
 - Musí být jasně časově definován, je většinou individuální
- **Konec sledování**
 - Konec studie, studia nebo grantu
 - Musí být opět jasně definován

NUMBER AT RISK: DŮVĚRYHODNOST KŘIVKY PŘEŽITÍ

- Počet pacientů v křivce přežití ubývá s časem, jednak výskytem události (úmrtí, progrese), jednak cenzorováním (po ukončení sledování žijícího pacienta již o něm dále nic nevíme)
- Výška „skoků“ v křivce přežití je závislá na počtu pacientů, kteří zůstávají v riziku, tedy počtu pacientů ve sledování → křivka je v každém čase podložena jiným počtem pacientů → různé části jsou jinak relevantní.

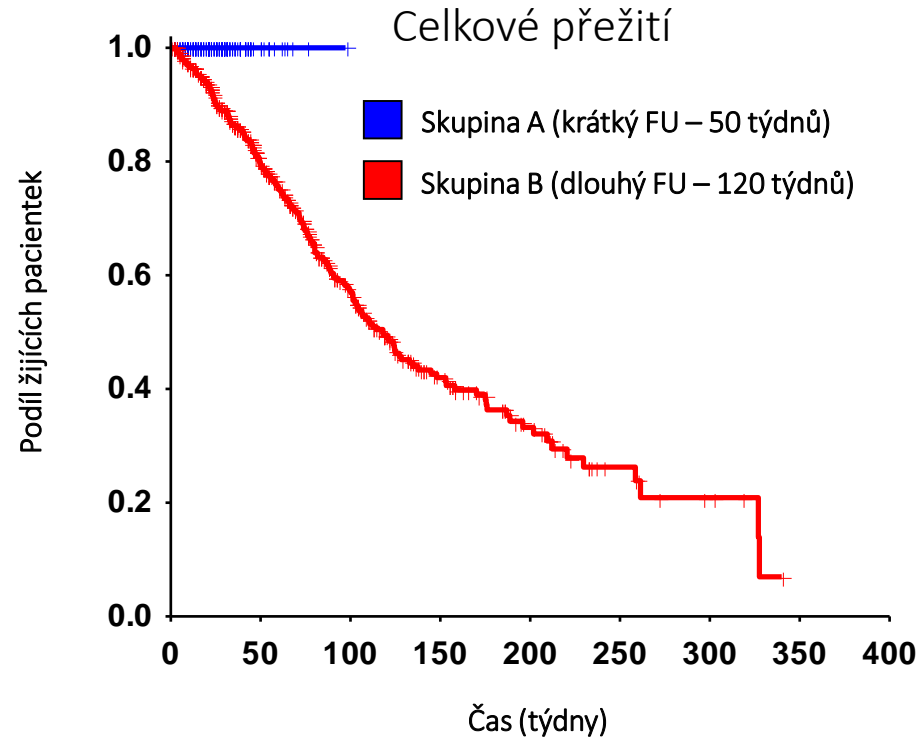


Je nutné mít na paměti nízkou věrohodnost „konce“ křivky přežití – zůstává-li ve studii 10 pacientů nebo méně, „skoky“ v přežití s každou další událostí jsou dramatické.

No at risk	T0	T1	T2	T3	T4	T5	T6
B	200	120	50	15	3	1	0
A	400	310	220	150	40	20	18

CENZOROVÁNÍ A DÉLKA SLEDOVÁNÍ PACIENTŮ

- Přítomnost cenzorování neumožňuje použít na data přežití klasické popisné a srovnávací metody (např. t test).
- Cenzorovaná pozorování nemůžeme vyhodit – obsahují informaci!
- Podíl cenzorovaných pozorování je měřítkem kvality studie a doby sledování pacientů

