

# Plánování, organizace a hodnocení klinických studií

## Statistické zpracování dat a jejich interpretace

# Struktura přednášky

- I. **Analýza dat klinických studií**
- II. **Cílové parametry hodnocení**
- III. **Experimentální design klinických studií**
- IV. **Randomizace**
- V. **Zaslepení studie**
- VI. **Interim analýza, subgroup analýzy**
- VII. **Principy statistické analýzy dat**
- VIII. **Statistická vs. klinická významnost**
- IX. **Optimalizace velikosti vzorku**
- X. **Aspekty korektního zpracování dat KHL**
- XI. **Příklady**

# I. Analýza dat klinických studií

# Medicína založená na důkazech

- Úkolem zdravotního systému je zajistit dostupnými prostředky nejlepší možný zdravotní a psychický stav národa.
- K naplňování tohoto úkolu by měl pomoci princip nazvaný medicína založená na důkazech („evidence based medicine“).
- **Medicína založená na důkazech je proces zabývající se systematickým hledáním, hodnocením a hlavně využitím současných výsledků klinického výzkumu při poskytování péče jednotlivým pacientům.**
  - Poskytování důkazů pomocí klinického výzkumu a vědecké literatury.
  - Vytváření klinických doporučení (založených na důkazech) a jejich distribuce.
  - Implementace účinných a efektivních postupů pomocí výuky a řízení kvality.
  - Hodnocení dodržování doporučených postupů pomocí klinických auditů, indikátorů kvality a výsledků léčebné péče.

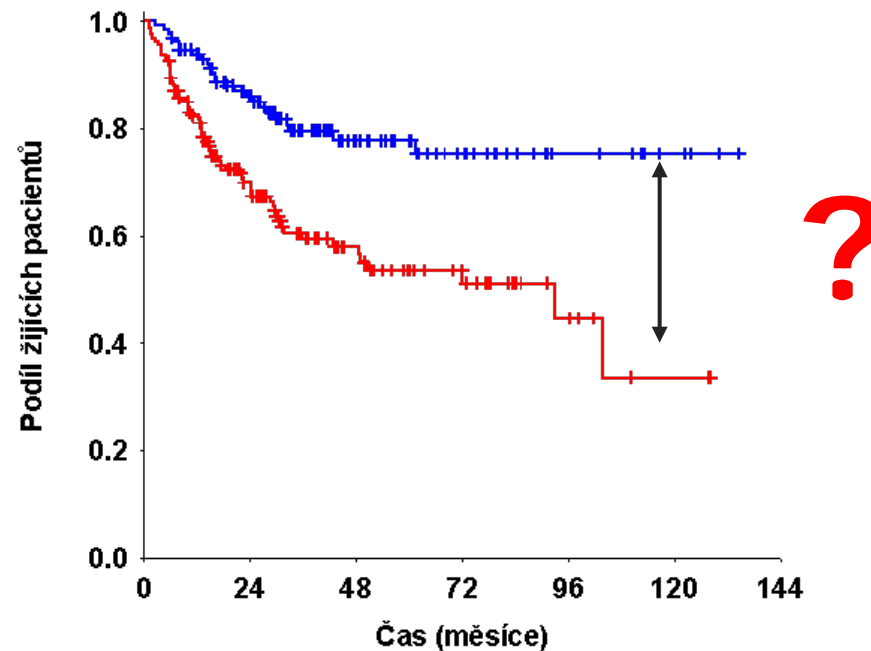
# Cíl klinických studií

## Primární cíl:

- Ověření **účinnosti** nebo **bezpečnosti** nového léčebného postupu nebo intervence.

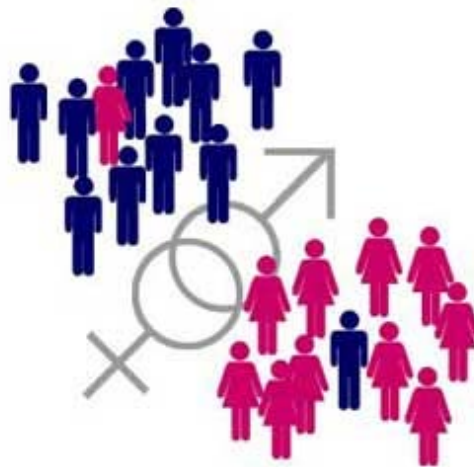
## Otázka:

- Jsou dosažené výsledky reálné nebo náhodné?



# Bias (zkreslení) v klinických studiích

- V hodnocení klinických studií se snažíme vyhnout zkreslení výsledků, tzv. „**bias**“.
- V případě klinických studií se jedná o **zkreslení výsledků** jinými faktory než těmi, které jsou cíli studie.
  1. Statistické srovnání není nikdy 100% spolehlivé, existuje náhoda a tedy i pravděpodobnost chybného úsudku – to nelze ovlivnit.
  2. My však chceme použít adekvátní metody pro odstranění vlivů, které by zkreslily výsledky a nebyly přitom náhodné (např. **zastoupení pohlaví**).



Úspěch a validita klinické studie jsou závislé na mnoha aspektech, některé z nich jsou však klíčové:

## 0. Správná formulace klinické hypotézy – endpoint

---

### 1. Použití souběžné kontrolní skupiny

### 2. Randomizace

### 3. Zaslepení studie

### 4. Optimalizace velikosti vzorku

## 0. Správná formulace klinické hypotézy

Je nutné definovat, co vlastně hodnotíme, neboť od hypotézy se odvíjí nejen uspořádání celé studie, ale i forma jejího vyhodnocení.

---

## 1. Použití souběžné kontrolní skupiny

Účinnost i bezpečnost léčby musí být hodnoceny v kontextu kontrolní skupiny pacientů, kteří nebyli exponováni experimentální léčbě, ale jsou srovnatelní v ostatních klinických faktorech.

## 2. Randomizace

Přiřazování subjektů do skupin musí být náhodné, protože pouze přiřazování, které není zatíženo subjektivitou, může umožnit relevantní srovnání primárních cílů studie.

## 3. Zaslepení

Cílem zaslepení studie je vyhnout se subjektivnímu hodnocení výsledků léčby nebo experimentu při vědomí pacienta a/nebo lékaře, do které léčebné skupiny pacient patří.

## 4. Optimalizace velikosti vzorku

Počet subjektů zapojených do klinické studie musí být optimalizován, aby nedošlo k ovlivnění statistického testu, případně aby tento test měl vůbec smysl.



# Validita klinické studie

- Medicína založená na důkazech – zajímají nás pouze „kvalitní“ důkazy.
- Hlavním aspektem kvality je **validita získaných výsledků**.
- **Interní validita studie:** odráží, jak moc lze rozdíly v účinnosti a bezpečnosti pozorované u srovnávaných skupin přisuzovat sledované intervenci. Chceme minimalizovat nenáhodnou chybu (zkreslení).
- **Externí validita studie:** odráží zobecnitelnost (z hlediska korektnosti) výsledků na jiné populace a experimentální podmínky.

# Interní validita studie

→ Interní validita studie jednoznačně souvisí se **zkreslením výsledků** (*bias*):

*„Bias is any process at any stage of inference tending to produce results that differ systematically from the true values.”*

→ V klinických studiích rozlišujeme čtyři typy zkreslení:

→ **Výběrové zkreslení** (*selection bias*)

→ **Prováděcí zkreslení** (*performance bias*)

→ **Detekční zkreslení** (*detection bias*)

→ **Úbytkové zkreslení** (*attrition bias*)

# Interní validita studie

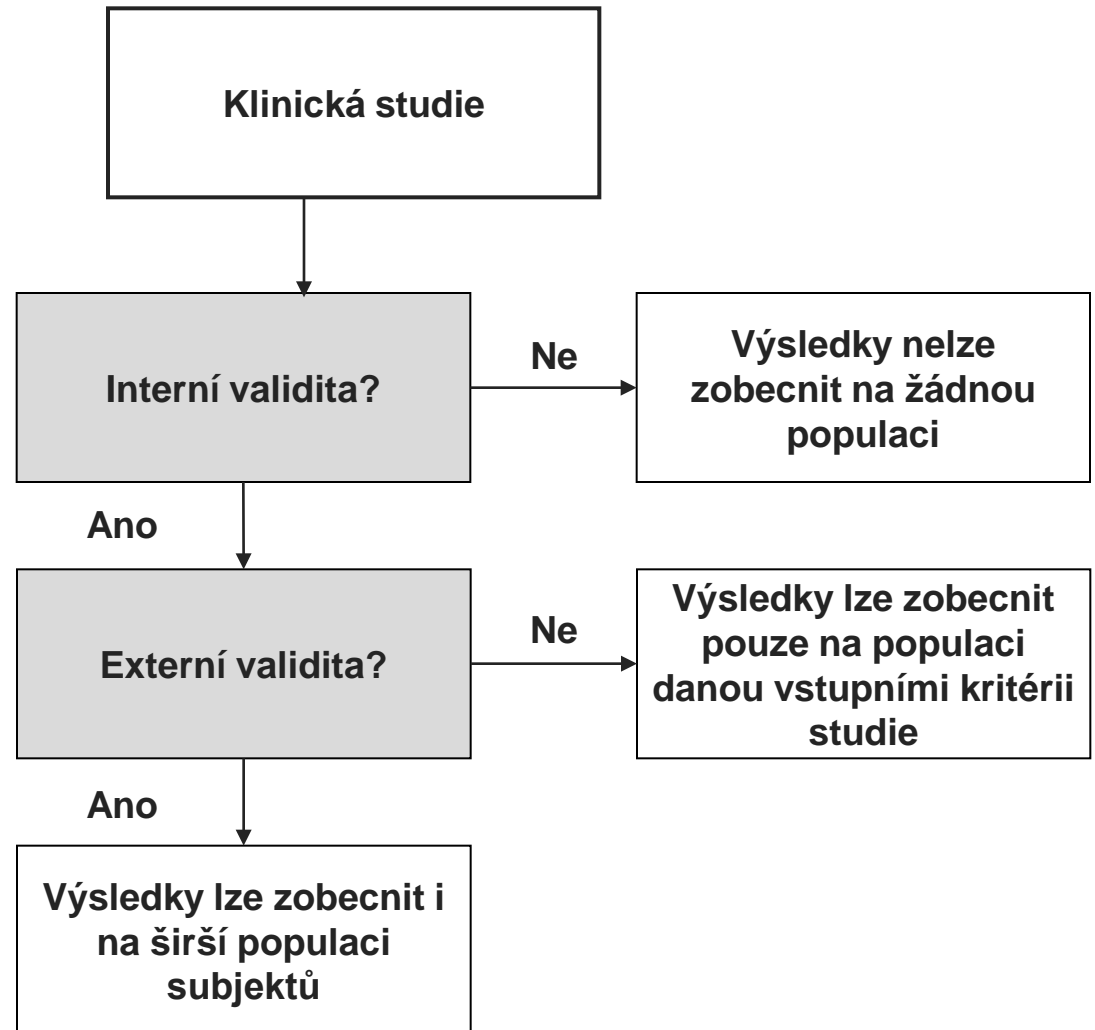
- **Výběrové zkreslení** (*selection bias*): vzniká ve chvíli, kdy srovnávané skupiny nejsou srovnatelné stran nějakého faktoru, který ovlivňuje výsledek. Vznik výběrového zkreslení se snažíme minimalizovat použitím randomizace.
- **Prováděcí zkreslení** (*performance bias*): vzniká ve chvíli, kdy srovnávané skupiny mají různé experimentální podmínky (samozřejmě kromě sledované intervence) – např. dostávají jinou léčbu/péči.
- **Detekční zkreslení** (*detection bias*): vzniká ve chvíli, kdy je u srovnávaných skupin různě hodnocena cílová proměnná. Vznik prováděcího a detekčního zkreslení se snažíme minimalizovat použitím zaslepení.
- **Úbytkové zkreslení** (*attrition bias*): vzniká ve chvíli, kdy srovnávané skupiny nejsou srovnatelné stran subjektů, které předčasně opustily studii a nelze u nich hodnotit cílovou proměnnou. Vznik úbytkového zkreslení se snažíme minimalizovat použitím *intention-to treat* analýzy.

# Externí validita studie

- Externí validita studie souvisí s **možností zobecnění jejích výsledků** na jinou populaci subjektů v jiném prostředí a za jiných podmínek.
- **Jiná populace subjektů**: vliv věku, pohlaví, tíže onemocnění, rizikových faktorů, komorbidit
- **Jiné prostředí**: systém zdravotnictví (USA×EU), organizace léčebné péče, specializace a zkušenosti lékařů
- **Jiné podmínky – léčba**: dávkování, podání, pravidelné sledování, další léčebné modality, souběžná léčba
- **Jiné podmínky – měření výsledků léčby**: definice cílové proměnné a její měřitelnost v daném prostředí

# Interní a externí validita studie

→ Interní validita studie je nutným předpokladem externí validity.



## II. Cílové parametry hodnocení

# Cílové parametry testování hypotéz v klinických studiích

1. Klinický parametr: Většinou je hodnocena změna v daném parametru nebo dosažení určité hladiny.
2. **Léčebná odpověď a odvozené parametry:** Hodnocení ústupu nádoru jako odpovědi na léčbu a délky času trvání této redukce. Tento endpoint je nejčastěji používaný ve studiích fáze II ale i v ostatních typech studií.
3. Time-to-event: Hodnotí délku času, po kterou je subjekt bez sledované události, např. bez zhoršení primárního onemocnění. Často používaný endpoint pro studie fáze III. OS, PFS, DFS.
4. Kvalita života: Parametr založený na subjektivním zhodnocení celkového zdravotního stavu pacienta a ovlivnění jednotlivých aspektů jeho života. Jedná se o současné zhodnocení přínosů léčby a negativního vlivu nežádoucích účinků.