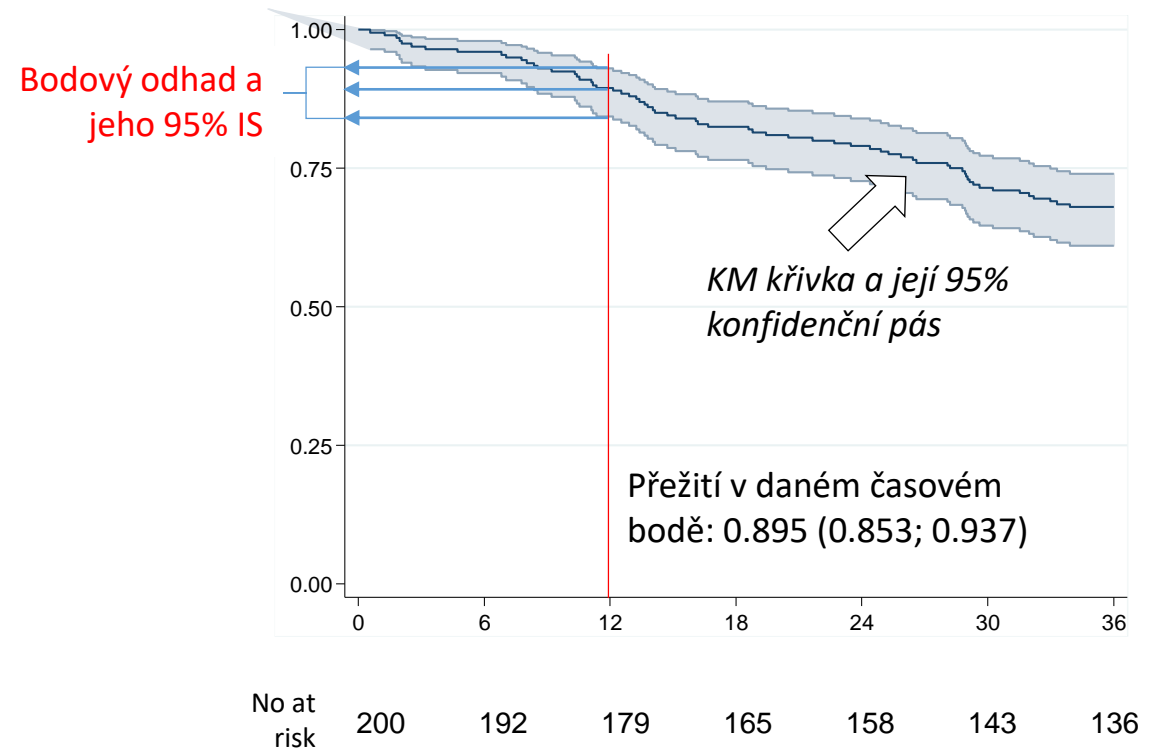
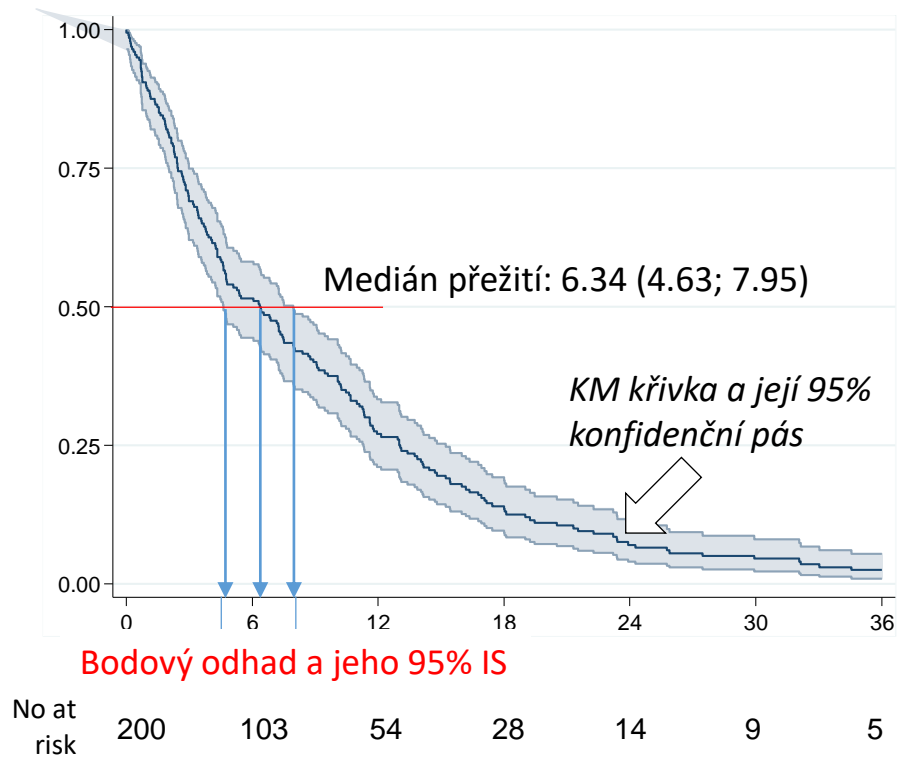


Ve studii musí být reportována doba sledování a podíl cenzorovaných pacientů.

Doba sledování musí být dostatečná pro projevení se sledované události (začleněno v designu studie).