# Př.: Hodnocení léčebné odpovědi na základě RECIST kritérií

Cílové léze (Target lesion) <sup>1</sup>		
Kompletní remise (odpověď)	CR	Vymizení všech cílových lézí
Parciální remise (odpověď)	PR	Minimálně 30% pokles součtu nejdelších průměru (LD) ve srovnání s úvodním měřením
Progrese	PD	Minimálně 20% nárůst součtu nejdelších průměru (LD) nebo objevení se 1 či více nových lézí
Stabilizace onemocnění	SD	Nesplnění kritérií pro PD, PR nebo CR
Ne-cílové léze (Non-target lesions) <sup>2</sup>		
Kompletní remise (odpověď)	CR	Vymizení všech ne-cílových lézí a normalizace nádorových markerů
Nekompletní remise/stabilizace	Non-CR/SD	Přetrvávání jedné nebo více necílových lézí a/nebo přetrvávání hladin nádorových markerů nad normu
Progrese	PD	Jednoznačná progrese existujících necílových lézí nebo objevení se 1 či více nových lézí

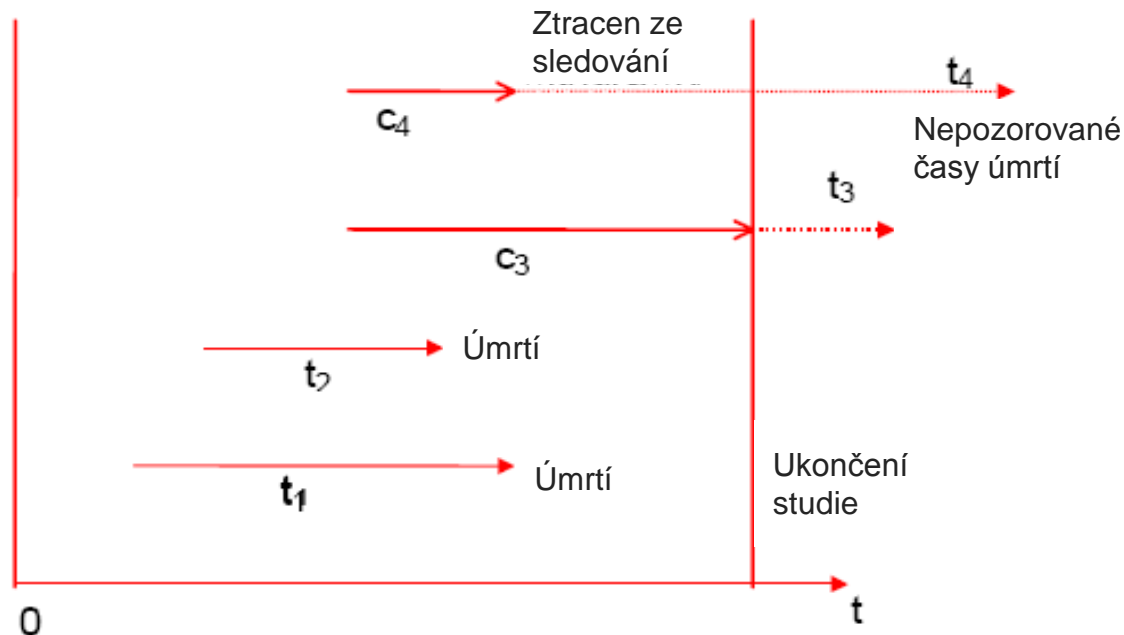
## Celková terapeutická odpověď

Cílové léze	Ne-cílové léze	Nové léze	Celková odpověď
CR	CR	Ne	CR
CR	non-CR/SD	Ne	PR
PR	non-PD	Ne	PR
SD	non-PD	Ne	SD
SD	jakákoliv odpověď	Ano nebo Ne	PD
jakákoliv odpověď		Ano nebo Ne	PD
jakákoliv odpověď	jakákoliv odpověď	Ano	PD

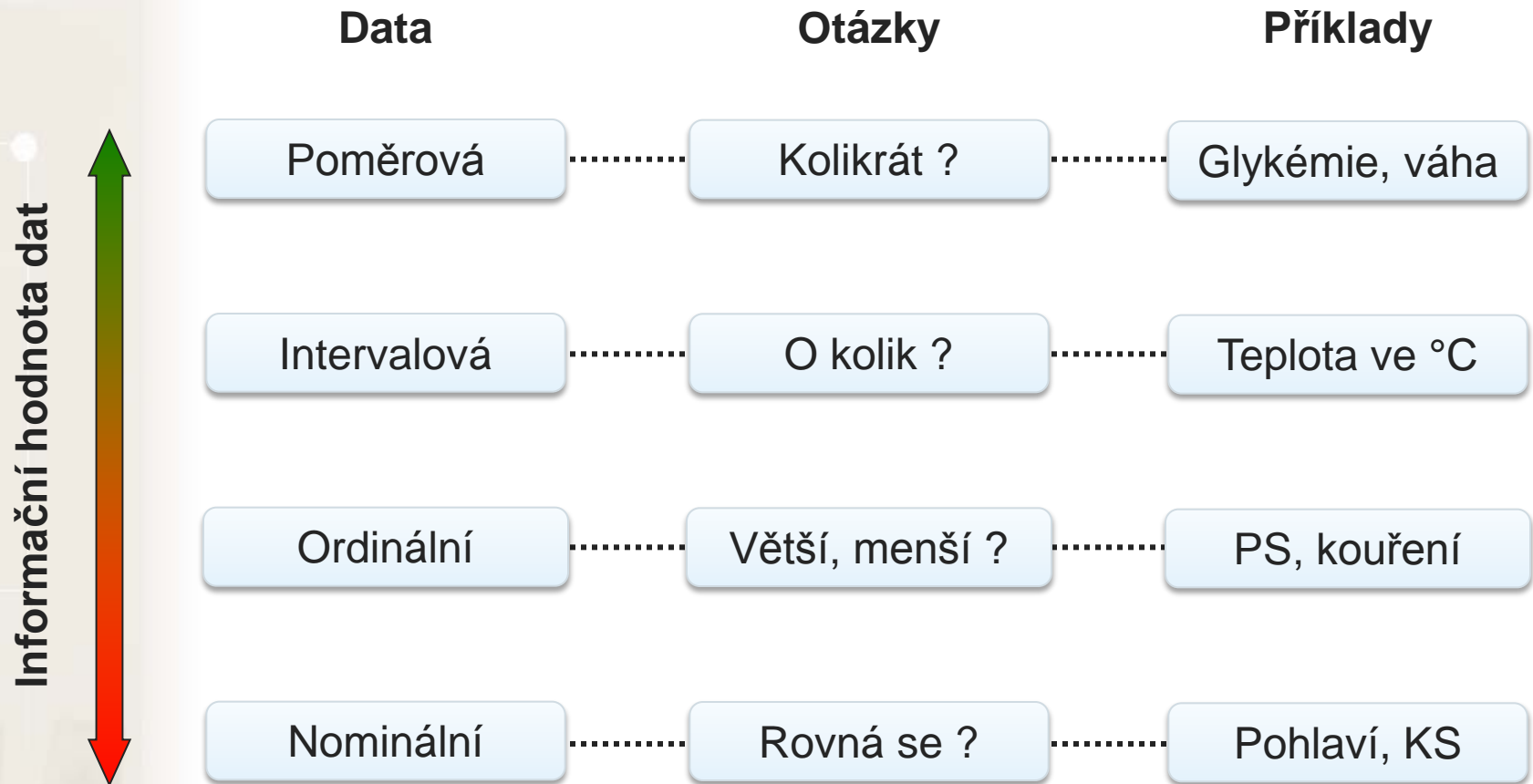


# Přežití jako endpoint

- Studujeme čas předcházející události, která nás zajímá
- Na rozdíl od mortality nechceme pouze počty událostí (událost = smrt, progrese nemoci, relaps ...)
- Musí být jasně časově definován počátek (je většinou individuální) i konec sledování.
- Cenzorování je typické pro analýzu přežití (kompletní čas přežití nemusí být pozorován u všech subjektů)



# Typy dat dle škály hodnot



# III. Experimentální design klinických studií

# Design klinických studií

- Klíčová věc při plánování klinické studie, od které se odvíjejí další kroky.
- Špatný design většinou vede ke znehodnocení výsledků celé studie.
- Správný design je základním předpokladem úspěšného průběhu studie.

## **Design studie musí především adekvátně zohledňovat**

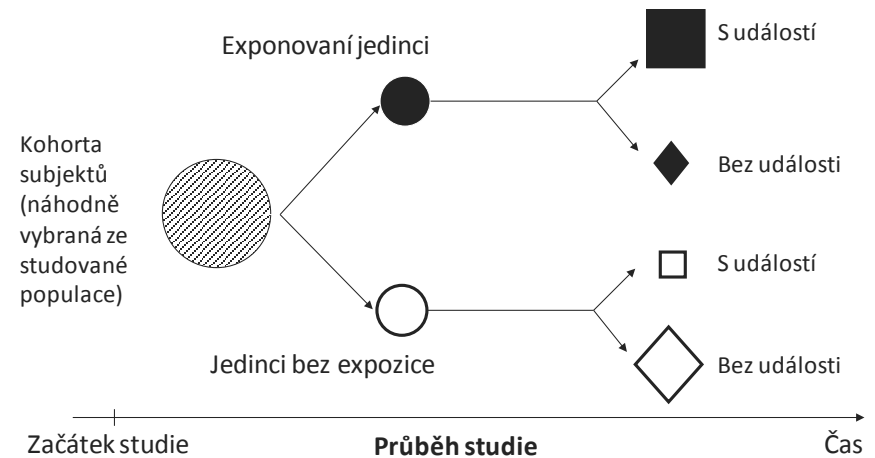
stanovené klinické hypotézy.

- Použitý experimentální design určuje jak organizaci studie tak její následné statistické zpracování.

# Prospektivní a retrospektivní studie

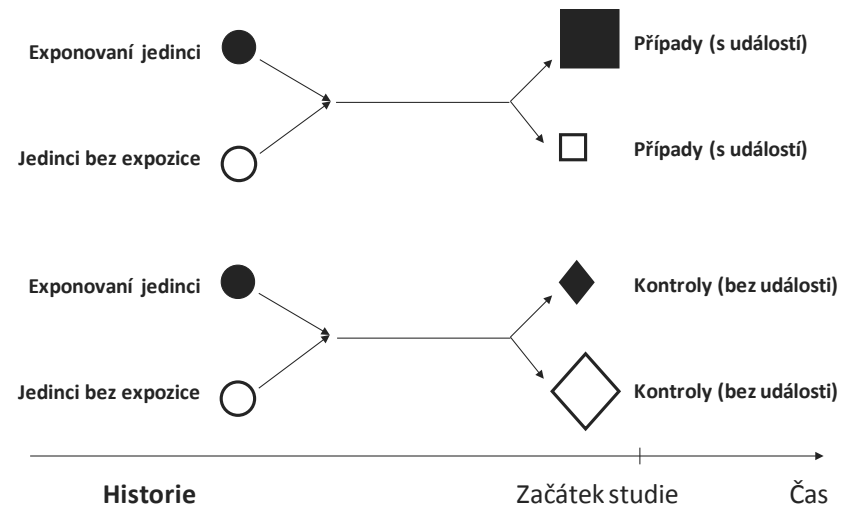
## ➤ Prospektivní studie

- U některých subjektů je rizikový faktor přítomen a u jiných ne → sledujeme v čase, zda se vyskytne událost.



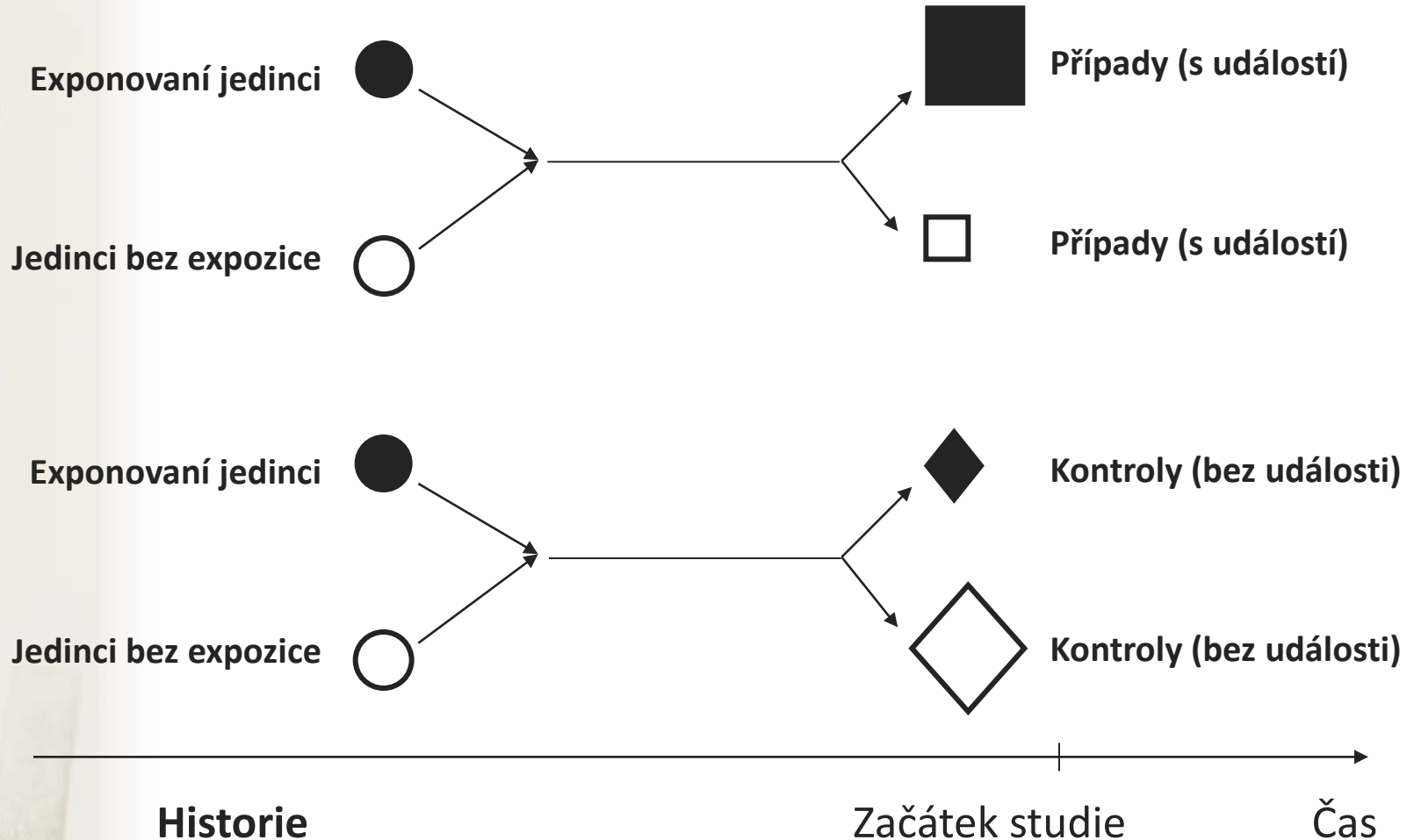
## ➤ Retrospektivní studie

- U některých subjektů se událost vyskytla a u jiných ne → zpětně hodnotíme, zda se liší s ohledem na nějaký rizikový faktor.