POPIS KŘIVKY PŘEŽITÍ: MEDIÁN PŘEŽITÍ A PODÍL PŘEŽIVŠÍCH

- **Popis křivky přežití: zásadní měřítko efektivity léčby v dnešních klinických studiích**
 1. čas jehož se dožije dané procento pacientů (medián přežití = čas jehož se dožije 50% pac.)
 2. Podíl přeživších v daném čase (např. podíl přeživších v 1, 2, 3 atd. letech)
- Popis křivky přežití by měl být doplněn jeho intervalem spolehlivosti = informace o spolehlivosti a důvěryhodnosti výsledku



ROLE TESTOVÁNÍ HYPOTÉZ

Klinický výzkum je základem medicíny založené na důkazech.

```
graph TD; A[Klinický výzkum je základem medicíny založené na důkazech.] --> B[Biostatistika je dnes nedílnou součástí klinického výzkumu.]; B --> C[Testování hypotéz je základem biostatistiky.];
```

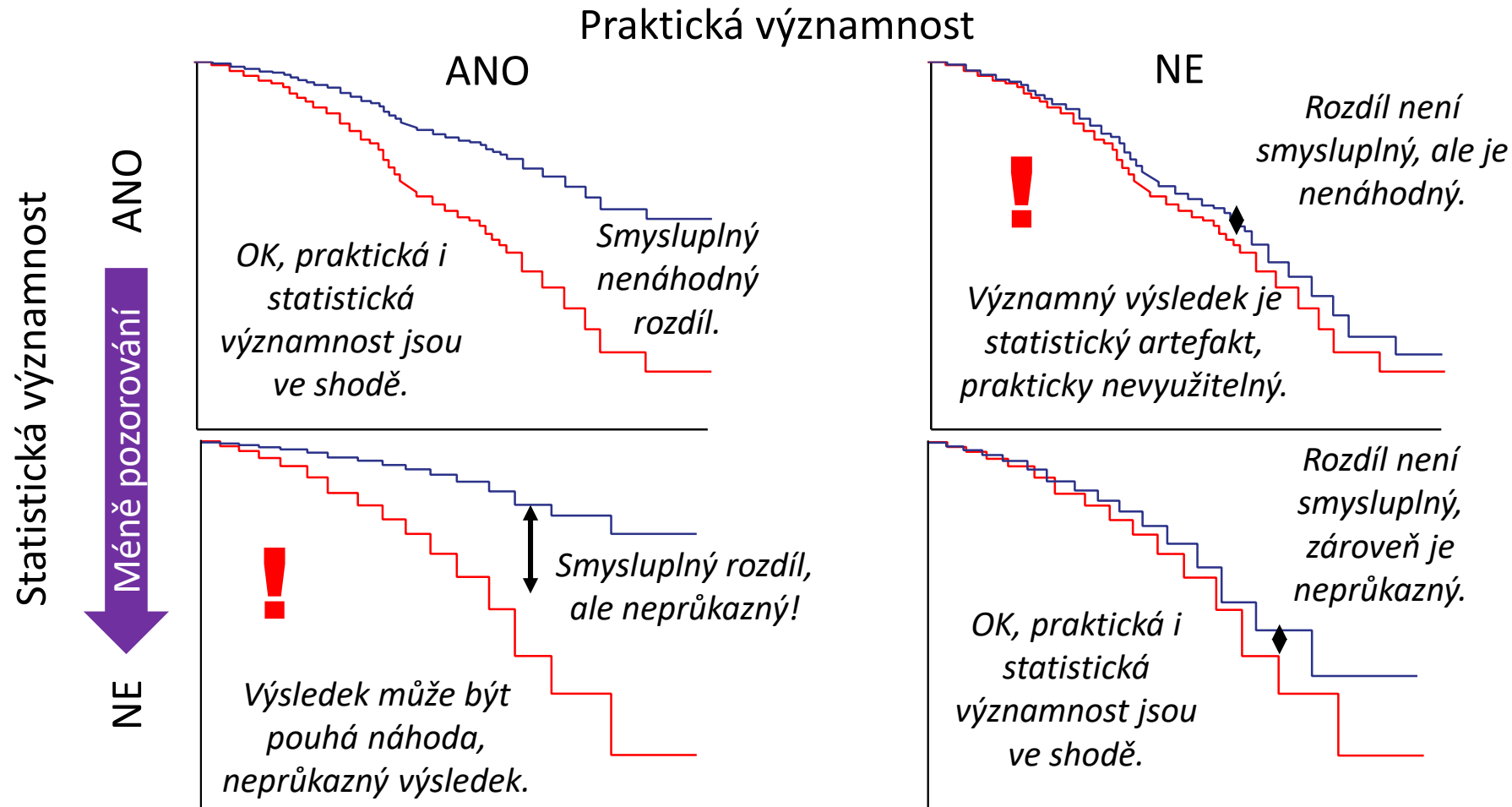
Biostatistika je dnes nedílnou součástí klinického výzkumu.

Testování hypotéz je základem biostatistiky.

KLÍČOVÉ PRINCIPY – VÝZNAMNOST

- Výsledky analýzy dat studie nejsou perfektním obrazem reality.
- Statistická významnost není totéž co kauzalita.
- Statistická významnost pouze indikuje, že pozorované rozdíly nebo vztahy nejsou náhodné (vzhledem k testované hypotéze).
- Praktická významnost, tedy významnost z pohledu lékaře nebo biologa je také významná.
- Statistická významnost je ovlivněna velikostí vzorku.

KLÍČOVÉ PRINCIPY – VÝZNAMNOST

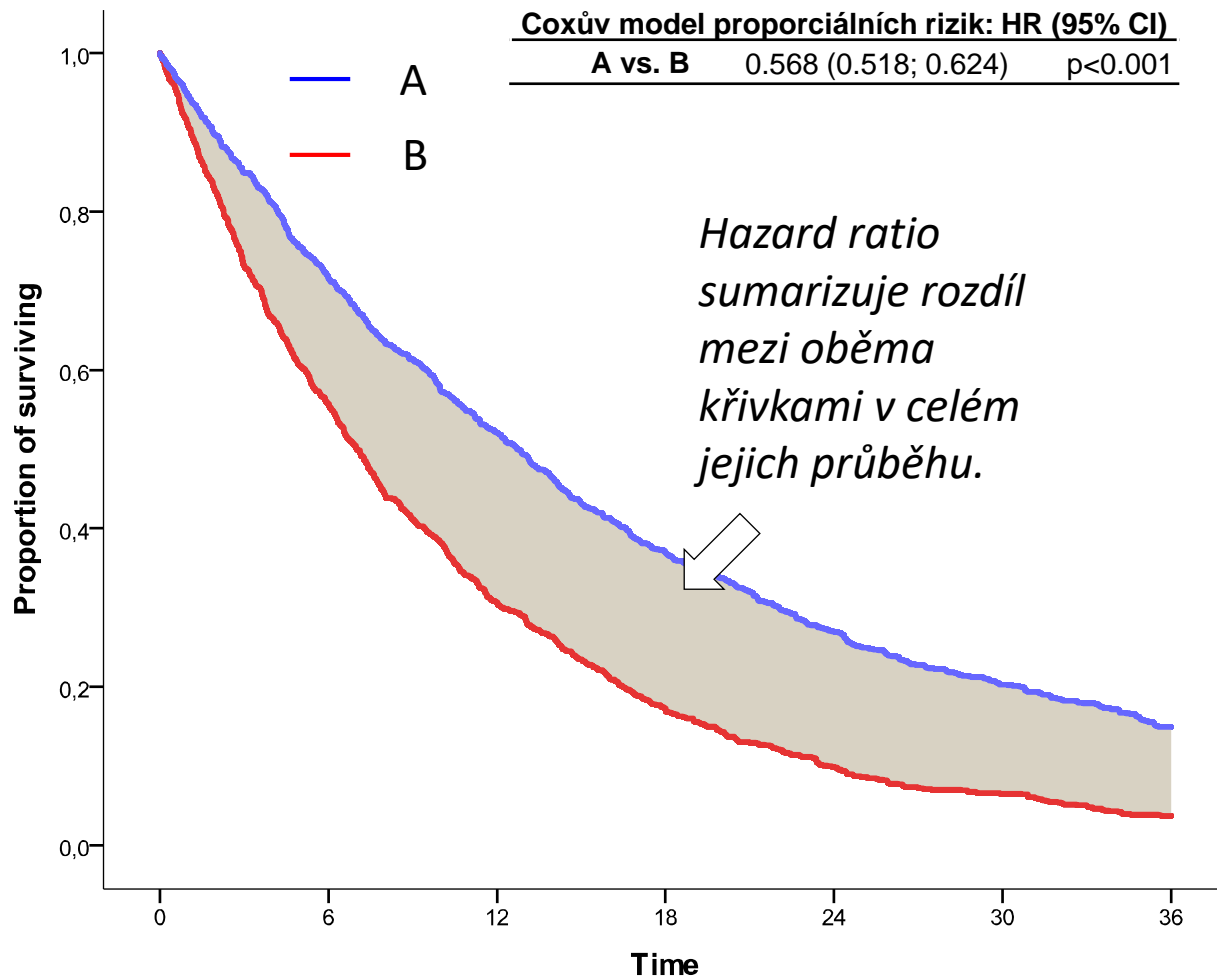


➡ Statisticky nevýznamný výsledek neznamena, že pozorovaný rozdíl ve skutečnosti neexistuje! Může to být způsobeno nedostatečnou informací v pozorovaných datech!