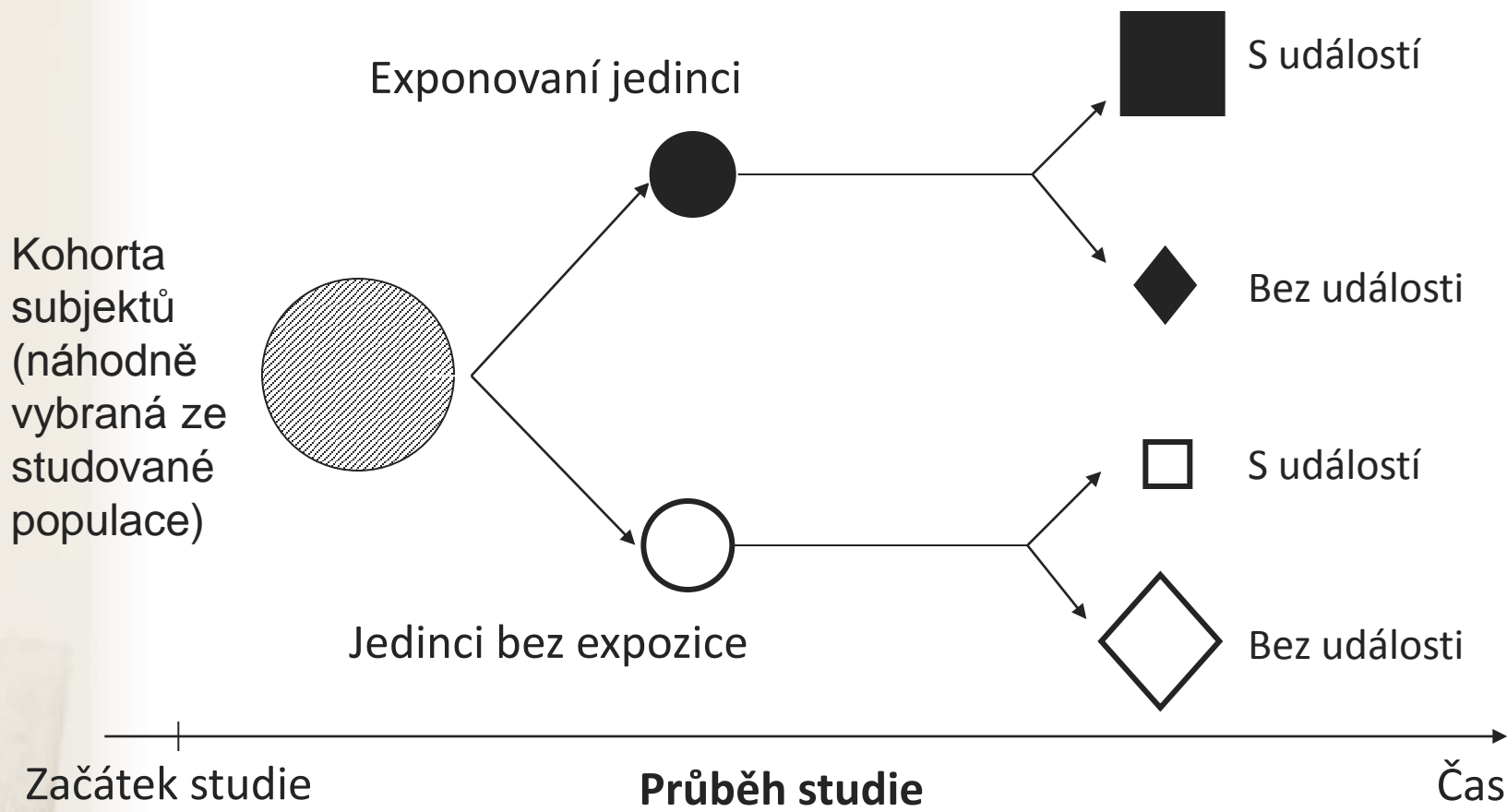
# Case-control (případ-kontrola) studie

Retrospektivní sledování souboru osob, u nichž došlo ke sledovanému jevu (např. vzniku onkologického onemocnění). Tuto skutečnost vztahujeme k výskytu určitého faktoru (rizikového nebo ochranného) v minulosti.



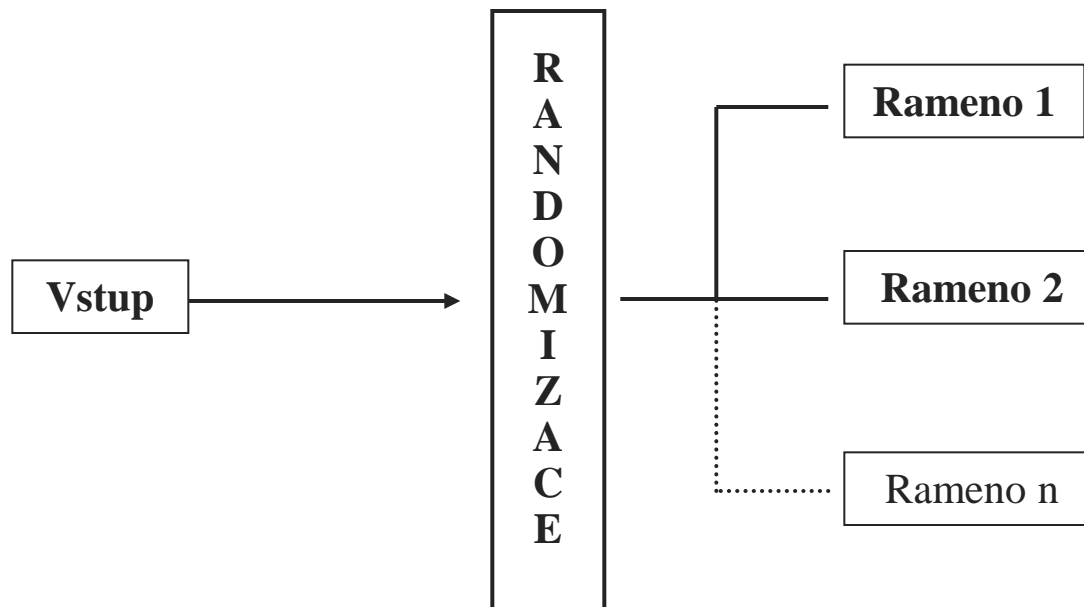
# Kohortní studie

Prospektivní sledování souboru osob, z nichž část byla vystavena působení určitého faktoru. Po adekvátní době sledování je zhodnocen výskyt sledovaného jevu ve skupině ovlivněné definovaným faktorem a ve skupinu bez něj.



# Paralelní design randomizované studie

Tento typ uspořádání je charakteristický pro studie fáze III kde je srovnávána účinnost a bezpečnost dvou nebo více terapeutických postupů. Základním prvkem designu je randomizace subjektů do jednotlivých léčebných skupin. V průběhu celé studie je zachováno jedinečné přiřazení subjektů do léčebných skupin.





# Paralelní design - příklad

Breast Cancer Res Treat (2008) 109:481–489

DOI 10.1007/s10549-007-9672-y

CLINICAL TRIAL

## A randomized phase-III trial of docetaxel/capecitabine versus doxorubicin/cyclophosphamide as primary chemotherapy for patients with stage II/III breast cancer

Keun Seok Lee · Jungsil Ro · Byung-Ho Nam · Eun Sook Lee ·  
Youngmee Kwon · Hye Soo Kwon · Ki Wook Chung · Han Sung Kang ·  
Eun A Kim · Seok Won Kim · Kyung Hwan Shin · Seok-Ki Kim

Received: 14 June 2007 / Accepted: 26 June 2007 / Published online: 26 July 2007

© Springer Science+Business Media B.V. 2007

**Abstract** We aimed to determine the efficacies of a non-anthracycline-containing regimen, docetaxel/capecitabine (TX), in comparison with an anthracycline-containing regimen, doxorubicin/cyclophosphamide (AC), as primary chemotherapy for node-positive early stage breast cancer. In this phase-III single center randomized study, we randomized 209 women with axillary node positive, stage II/III breast cancer to receive four cycles of either TX or AC

into a gain in DFS, the patients who achieved pCR in LN developed significantly less recurrence.

**Keywords** Breast cancer · Capecitabine · Clinical trial · Docetaxel · Phase III

**Introduction**

# Paralelní design - příklad

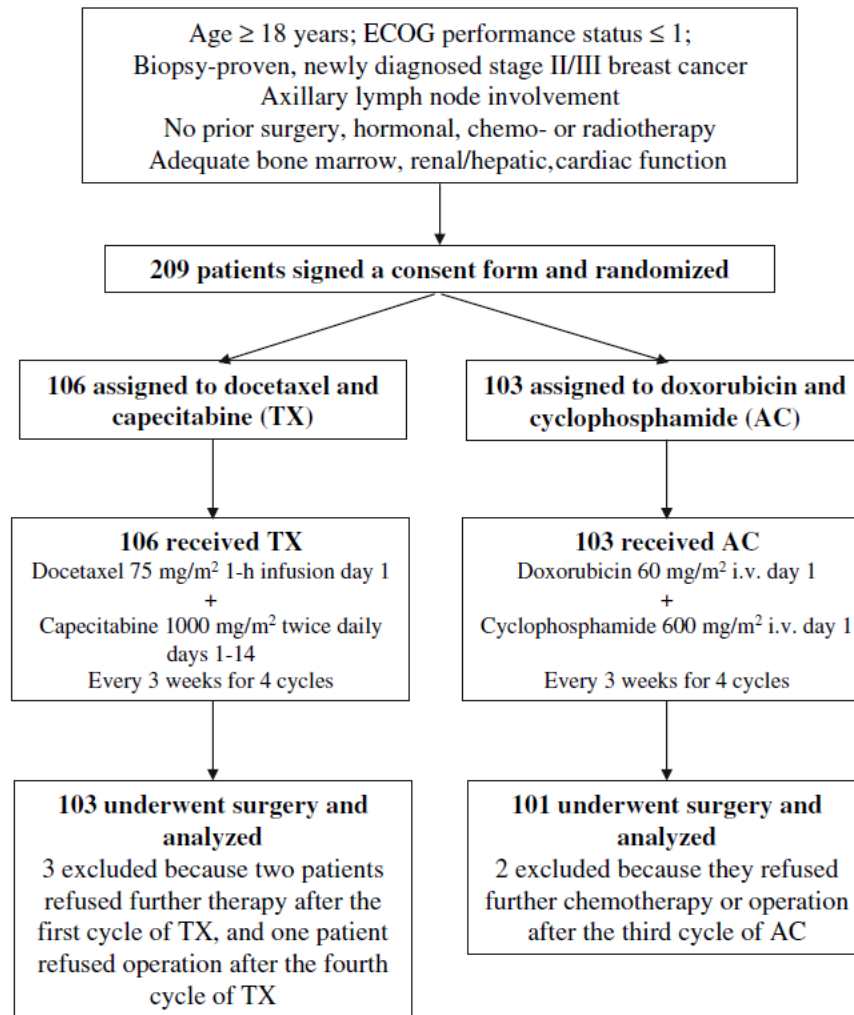
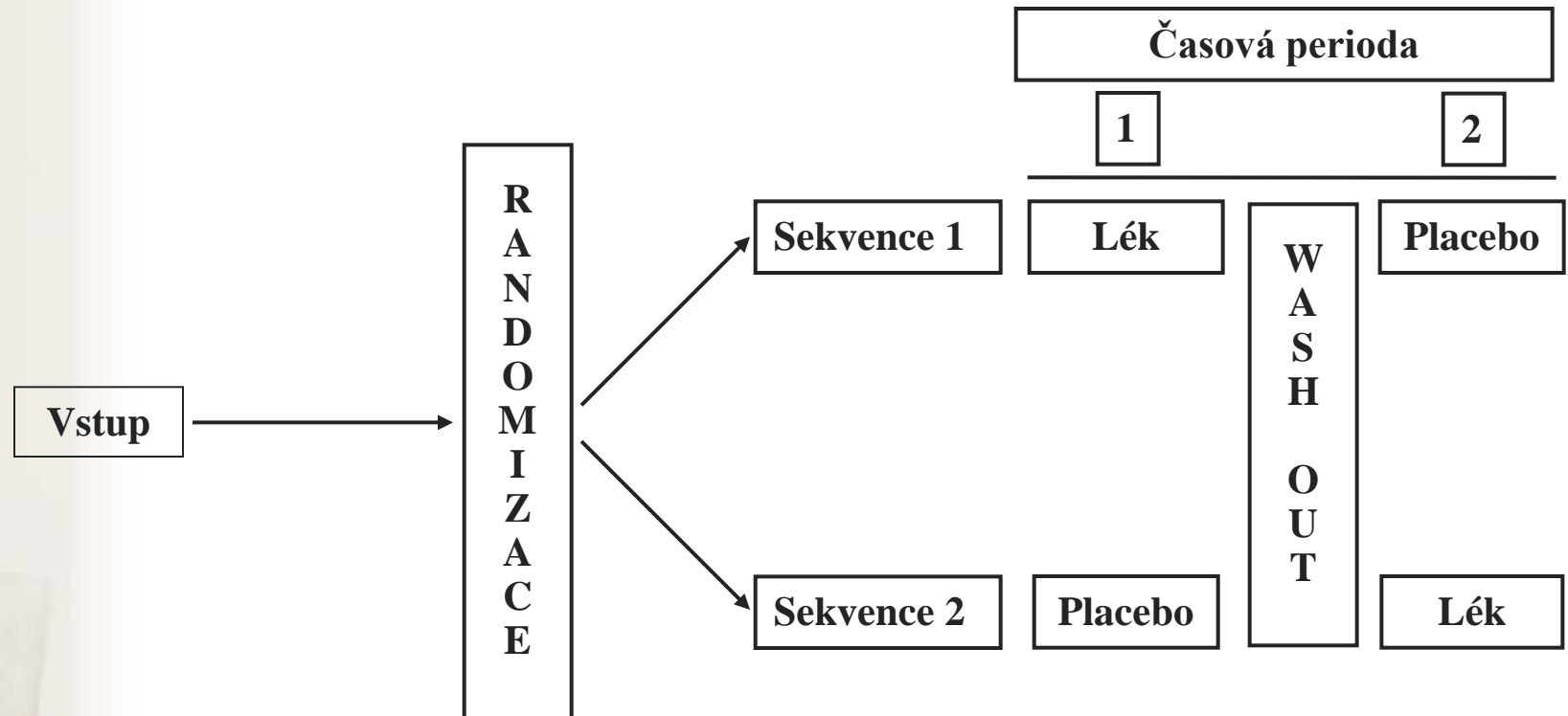


Fig. 1 Study profile

# Cross-over design randomizované studie

Tento typ uspořádání je nejčastěji používán ve studiích kde předpokládáme velkou interindividuální variabilitu v účinnosti a bezpečnosti hodnocených léčivých přípravků. Základem je časově oddělená aplikace všech srovnávaných přípravků všem subjektům.



# Cross-over design - příklad

Journal of the American College of Cardiology  
© 2006 by the American College of Cardiology Foundation  
Published by Elsevier Inc.

Vol. 47, No. 10, 2006  
ISSN 0735-1097/06/\$32.00  
doi:10.1016/j.jacc.2005.12.056

## CLINICAL RESEARCH

## Clinical Trials

### Biventricular Versus Conventional Right Ventricular Stimulation for Patients With Standard Pacing Indication and Left Ventricular Dysfunction

#### The Homburg Biventricular Pacing Evaluation (HOBIPACE)

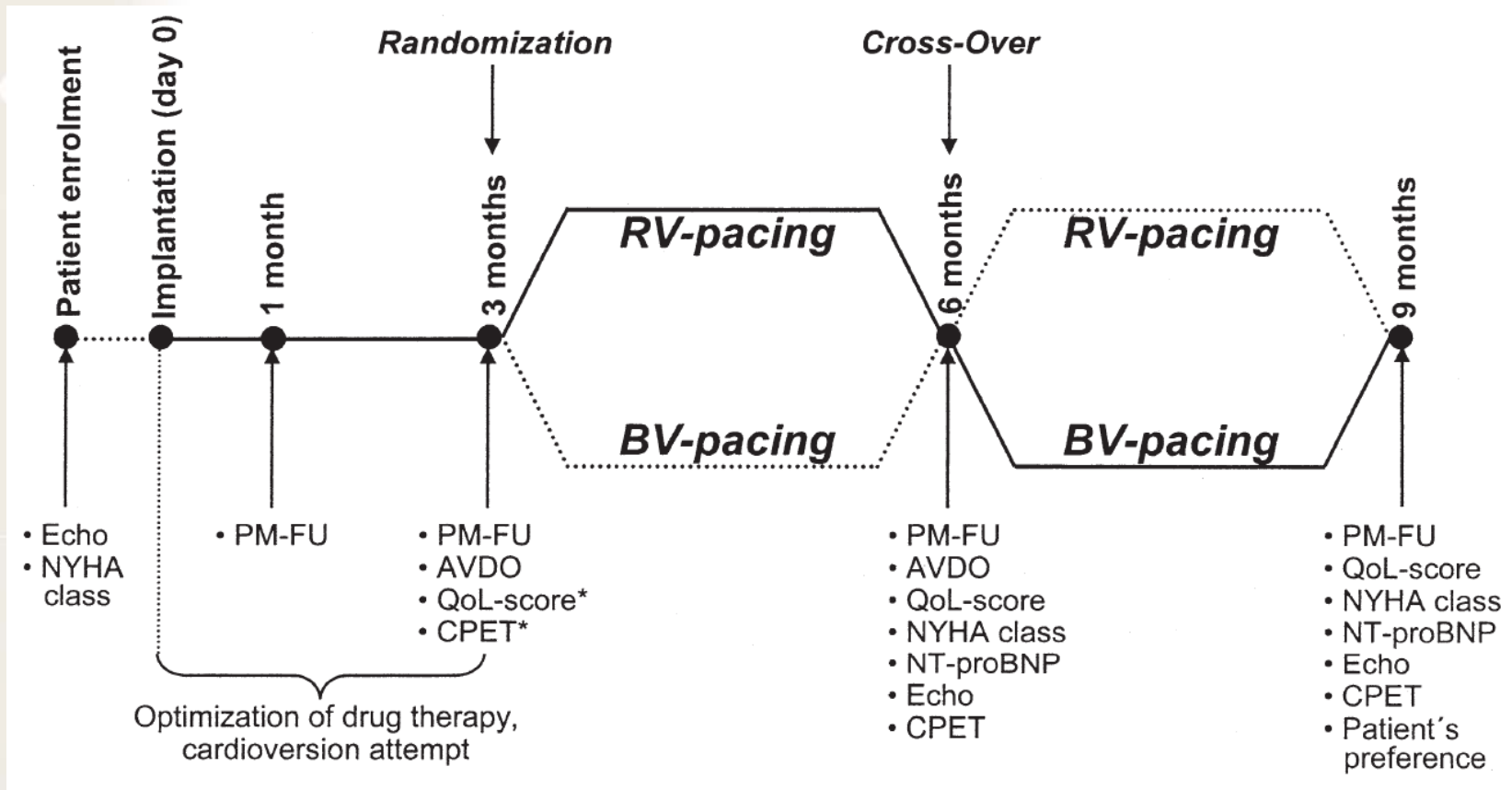
Michael Kindermann, MD,\* Benno Hennen, MD,\* Jens Jung, MD,† Jürgen Geisel, MD,‡  
Michael Böhm, MD,\* Gerd Fröhlig, MD\*

*Homburg/Saar and Worms, Germany*

- OBJECTIVES** The Homburg Biventricular Pacing Evaluation (HOBIPACE) is the first randomized controlled study that compares the biventricular (BV) pacing approach with conventional right ventricular (RV) pacing in patients with left ventricular (LV) dysfunction and a standard indication for antibradycardia pacing in the ventricle.
- BACKGROUND** In patients with LV dysfunction and atrioventricular block, conventional RV pacing may yield a detrimental effect on LV function.
- METHODS** Thirty patients with standard indication for permanent ventricular pacing and LV dysfunction defined by an LV end-diastolic diameter  $\geq 60$  mm and an ejection fraction  $\leq 40\%$  were included. Using a prospective, randomized crossover design, three months of RV pacing were compared with three months of BV pacing with regard to LV function, N-terminal pro-B-type natriuretic peptide (NT-proBNP) serum concentration, exercise capacity, and quality of life.

# Cross-over design - příklad

**Study protocol.** HOBIPACE was a monocentric, prospective, randomized, single-blinded crossover comparison between RV and BV pacing. It was an investigator-driven trial



# Rozdíl mezi paralelním a cross-over designem

- Výhodou cross-over designu je aplikace standardní i experimentální léčby na všech subjektech, tedy každý subjekt je zároveň sám sobě kontrolou → cross-over design vyžaduje menší celkový počet subjektů.
- Výhodou paralelního uspořádání je jeho jednoduchost, a to jak při organizaci studie, tak při zpracování výsledků.
- Cross-over design nelze aplikovat vždy (např. v onkologické léčbě nebo při srovnávání většího počtu srovnávaných léčiv).
- U cross-over designu je nutné brát v úvahu tzv. carry-over efekt.

# Další experimentální designy

## ➤ Faktoriální design

- Vyhodnocujeme účinnost/bezpečnost dvou a více léčebných postupů zároveň
- Subjekty jsou přiřazovány ke všem možným léčebným kombinacím

## ➤ Adaptivní designy

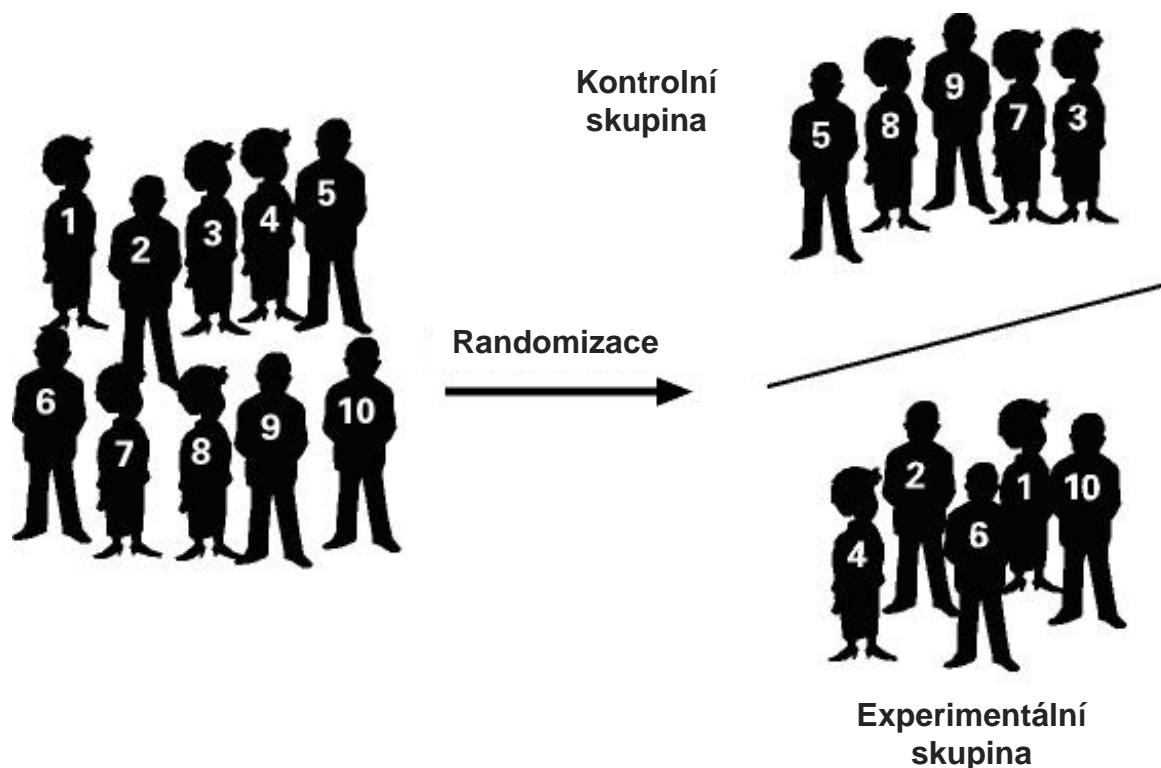
- Na rozdíl od „statických“ designů využívají průběžné informace zaznamenané ve studii k modifikaci jejích aspektů (např. velikosti vzorku)

## IV. Randomizace



# Randomizace

Metodika a proces náhodného (pseudonáhodného) rozdělování subjektů hodnocení (pacientů) do dvou nebo více léčebných skupin.



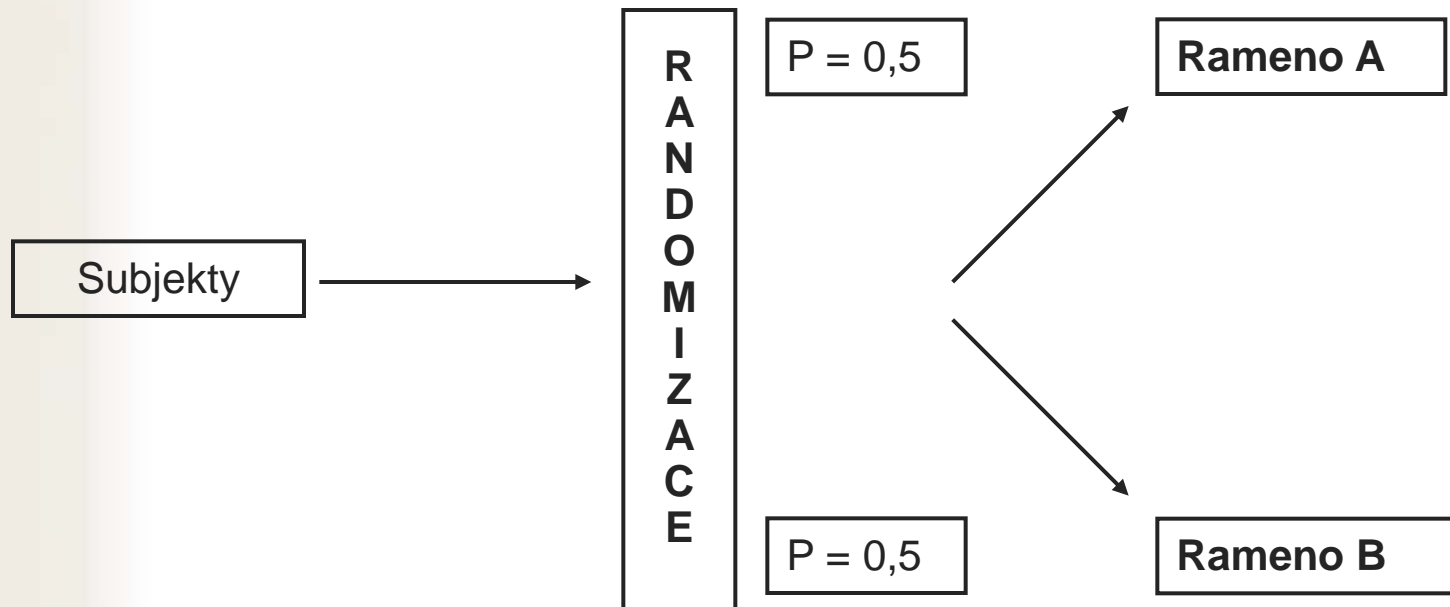
# Cíle a smysl randomizační procedury

- Zamezit subjektivnímu a selektivnímu rozdělování pacientů (subjektů) do léčebných skupin.
- Zamezit systematické chybě, zkreslení
- Splnit základní **předpoklad validního použití statistických testů** - náhodný výběr z populace
- Zabezpečit požadovaný poměr počtu pacientů v jednotlivých léčebných skupinách
- Možnost kontroly rovnoměrné distribuce prognostických faktorů v léčebných skupinách
- Randomizace však **není zárukou** rovnoměrné distribuce prognostických faktorů – i při minimálním rozdílu v klíčovém parametru mezi jednotlivými skupinami je nutné tento rozdíl brát při statistickém hodnocení v potaz

# Kompletní randomizace

**Princip: Náhodné rozdělování pacientů do definovaných skupin**

Např. klinická studie se dvěma rameny: A,B;  $N = 100$ .



**Nevhodná!**

**Riziko nevyváženého počtu subjektů hodnocení v léčebných skupinách**

Při použití kompletní randomizace je pravděpodobnost shodného počtu SH ve dvou léčebných skupinách se 100 SH celkem pouze 8%.

# Permutační bloková randomizace

**Princip metody:** pravidelné opakování bloků složených z identifikátorů skupin zajišťující stejný počet subjektů v obou skupinách.

**Př.:** Klinická studie se dvěma rameny: A,B     $N = 80$

- 1) Volba velikosti bloku:  $B = 4$
- 2) Permutace: Celkem 6 možností bloku:

1	AABB
2	BBAA
3	ABBA
4	BAAB
5	ABAB
6	BABA

- 3) Sestavení randomizačního plánu jako sekvence 20 náhodně po sobě jdoucích bloků

# Stratifikovaná permutační bloková randomizace

Principem je provedení permutační blokové randomizace v rámci definovaných podskupin pacientů.