DŮLEŽITÉ POZNÁMKY K TESTOVÁNÍ HYPOTÉZ

- **Nezamítnutí nulové hypotézy neznamená automaticky její přijetí!** Může se jednat o situaci, kdy pro zamítnutí nulové hypotézy nemáme dostatečné množství informace.
- **Dosažená hladina významnosti testu (ať už 0,05, 0,01 nebo 0,10) nesmí být slepě brána jako hranice pro existenci/neexistenci testovaného efektu.** Neexistuje jasná hranice pro významnost či nevýznamnost – často je velmi malý rozdíl mezi p-hodnotou 0,04 a p-hodnotou 0,06.
- **Malá p-hodnota nemusí znamenat velký efekt.** Hodnota testové statistiky a odpovídající p-hodnota může být ovlivněna velkou velikostí vzorku a malou variabilitou pozorovaných dat.
- **Výsledky testování musí být nahlíženy kriticky** – jedná se o závěr založený „pouze“ na jednom výběrovém souboru.
- **Statistická významnost a predikční/klasifikační schopnost modelu nejsou totéž;** statistická významnost je pouze nutná, nikoliv dostačující podmínka pro dobrou predikci/klasifikaci pacientů.

HAZARD RATIO



- Je standardním postupem pro popis a testování významnosti celkového rozdílu mezi křivkami přežití (nebo obecně pro testování vlivu charakteristik pacientů na přežití)
- Spjato s tz. Coxovým modelem proporcionálních rizik
- Hodnoty
 - 1 = není žádný rozdíl
 - >1 = rizikový faktor
 - <1 = ochranný faktor
- Je vhodné jej doplnit intervalem spolehlivosti a statistickou významností
- V našem příkladu příslušnost pacienta ke skupině A snižuje jeho riziko úmrtí

HAZARD RATIO – PŘÍKLAD VÝSTUPU

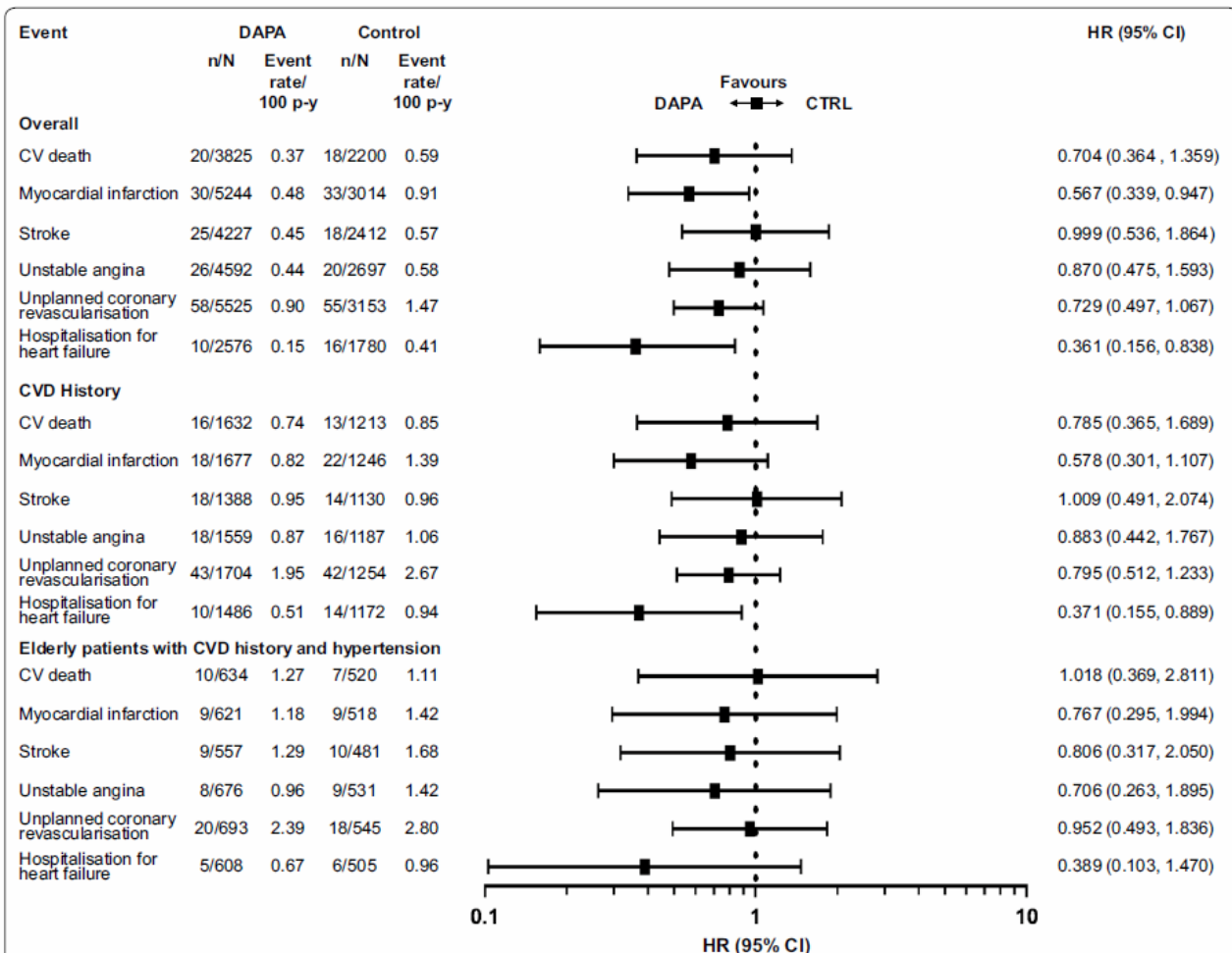
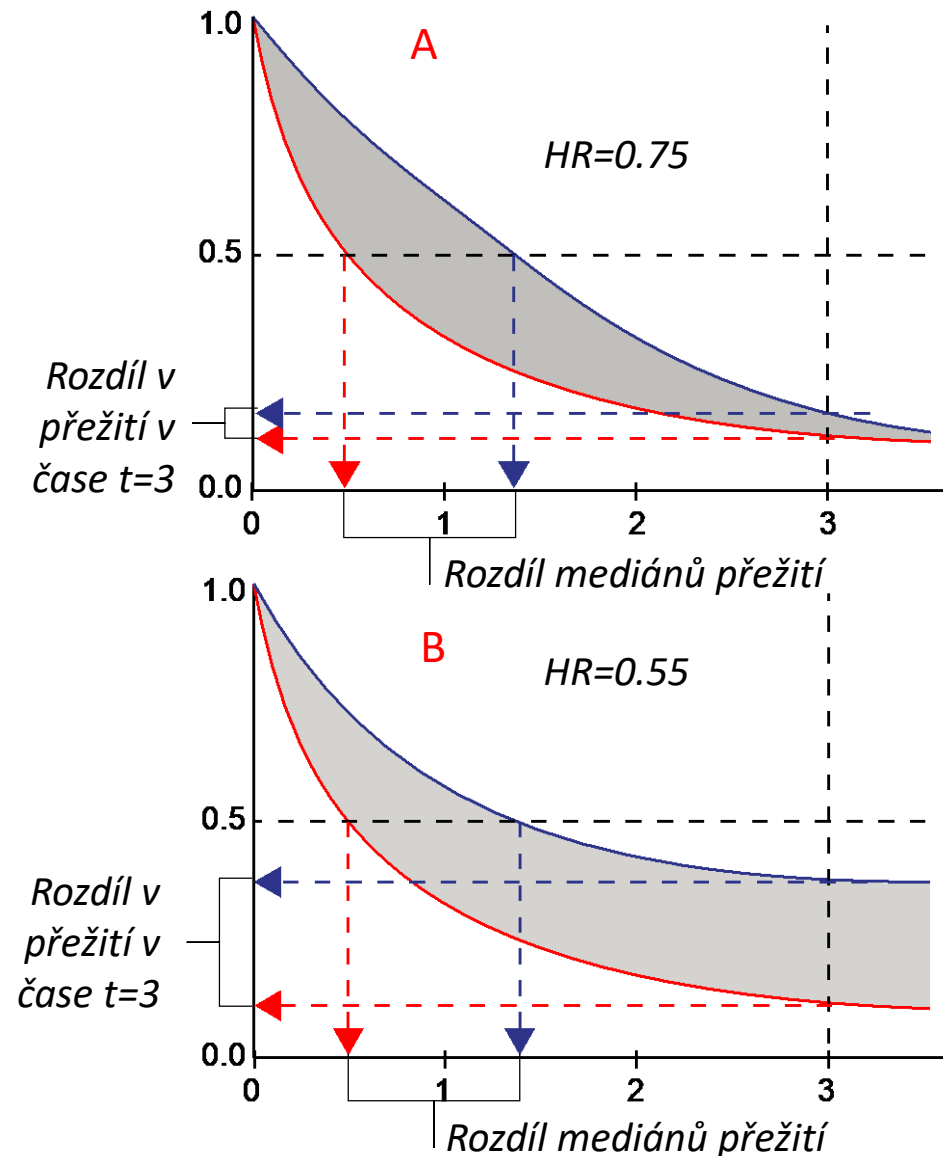


Fig. 3 CV events. Data presented for the overall population, the subgroup of patients with a history of CVD (CVD history) and the subgroup of elderly patients aged ≥ 65 years with a history of CVD and hypertension (Elderly patients with CVD risk). *n* is the number of patients with an event; *N* is number of patients in treatment group. *CI* confidence interval, *CTRL* control, *CV* cardiovascular, *CVD* CV disease, *DAPA* dapagliflozin, *HR* hazard ratio, *p-y* = patient years

Sonesson et al. Cardiovascular effects of dapagliflozin in patients with type 2 diabetes and different risk categories: a meta-analysis. *Cardiovasc Diabetol* (2016) 15:37