Př.: KH: N = 80 SH, 2 ramena (A,B), blok o velikosti B = 4

1) Volba stratifikačních kriterií:

- pohlaví: muž x žena
- věk:  $\leq 50$  x  $> 50$

2) Vytvoření čtyř podskupin SH na základě všech možných kombinací prognostických faktorů

3) Bloková randomizace v rámci podskupin:

Blok	Skupina			
	Muži		Ženy	
	$\leq 50$	$> 50$	$\leq 50$	$> 50$
1	A	A	B	A
	B	A	B	A
	B	B	A	B
	A	B	A	B
2	B	A	B	A
	B	B	A	B
	A	A	A	A
	A	B	B	B
	.	.	.	.

- Stratifikační faktor musí být parametr zjistitelný před randomizací subjektů
- Maximální efektivity randomizace je dosaženo při volbě 2-3 stratifikačních parametrů (dle rozsahu studie)
- Stratifikační parametry musejí být nestochastické proměnné
- Stratifikací nelze kontrolovat všechny prognostické faktory, tento problém je předmětem post-randomizačních vícerozměrných statistických analýz

# Adaptivní randomizace

**Princip:** Každý SH je randomizován do léčebné skupiny na základě rozložení prognostických faktorů z dosavadního průběhu KH

**Př.:** Klinická studie se třemi sledovanými prognostickými faktory.  
Dosavadní průběh:

		Terapie	
		Placebo	Lék
n		106	107
Věk	≤50	57	56
	>50	49	51
Klin. par. 1	≤10	45	44
	>10	61	63
Stadium	1	25	26
	2	52	51
	3	29	30

**Další SH:** Věk = 68 let; Klinický parametr 1 = 7,4; Stadium = 3

Počet SH se stejnou kombinací prognostických faktorů:

Placebo rameno:  $49+45+29 = 123$

Lék:  $51+44+30 = 125$

SH bude s pravděpodobností  $P > 0,5$  (3/4 nebo 2/3) randomizován do ramene s placebem.

# Limitace randomizačních procedur

- **Randomizace** nezaručuje externí validitu klinické studie, ta je dána především vstupními kritérii
- **Randomizace** nezaručuje rovnoměrnou distribuci všech prognostických faktorů v léčebných skupinách
- **Randomizace** je pouze jedna z řady metod využívaných k minimalizaci rizika zkreslení výsledků klinických studií – tzn. není samospásná
- Výběr vhodné randomizační procedury je limitován a předurčen designem klinické studie

# V. Zaslepení studie



# Blinding - zaslepení

- Zaslepení studie je proces, kdy pacient a/nebo investigátor nevědí, jakou léčbu pacient dostává (při paralelním uspořádání studie není znám podaný lék, při change-over uspořádání studie není známo pořadí podaných léků).
- To ovšem neznamena, že by pacienti nevěděli, jaké léky jsou ve studii použity!
- **Cílem zaslepení je vyhnout se úmyslnému i neúmyslnému zkreslení výsledků studie.**
  - Pacient jinak hodnotí svůj zdravotní stav, když ví, je-li mu podávána nová léčba nebo placebo.
  - Lékař může být tímto vědomím také ovlivněn, např. při použití kategoriálního hodnocení.

# Blind (zaslepené) studie



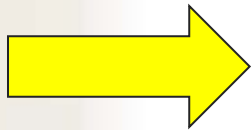
## Single blind (jednoduše slepá) studie

- pacient neví jaký lék dostává, zatímco lékař jej zná
- téměř se nepoužívají, pouze v případech, kdy nelze jinak
- např. v chirurgii (lékař musí vědět, jak pacienta operovat)



## Double blind (dvojitě slepá) studie

- pacient ani lékař neznají léčbu přidělenou pacientovi
- drtivá většina studií je dvojitě zaslepená
- např. v onkologii při testování nových modalit léčby rakoviny



## Triple blind (trojitě slepá) studie

- pacient, lékař ani statistik studie neznají léčbu přidělenou pacientovi
- modifikace dvojitě zaslepené studie
- z hlediska objektivity studie je to další zlepšení

## VI. Interim analýza, subgroup analýzy

# Subgroup analýza x interim analýza I

## Subgroup analýza:

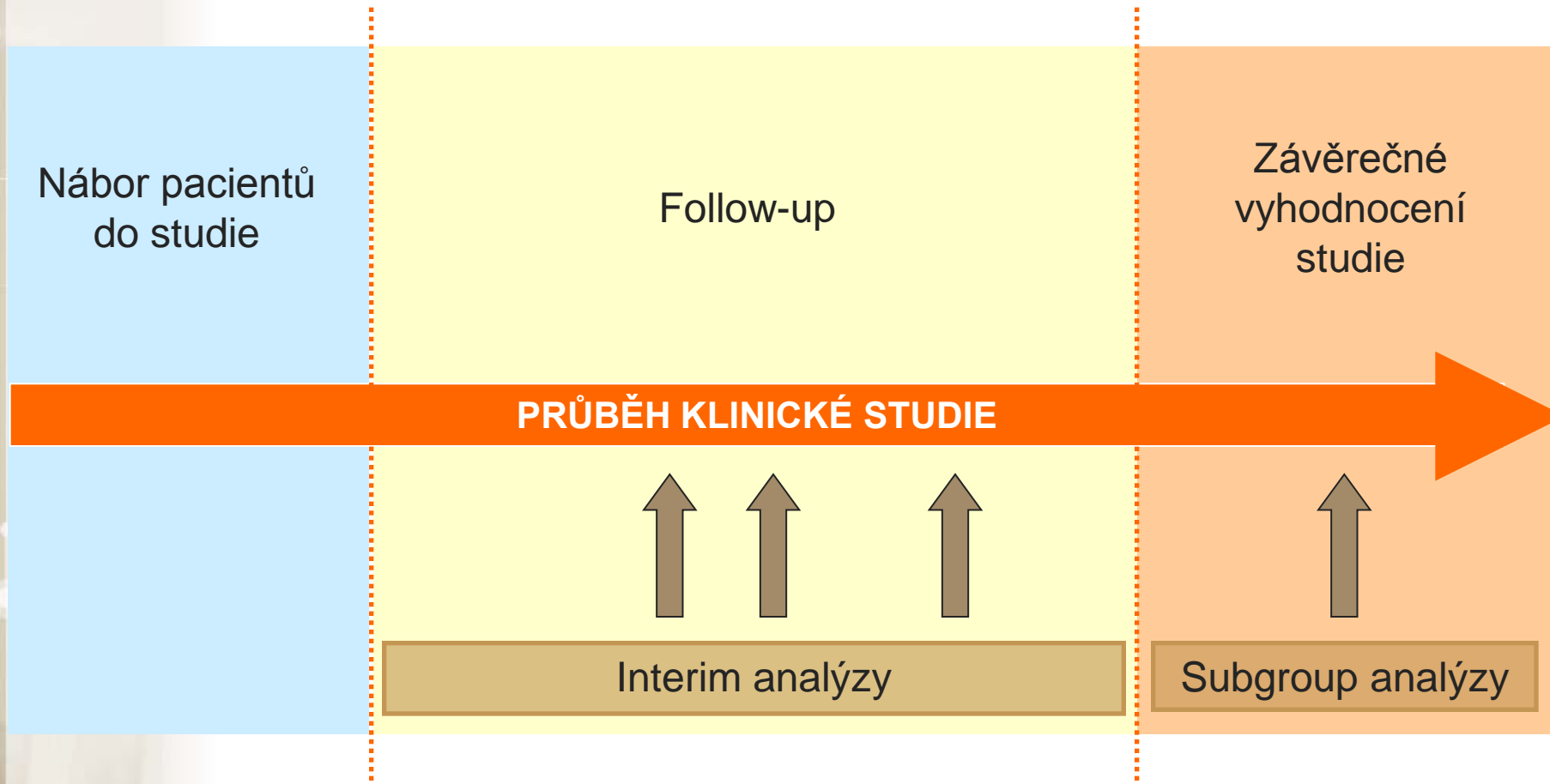
Analýza bezpečnosti a účinnosti léčby prováděná na podskupinách subjektů hodnocení definovaných na základě stanovených vstupních kritérií (jednoho či více).

## Interim analýza:

Rozhodnutí o dalším pokračování studie na základě zhodnocení platnosti vstupních předpokladů z dostupných dat. Tedy analýza bezpečnosti a účinnosti léčby prováděná před ukončením follow-up.

- V případě interim analýz i subgroup analýz se jedná o násobné testování hypotéz – výsledné p-hodnoty tedy musí být korigovány na počet provedených testů.
- Interim analýza – O'Brien-Fleming kritérium; Subgroup analýza – Bonferroniho korekce.

# Subgroup analýza x interim analýza II



# Subgroup analýzy v klinických studiích

- **Výhody:** Možnost nalezení skupiny pacientů, u kterých je/není léčba účinnější a bezpečnější – cílená terapie.
- **Nevýhody:** Zvýšení rizika průkazu falešně pozitivního výsledku – zjištění rozdílu v účinnosti/bezpečnosti tam, kde ve skutečnosti není.
  - **Násobné testování hypotéz:** provádíme více testů zároveň, chceme ovšem, aby celková pravděpodobnost falešně pozitivního výsledku zůstala  $\alpha$  => musíme výsledné p-hodnoty korigovat na počet provedených testů.
  - **Síla testu:** velikost vzorku v klinické studii je optimalizována vzhledem k primárnímu endpointu studie. Subgroup analýzy jsou prováděny na podskupinách, které jsou menší, tudíž síla testu je v těchto případech nedostatečná.
  - **Adekvátní statistický test:** testovat rozdíl v efektivitě léčby v rámci skupin pacientů (definovaných např. pohlavím nebo věkem) lze několika způsoby, které však nemusí být správné.
- **Výsledky subgroup analýz nelze přeceňovat – v kontextu klinické studie je hlavním výsledkem studie vyhodnocení primárního endpointu studie. Na základě výsledku subgroup analýz nelze vytvářet klinická doporučení.**

# Subgroup analýzy: doporučení

- Počet plánovaných subgroup analýz by měl být uveden v protokolu studie (tedy před jejím zahájením).
- Pokud je to možné, před zahájením studie by měl být navržen optimální počet pacientů také pro plánované subgroup analýzy.
- Statistické zpracování by mělo být založeno na hodnocení **interakcí** sledovaných charakteristik s primárním endpointem, nikoliv na hodnocení endpointu v každé individuální podskupině zvlášť.
- Hladina významnosti pro testování hypotéz by v případě více testů měla být korigována dle jejich počtu.
- Výsledky subgroup analýz by měly být interpretovány jen jako hypotézy pro další výzkum.

# VII. Principy statistické analýzy dat



# Vzorkování a jeho význam ve statistice

- Statistika hovoří o realitě prostřednictvím vzorku!!!
- Statistické předpoklady korektního vzorkování



**Reprezentativnost:** struktura vzorku musí maximálně reflektovat realitu



**Nezávislost:** několikanásobné vzorkování téhož objektu nepřináší ze statistického hlediska žádnou novou informaci




# Velikost vzorku a přesnost statistických výstupů

Existuje skutečné rozložení a skutečný průměr měřené proměnné

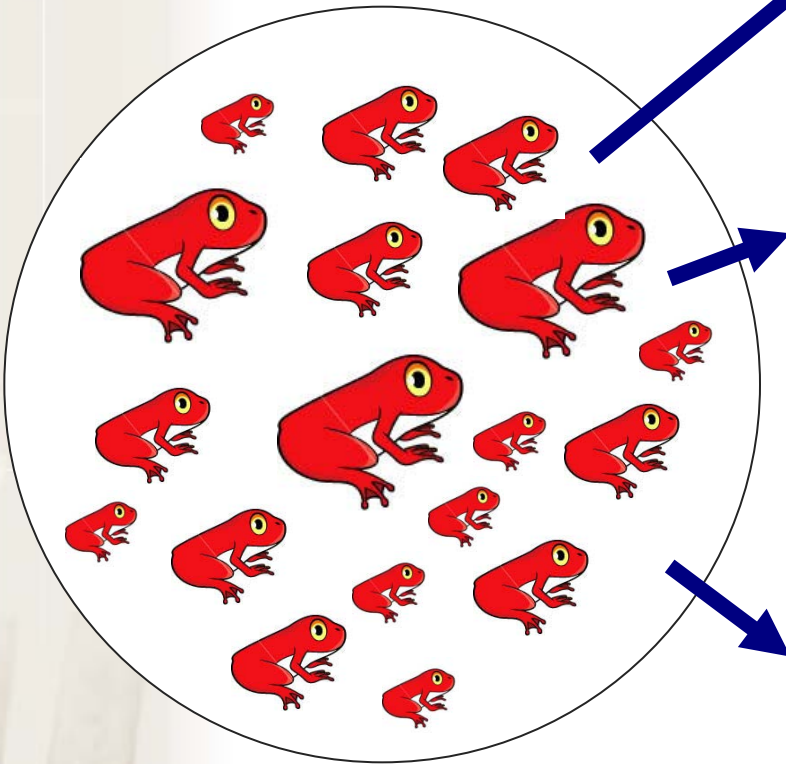
Z jednoho měření nezjistíme nic

Vzorek:  → ??????