RŮZNÉ CHARAKTERISTIKY KŘIVKY PŘEŽITÍ – RŮZNÁ INTERPRETACE

- Různé charakteristiky křivek přežití přináší různou informaci -> nezbytné rozhodnutí, která z nich je relevantní pro hodnocenou studii
- Příklad A vs. B
 - Shodný rozdíl mediánů přežití
ALE
 - Vizuálně rozdílný průběh křivek
 - Rozdílný podíl přeživších v hodnoceném časovém bodě
 - Rozdílná hazard ratia (odpovídají rozdílnému průběhu křivek v jednotlivých příkladech)
- Nezbytné zhodnotit přínos různých charakteristik křivek přežití k jejich interpretaci



INTERPRETACE VÝSTUPU ANALÝZY PŘEŽITÍ – SROVNÁNÍ DAT Z RŮZNÝCH STUDIÍ

- Přímé srovnání hodnot z různých studií je extrémně problematické
- Nezbytné detailní hodnocení definice populací ve studiích, zařazovací a vyřazovací kritéria
- Příklad A vs. B
 - Shodný rozdíl mediánů přežití
ALE
 - Vizuálně rozdílný průběh křivek
 - Rozdílný podíl přeživších v hodnoceném časovém bodě
 - Rozdílná hazard ratia (odpovídají rozdílnému průběhu křivek v jednotlivých příkladech)
- Nezbytné zhodnotit přínos různých charakteristik křivek přežití k jejich interpretaci

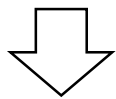
RCT A RWE

RCT: randomised clinical trials

Náhodný výběr,
randomizace,
zařazovací a
vyřazovací kritéria



vs.



Srovnáváme srovnatelné;
nemusí být plně
reprezentativní vůči cílové
populaci

RWE: real world evidence

Náhodný reprezentativní výběr
(nebo celopopulační data, např.
NZIS), srovnání existujících skupin



vs.



**Vzájemně se
doplňující
informace**

Popisujeme reálnou situaci;
srovnávané skupiny nemusí
být srovnatelné

RCT A RWE

RCT: randomised clinical trials

Otázka „Může to fungovat“?

- + průkaz účinnosti
 - + průkaz bezpečnosti
 - + nezbytné pro schválení nasazení léku v praxi
 - + zabezpečena srovnatelnost skupin
 - + minimalizován vliv zavádějících faktorů
- Účinnost a bezpečnost je prokázána za striktně definovaných podmínek, které nemusí plně reflektovat reálné podmínky populace pacientů
- Vliv zařazovacích a vyřazovacích kritérií na reprezentativnost vůči cílové populaci

RWE: real world evidence

Otázka „Funguje to“?

- + efektivita v klinické praxi = jak se lék chová při nasazení na reálné populaci pacientů
 - + utilizace rozsáhlých datových zdrojů (studie, registry, administrativní data)
- Srovnávání skupin pacientů může vést ke srovnávání nesrovnatelného (skupiny nejsou produktem randomizace)
- Nezbytné vypořádat se s vlivem zavádějících faktorů (např. pomocí adjustace vlivu proměnných, propensity skóre apod.)

ZKRESLENÍ VE STUDIÍCH



Kontrola vlivu zavádějících faktorů

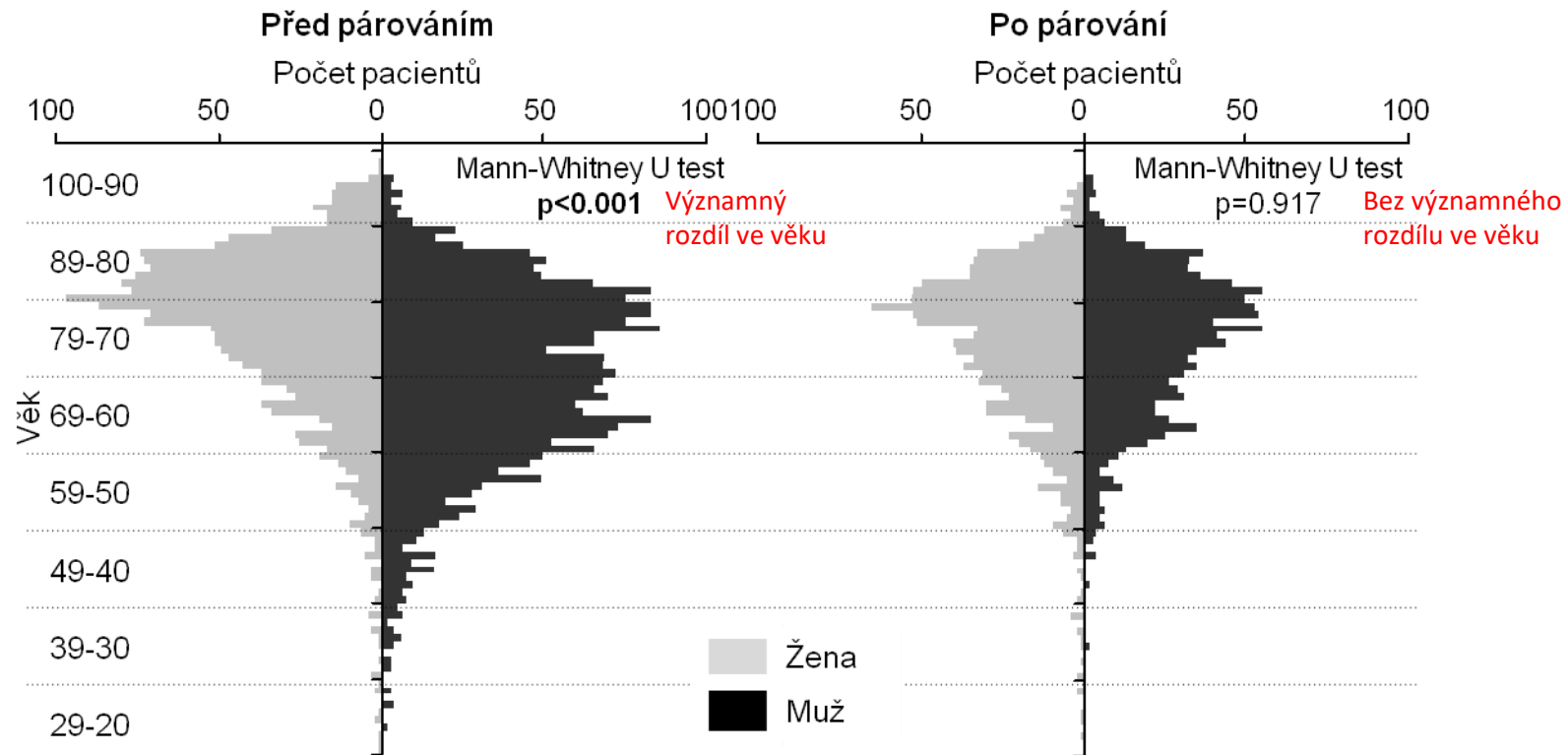
- Při plánování designu studie (strukturálním vyvažováním, standardizace, stratifikace, randomizace)
- Dodatečně (po proběhnutí experimentu) – adjustace, **propensity skóre**
 - Propensity skóre – odhad vlivu léčby oproštěný od vlivu zavádějících faktorů

PROPENSITY SKÓRE

- Nezávislost na definici ramene studie
- Odvozené skóre
 - Právěpodobnost, že subjekt byl zařazen do daného ramene studie
 - Subjekty jsou na základě této pravděpodobnosti párování
 - Získáme „quasi-randomizovaný“ experiment



PŘÍKLAD VÝPOČTU PROPENSITY SCORE



Srovnáváme srovnatelné

METAANALÝZA

- Statistická technika kombinující výsledky různých nezávislých studií
- Poskytuje přesný a souhrnný odhad efektů léčby
- Přesnější odhad reality než osamocená studie

- Metaanalýza je jen tak dobrá jako studie, na kterých je postavena !!!
 - Selektce publikovaných studií
 - Mnohonásobné publikování výsledků
 - Lze studie spojit v metaanalýze?
 - Různé cílové populace
 - Primární cíl hodnocení
 - Základní nastavení studií (např. v různých zemích)
 - Subjektivita při analýze podskupin studií
 - Analýza citlivosti metaanalýzy na různé charakteristiky studií

- Studie zahrnuté do metaanalýzy musí být zkontrolovány z hlediska výše popsaných kritérií
- Kromě metaanalýzy využívající pouze výsledky studií je v případě dostupných primárních dat studií možná tzv. pooled analýza – analýza ze sloučených dat

ANALÝZA DAT VE STUDIÍCH - SHRNU TÍ

- RCT (randomised clinical trial) zůstává standardem pro prokázání účinnosti, bezpečnosti a registrace léčiv
- Jednotlivá randomizovaná studie nicméně není schopna obsáhnout všechny aspekty reálného nasazení léku - > nezbytné využití všech dostupných datových zdrojů (s narůstajícím množstvím studií a údajů v registrech a dalších databázích narůstá i potřeba jejich souhrnného zpracování)
- RWE (real world evidence) – využití dat z reálného nasazení léku pro rozhodnutí o jeho efektivitě v klinické praxi
- Metaanalýza – systematický a kvantitativní přístup k hledání odpovědí na klinické otázky - vzhledem k závislosti na kvalitách využitých studií je ovšem nezbytný kritický přístup k jejím závěrům (splnění podmínky kvality studií)
- Ve všech případech je nezbytný kritický přístup zohledňující 5 klíčových principů biostatistiky, ty zůstávají v platnosti pro RCT, RWE i metaanalýzy



DĚKUJI ZA POZORNOST