Vzorek určité velikosti poskytuje odhad reálné hodnoty s definovanou spolehlivostí

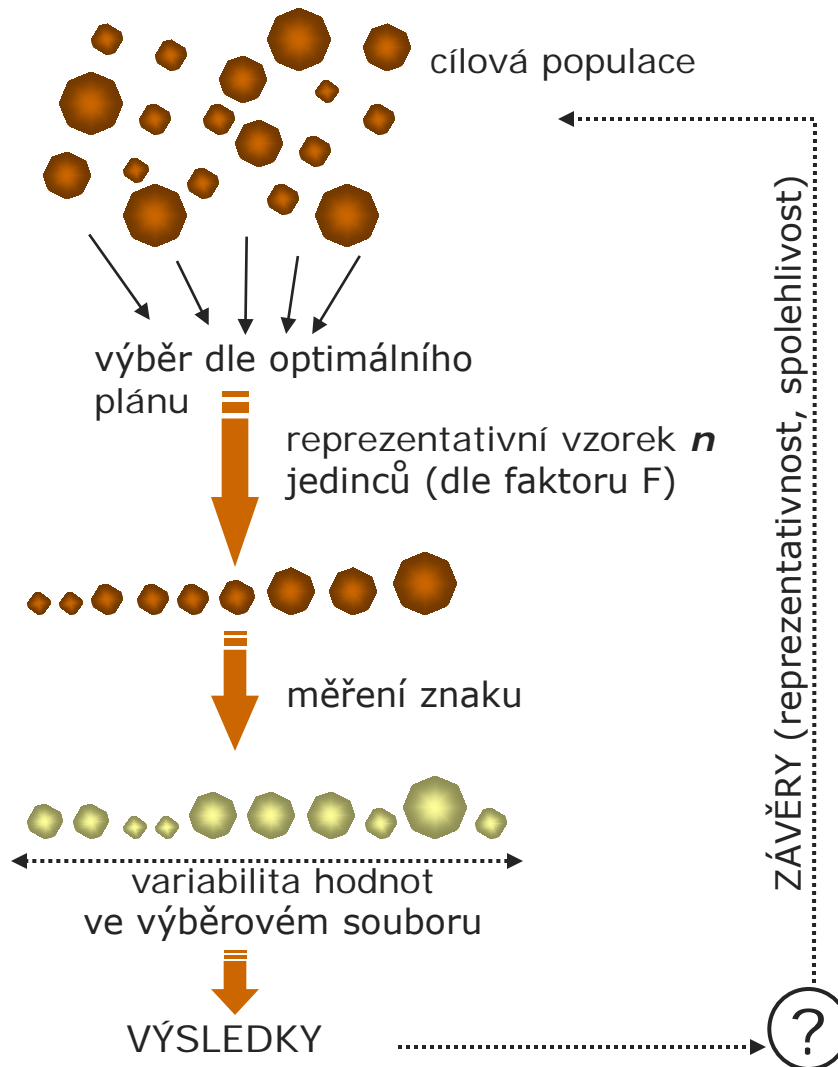
Vzorek:  → Odhad průměru atd.

Vzorkování všech existujících objektů poskytne skutečnou hodnotu dané popisné statistiky, nicméně tento přístup je ve většině případech nereálný.

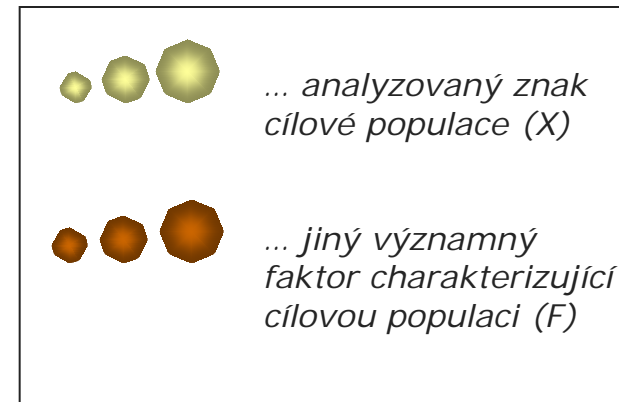


# Variabilita – její popis

**Účel analýzy:** Popisný (*např. Exploratorní studie*)

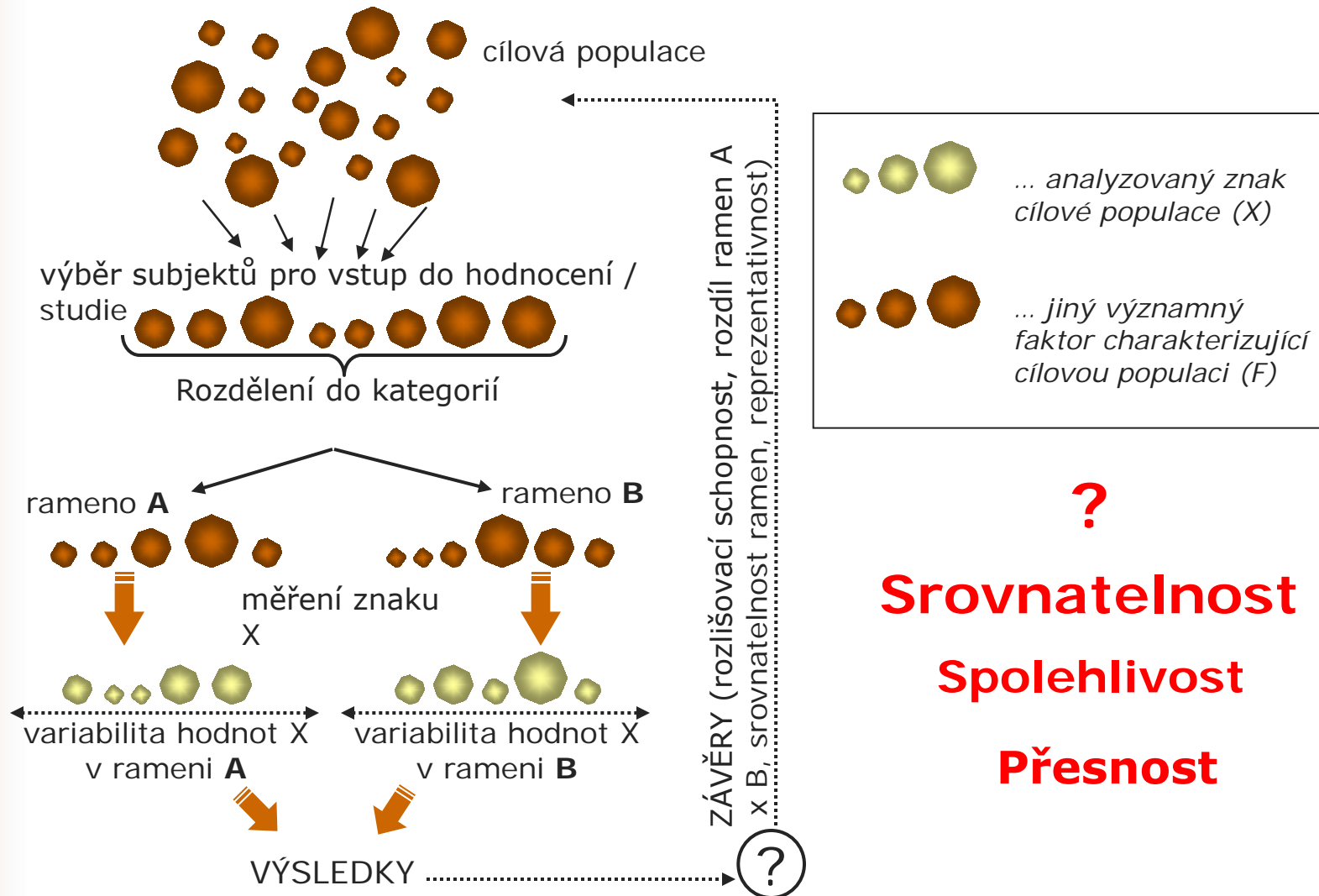


**?**  
**Reprezentativnost**  
**Spolehlivost**  
**Přesnost**



# Variabilita – srovnání dvou skupin

**Účel analýzy:** Srovnávací (např. Konfirmační studie)



# Testování hypotéz

- Testování hypotéz se zabývá rozhodováním o platnosti stanovených hypotéz na základě pozorovaných dat.
- **Platnost hypotéz ověřujeme pomocí statistického testu** – rozhodovacího pravidla, které každému náhodnému výběru přiřadí právě jedno ze dvou možných rozhodnutí –  $H_0$  nezamítáme nebo  $H_0$  zamítáme.
- **Nulová hypotéza** („null hypothesis“) – tvrzení o neznámých vlastnostech rozdělení pravděpodobnosti sledované náhodné veličiny (na cílové populaci). Může být tvrzením o parametrech rozdělení nebo tvaru rozdělení pravděpodobnosti.
- **Alternativní hypotéza** – tvrzení o neznámých vlastnostech rozdělení pravděpodobnosti sledované náhodné veličiny, které popírá platnost nulové hypotézy. Vymezuje, jaká situace nastává, když nulová hypotéza neplatí.

# Příklady – hypotézy

1. Urychluje použití antibiotika ve srovnání s použitím běžné dezinfekce hojení rány?

Střední doba hojení s antibiotiky:  $\theta_1$   $H_0 : \theta_1 = \theta_2$

Střední doba hojení bez antibiotik:  $\theta_2$   $H_1 : \theta_1 \neq \theta_2$

2. Je průměrný objem prostaty mužů nad 70 let stejný jako průměrný objem prostaty celé mužské populace?

Střední objem prostaty mužů nad 70 let :  $\theta_1$   $H_0 : \theta_1 = \theta_0$

Populační hodnota (konstanta):  $\theta_0$   $H_1 : \theta_1 > \theta_0$

# Proč nulová hypotéza vyjadřuje nepřítomnost efektu?

- Nulová hypotéza odráží fakt, že se něco nestalo nebo neprojevalo → je stanovena obvykle jako opak toho, co chceme experimentem prokázat.
- **Nulová hypotéza je postavena tak, abychom ji mohli pomocí pozorovaných hodnot vyvrátit.**
- Pro zamítnutí platnosti nulové hypotézy nám totiž stačí najít jeden příklad, kdy nulová hypotéza neplatí – tím příkladem má být náš náhodný výběr (naše pozorovaná data).
- Zamítnout nulovou hypotézu je jednodušší než nulovou hypotézu potvrdit.

# Hypotézy v klinických studiích

➔ Co chceme prokázat:

1. Experimentální léčba je lepší než standardní (placebo)
  - **hodnocení superiority** (*superiority trial*)
2. Experimentální léčba není horší než standardní (placebo)
  - **hodnocení non-inferiority** (*non-inferiority trial*)
3. Experimentální léčba je „stejně dobrá“ jako standardní (placebo)
  - **hodnocení ekvivalence** (*equivalence trial*)



# Hodnocení superiority

→ Chceme prokázat, že experimentální léčba je lepší než standardní (placebo).

→ Hypotézy:  $H_0 : \theta_1 = \theta_2$   $\longrightarrow$   $H_0 : \theta_1 - \theta_2 = 0$   
 $H_1 : \theta_1 \neq \theta_2$   $\longrightarrow$   $H_1 : \theta_1 - \theta_2 \neq 0$

→ K hodnocení superiority používáme standardní testování hypotéz.

→ Při nevýznamném výsledku nelze říci, že experimentální léčba je lepší než standardní, ale zároveň nelze říci, že experimentální léčba je stejná jako standardní.

# Hodnocení non-inferiority

→ Chceme prokázat, že experimentální léčba není horší než standardní (placebo) .

→ Hypotézy:

$$H_0 : \theta_1 - \theta_2 > \Delta_C$$

$$H_1 : \theta_1 - \theta_2 \leq \Delta_C$$

→ Kde  $\Delta_C$  je expertně stanovená klinicky/biologicky podstatná odchylka účinnosti/bezpečnosti obou srovnávaných léčiv.

→ Pro hodnocení non-inferiority otáčíme klasický koncept nulové a alternativní hypotézy.

# Hodnocení ekvivalence

→ Chceme prokázat, že experimentální léčba je srovnatelná se standardní (placebo).

→ Hypotézy:  $H_0 : \theta_1 - \theta_2 > \Delta_C$  nebo  $\theta_1 - \theta_2 < -\Delta_C$   
 $H_1 : -\Delta_C \leq \theta_1 - \theta_2 \leq \Delta_C$

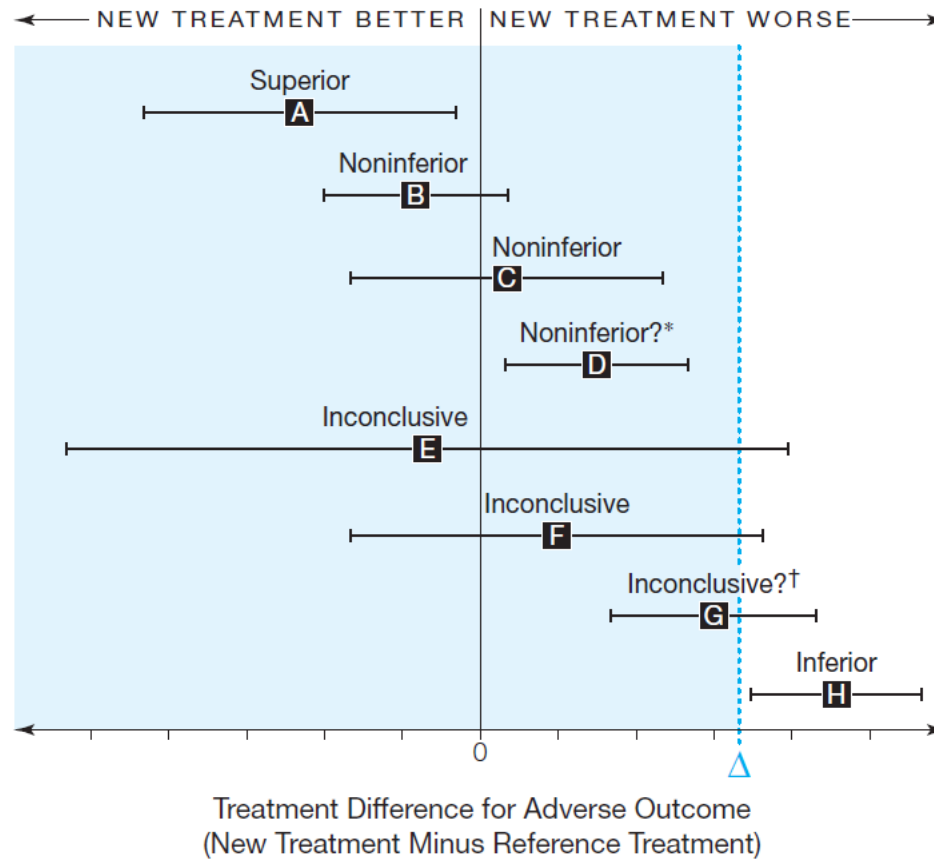
→ Kde  $\Delta_C$  je expertně stanovená klinicky/biologicky podstatná odchylka účinnosti/bezpečnosti obou srovnávaných léčiv.

→ Také pro hodnocení ekvivalence otáčíme klasický koncept nulové a alternativní hypotézy.

# Problém s hodnocením ekvivalence a non-inferiority

- ➔ **Pro hodnocení ekvivalence a non-inferiority nelze použít standardní přístup v testování hypotéz** – pro dané  $\Delta_C$  totiž lze vždy nalézt dostatečnou velikost souboru pacientů tak, abychom prokázali, že srovnávaná léčiva/postupy jsou vzhledem k hodnocené proměnné rozdílná.
- ➔ Pro vyhodnocení hypotéz se používá přístup pomocí intervalů spolehlivosti.
- ➔ Lze je však použít jako náhradu testování hypotéz?

# Superiorita a non-inferiorita



**Zdroj:** Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA. 2006 Mar 8;295(10):1152-60. Statistics and Informatics Services Group, Department of Reproductive Health and Research, World Health Organization, Geneva.

# Důvody pro hodnocení non-inferiority

- ➔ Je třeba vždy pečlivě zvážit, zda je vhodnější hodnocení non-inferiority nebo superiority.
  
- ➔ Důvody pro hodnocení non-inferiority:
  - 1. Nepředpokládáme žádný podstatný rozdíl mezi léčivy/postupy**
    - ➔ Nové léčivo/postup je ze stejné skupiny jako standardní léčivo/postup
  
  - 2. V nějakém ohledu předpokládáme rozdíl mezi léčivy/postupy**
    - ➔ Nové léčivo/postup má méně nežádoucích účinků
    - ➔ Nové léčivo/postup je méně invazivní
    - ➔ Nové léčivo/postup je levnější

# Hranice non-inferiority (*non-inferiority margin*)

- Stanovit hranici pro hodnocení non-inferiority není jednoduché, určitě by měla být stanovena nějakým odůvodnitelným způsobem.
- **Přímé stanovení**
  - Stanovení úvahou na základě přímého srovnání účinnosti/bezpečnosti srovnávaných látek – přímého stanovení maximálního možného snížení účinnosti/bezpečnosti.
- **Nepřímé stanovení**
  - Stanovení úvahou na základě srovnání účinnosti/bezpečnosti standardního postupu s placebem (historickou kontrolou) – stanovení provedeme na základě publikovaných dat o účinnosti standardního postupu.

# Hranice non-inferiority - příklad

## → Přímé stanovení

- Stanovení maximálního možného snížení účinnosti/bezpečnosti:
- Lék A má 85% účinnost v léčbě infekce *Helicobacter Pylori* a pro nás je maximální přípustné snížení účinnosti o 15 %. Lék B tedy musí prokázat vyšší účinnost než 70 %.

## → Nepřímé stanovení

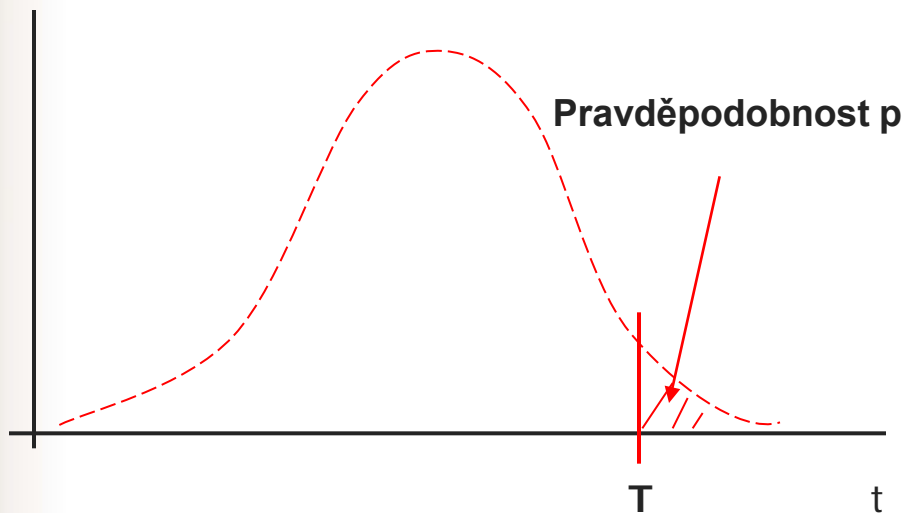
- Stanovení na základě publikovaných dat o účinnosti standardního postupu:
- Lék A snižuje riziko infekce *Helicobacter Pylori* o 50 % proti placebo (jinými slovy placebo zvyšuje riziko infekce o 100 %) a pro nás je důležité zachování alespoň poloviny této účinnosti. Lék B tedy musí prokázat relativní riziko vzniku infekce *Helicobacter Pylori* vzhledem k léku A ve výši maximálně 1,5.



## VIII. Statistická vs. klinická významnost

# Statistická významnost – p-hodnota

- Často je ve statistických softwarech jediným výstupem testování tzv. **p-hodnota**.
- Co vyjadřuje?
- **Vyjadřuje pravděpodobnost získání stejně velké nebo extrémnější hodnoty testové statistiky v případě, že platí nulová hypotéza, tj. v případě, že mezi skupinami není žádný rozdíl.**



Tedy p-hodnota vyjadřuje pravděpodobnost, že pozorovaný rozdíl je v rámci předpokládaného modelu dán pouze náhodou.

- Nejčastěji používanou hladinou, se kterou p-hodnotu srovnáváme je 0,05 (hladina významnosti  $\alpha$ ). Ta odpovídá přijímanému riziku falešně pozitivního výsledku.
- Při provádění více testů zároveň je nutné srovnávat p-hodnoty s korigovanou hladinou  $\alpha$ , abychom se vyhnuli nárůstu pravděpodobnosti získání falešně pozitivního výsledku.

# Praktická a statistická významnost

- Analytické výsledky studie nemusí odpovídat realitě a skutečnosti. Statistická významnost jednoduše nemusí znamenat příčinný vztah!
- Statistická významnost pouze indikuje, že pozorovaný rozdíl není náhodný (ve smyslu stanovené hypotézy).
- Stejně důležitá je i praktická významnost, tedy významnost z hlediska lékaře nebo biologa.
- Statistickou významnost lze ovlivnit velikostí vzorku.

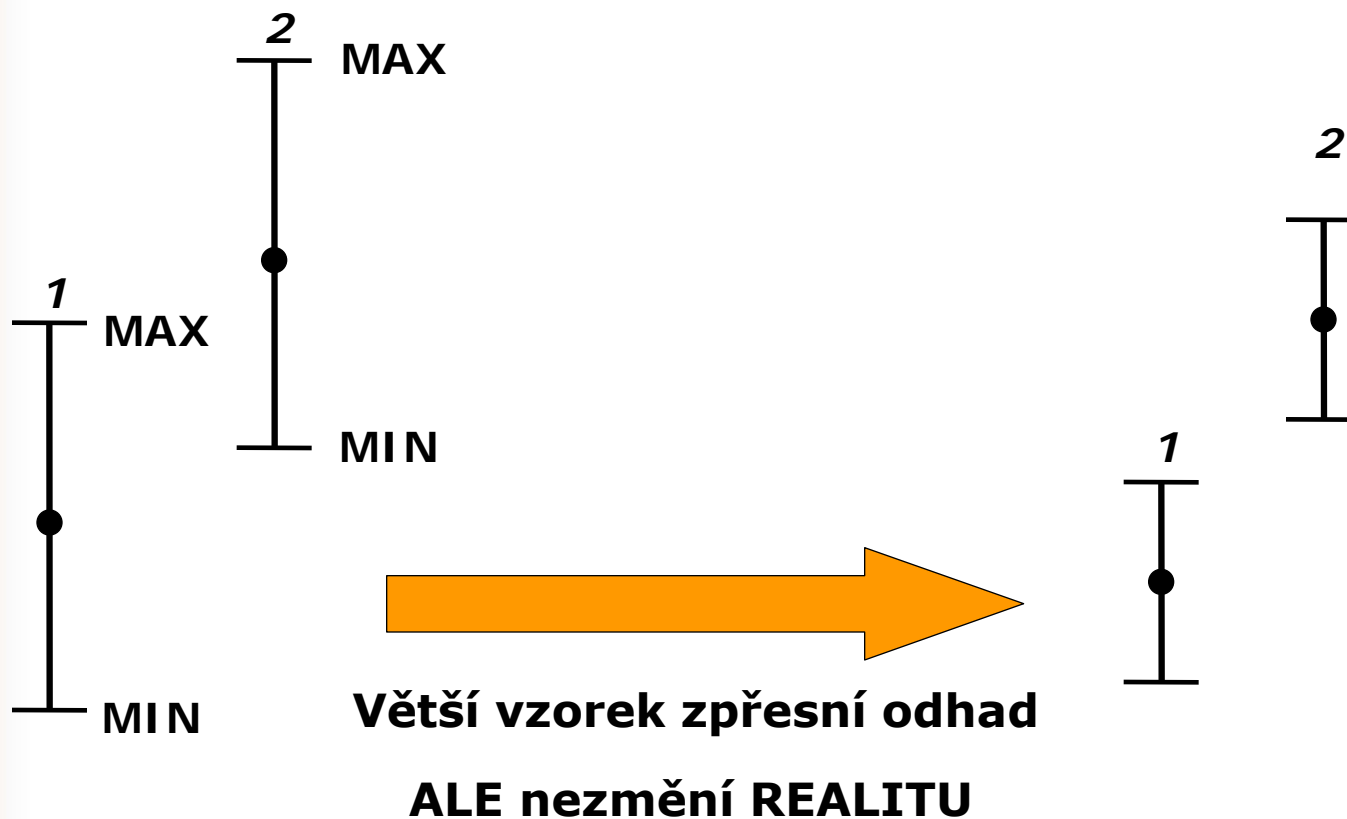
# Praktická a statistická významnost

- Samotná statistická významnost nemá žádný reálný význam, je pouze měřítkem náhodnosti hodnoceného jevu
- Pro vyhodnocení reálné významnosti je nezbytné znát i reálně významné hodnoty

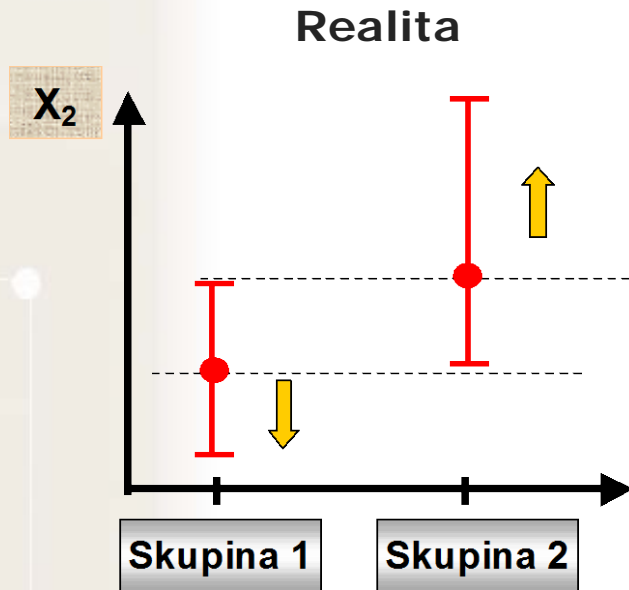
		<b>Praktická významnost</b>	
		ANO	NE
<b>Statistická významnost</b>	ANO	OK, praktická i statistická významnost je ve shodě, jednoznačný závěr	Významný výsledek je statistický artefakt velkého vzorku, prakticky nevyužitelné
	NE	Výsledek může být pouhá náhoda, neprůkazný výsledek	OK, praktická i statistická významnost je ve shodě, jednoznačný závěr

# Statistická významnost - indikační nástroj

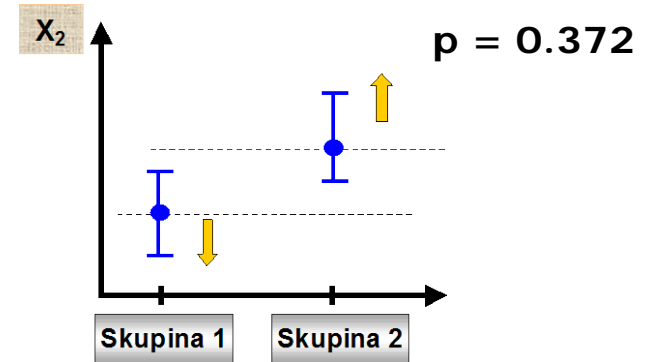
Statistická významnost indikuje, že pozorovaný rozdíl nevznikl náhodou. Otázkou je, zda jde o rozdíl věcně a klinicky podstatný a zda byl pozorován nezkresleně objektivními postupy.



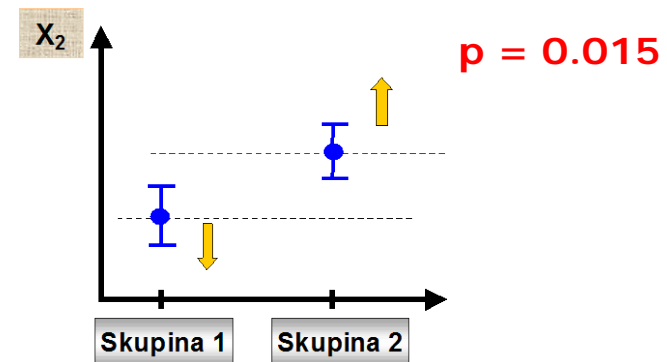
# Statistická významnost - indikační nástroj



**Srovnávací test ( $n_1$ )**



**Srovnávací test ( $n_2 > n_1$ )**

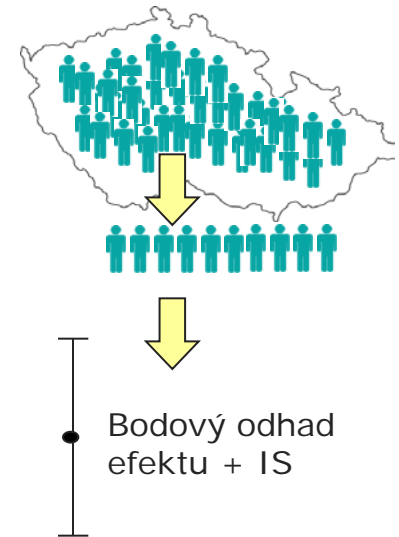
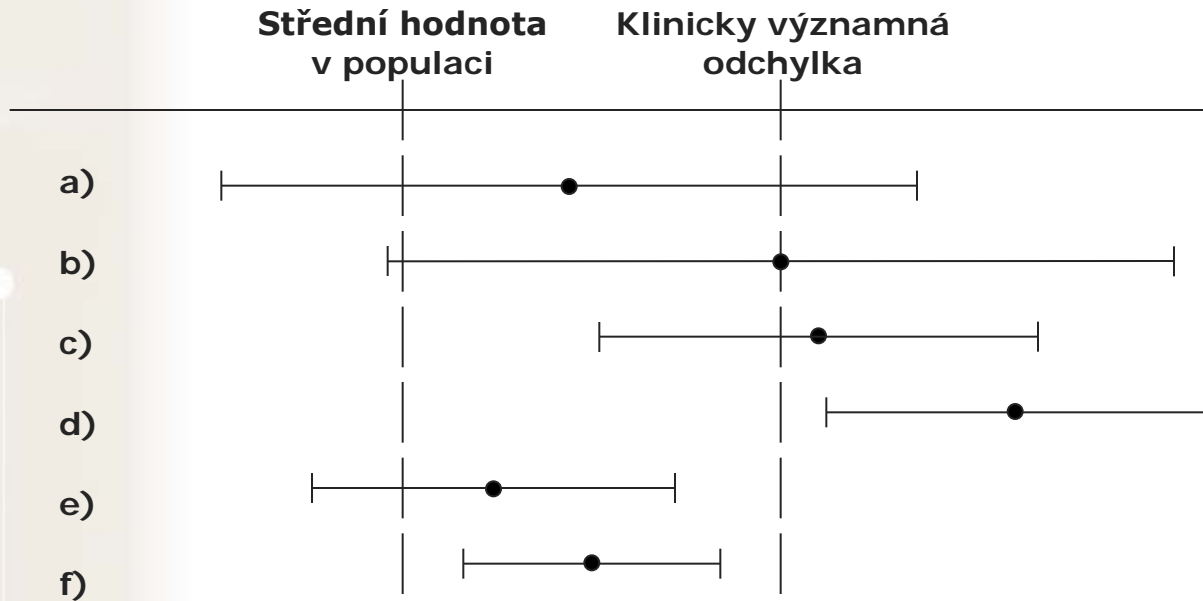


**Variabilita původních dat**



**Variabilita odhadu středových statistik**

# Statistická vs. klinická významnost



Možnost	Statistická významnost	Klinická významnost
a)	ne	možná
b)	ne	možná
c)	ano	možná
d)	ano	ano
e)	ne	ne
f)	ano	ne

# Důležité poznámky k testování hypotéz

- **Nezamítnutí nulové hypotézy neznamena 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.



# IX. Optimalizace velikosti vzorku

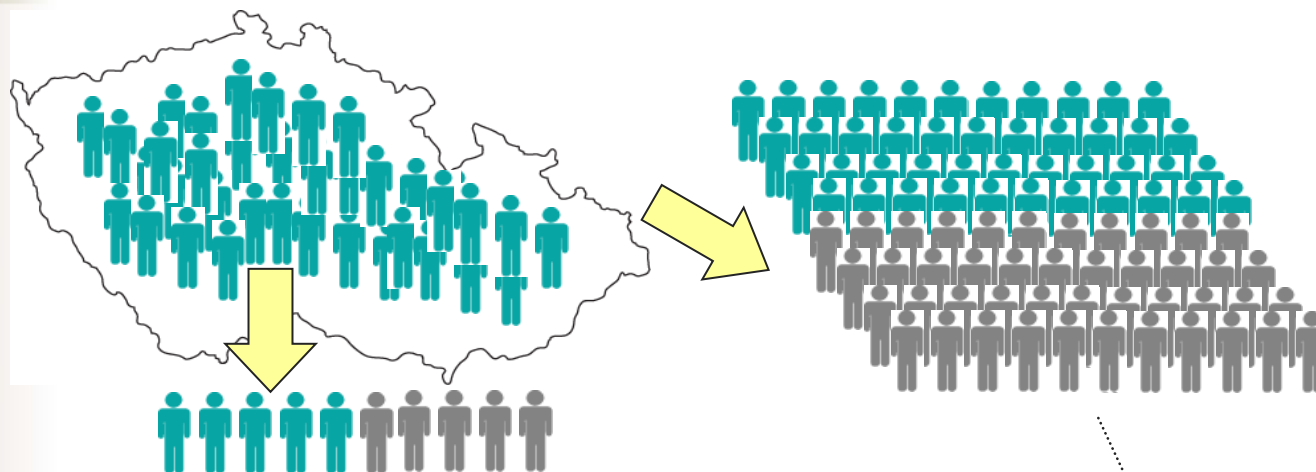
# Proč je důležité optimalizovat prospektivně velikost vzorku?

- **Etické aspekty** – nelze zbytečně léčit lidi
  - **Statistické vlastnosti** – při velkém N lze prokázat cokoliv
  - **Ekonomické aspekty** – zbytečné plýtvání prostředky
- ➔ Požadavky legislativních autorit (SÚKL, FDA) na výpočet velikosti vzorku

## Rizika neplánovaného počtu subjektů hodnocení:

- **Malý vzorek** – ztráta času, nemožnost prokázat rozdíl mezi srovnávanými skupinami pacientů
- **Velký vzorek** – ztráta času a prostředků, průkaz klinicky nevýznamného rozdílu mezi srovnávanými skupinami pacientů

# Vliv velikosti vzorku na výsledky testování



Dvě skupiny pacientů s nepatrným rozdílem v dané charakteristice, který ale není klinicky významný.

$$N_1 = 10, N_2 = 10$$
$$p = 0.797$$

$$N_1 = 100, N_2 = 100$$
$$p = 0.140$$

$$N_1 = 1000, N_2 = 1000$$
$$p < 0.001$$

**Statistická významnost způsobená velkým N**

# Jak tedy zní otázka při plánování klinických studií?

1. Kolik subjektů hodnocení potřebujeme pro korektní ověření klinické hypotézy (klinicky významného rozdílu)?
2. Nebudeme-li schopni tyto subjekty z různých důvodů zajistit, jaká je cena za snížení velikosti vzorku?

Tímto se zabývá  
power analýza

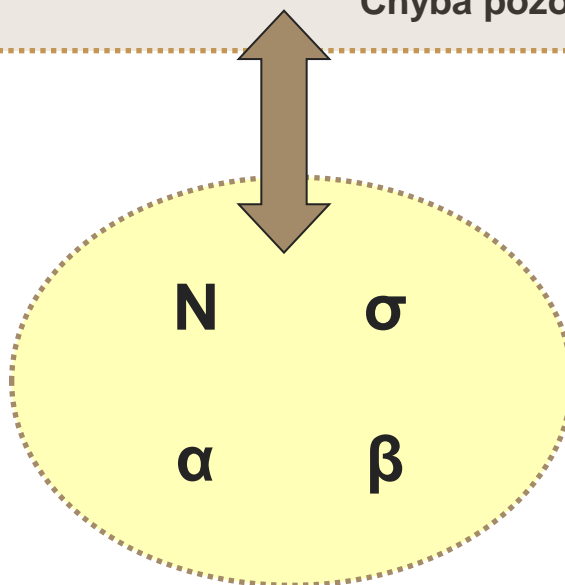
➤ Důvody pro nedostatečné N:

- Nedostatek prostředků
- Nedostatečná incidence – museli bychom čekat X let na dostatečný počet vhodných pacientů

# Princip testování hypotéz

- Hypotézy jsou testovány na základě testových statistik, které odrážejí, jak moc se náš výsledek liší od počátečních předpokladů = nulové hypotézy.

$$\text{Testová statistika} = \frac{\text{Pozorovaná hodnota} - \text{Očekávaná hodnota}}{\text{Chyba pozorované hodnoty}}$$



+ Velikost klinicky významného rozdílu

Testová statistika obsahuje (tj. vychází) z určitých parametrů. Pokud chceme vypočítat jeden z nich, ostatní musíme znát (expertně odhadnout).

# Princip výpočtu velikosti vzorku

K  
L  
I  
N  
I  
K

Klinické zadání



Klinická hypotéza



Odpovídající testová statistika

$\sigma$   
 $\alpha$   
 $1-\beta$



**Výpočet velikosti vzorku**

S  
T  
A  
T  
I  
S  
T  
I  
K

Výpočet velikosti vzorku je tedy vždy individuální záležitostí, která je šitá na míru danému klinickému hodnocení.

# Power analýza a optimalizace velikosti vzorku

- ☑ Power analýza (analýza síly testu) a optimalizace velikosti vzorku (sample size estimation) jsou dvě strany téže mince.
- ☑ Obě vycházejí z testování hypotéz, tedy hypotézy a příslušné testové statistiky, jednou však máme jako předpoklad požadovanou sílu testu a chceme optimalizovat  $N$ , podruhé jsme limitováni  $N$  a ptáme se, jaké jsme v našich podmínkách schopni dosáhnout síly testu.

**Power analýza**



**Odhad velikosti vzorku**



- ☑ Optimalizace velikosti vzorku může také souviset s dosažením určité přesnosti v odhadu cílového parametru.

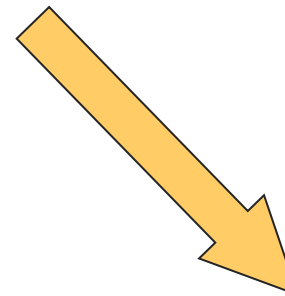
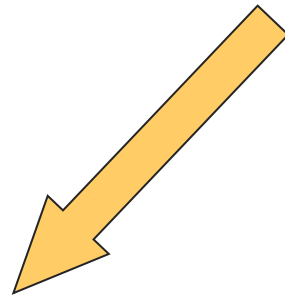
**Dosažení určité přesnosti  
(precision analysis)**

# Odhad velikosti vzorku – srovnání dvou výběrů I

- ☑ Cílem je ověřit klinickou hypotézu o nenulovém rozdílu ve sledovaném parametru mezi dvěma skupinami pacientů:

$$H_0: X_1 = X_2 + k$$

**Klinicky významný rozdíl**



## Spojitě proměnné

nabývající hodnot v určitém intervalu

- tlak,
- hladina cholesterolu
- počet leukocytů

## Binomické proměnné

nabývající jedné z možností úspěch/neúspěch

- snížení hladiny cholesterolu
- incidence zlomeniny



# Odhad velikosti vzorku – srovnání dvou výběrů II

- ☑ V případě **spojitých proměnných** vychází výpočet z následujících charakteristik:
  1.  $\alpha$  – požadovaná hladina významnosti,  $1-\beta$  – požadovaná síla testu
  2. **odhad variability** měřené proměnné v kontrolní skupině
  3. **odhad variability** měřené proměnné v experimentální skupině
  4. **klinicky významný rozdíl**, který chceme identifikovat
  
- ☑ V případě **binomických proměnných** vychází výpočet z následujících charakteristik:
  1.  $\alpha$  – požadovaná hladina významnosti,  $1-\beta$  – požadovaná síla testu
  2. **odhad incidence** sledované charakteristiky v kontrolní skupině
  3. **klinicky významný rozdíl**, který chceme identifikovat

# Srovnání dvou výběrů – příklad 1

- ☑ Chceme srovnat účinnost dvou preparátů snižujících hladinu cholesterolu v krvi. Primárním cílovým parametrem je tedy hladina LDL-cholesterolu.
- ☑ Předpokládáme, že rozdíl v procentuálním úbytku cholesterolu roven 8% je klinicky zajímavý, s tím, že v obou skupinách předpokládáme směrodatnou odchylku procentuálních úbytků 15%.

Rameno A (lék A); Rameno B (lék B)

$$\alpha = 0.05$$

$$\beta = 0.10 \rightarrow \text{síla testu: } 1 - \beta = 0.90$$

odhad variability v 1.skupině = 15%

odhad variability v 2.skupině = 15%

klinicky významný rozdíl mezi rameny = 8%

**N = 75 pacientů v jednom rameni**

# Srovnání dvou výběrů – příklad 2

- ☑ Chceme srovnat účinnost dvou preparátů pro léčbu infekce dýchacích cest. Primárním cílem studie je odpověď pacienta na podání léku (ano/ne).
- ☑ Předpokládáme, že standardní přípravek účinkuje u 80% pacientů s tím, že klinicky významný posun by bylo dosažení odpovědi u 90% pacientů, tedy zvýšení účinnosti o 10% při použití nového léku oproti standardu.

Rameno A (standardní lék): účinnost = 80%

Rameno B (nový lék): požadujeme účinnost = 90%

$\alpha = 0.05$

$\beta = 0.20 \rightarrow$  síla testu:  $1-\beta = 0.80$

odhad incidence úspěchů v kontrolní skupině = 80%

klinicky významný rozdíl mezi rameny = 10%

**N = 199 pacientů  
v jednom rameni**

pro  $\beta = 0.10 \rightarrow$  síla testu:  $1-\beta = 0.90$  je odhad počtu pacientů v 1 rameni = 266

# Odhad velikosti vzorku – analýza přežití

- ✓ V klinických studiích jsou přijímány jako klíčové parametry charakterizující přežití

## MEDIÁN PŘEŽITÍ a PRAVDĚPODOBNOST PŘEŽITÍ V DANÉM ČASE

- ✓ V případě srovnání mediánů je nulová hypotéza:

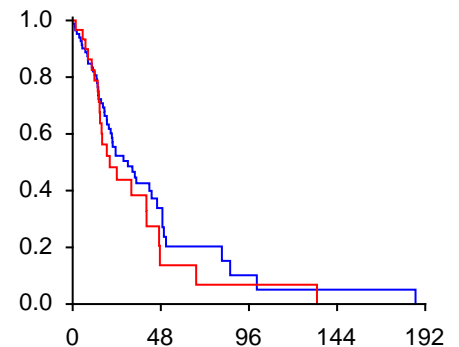
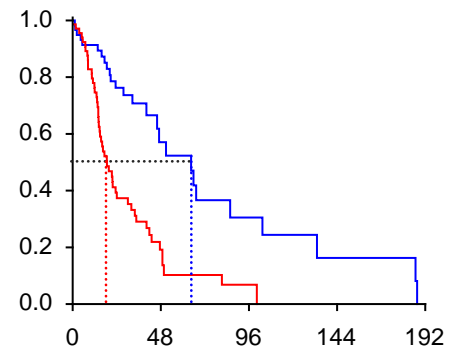
$$H_0: \text{medián}_1 = \text{medián}_2$$

Výpočet N na základě očekávaného **hazard ratio**, tedy odhadu podílu rizika v experimentální / kontrolní skupině.

- ✓ V případě srovnání křivek přežití je nulová hypotéza:

$$H_0: S_1(t) = S_2(t)$$

Výpočet N na základě odhadu **četnosti výskytu sledovaných událostí** v obou skupinách.



# Odhad velikosti vzorku – shrnutí

- ☑ Optimalizace velikosti vzorku je nutná z hlediska korektního statistického zpracování výsledků a dostatečné síle studie prokázat klinicky zajímavý rozdíl.
- ☑ Optimalizace velikosti vzorku je založena na principu testování klinických hypotéz.
- ☑ Každou hypotézu je nutné vyjádřit ve smyslu testové statistiky, která zahrnuje parametry nutné pro ověření hypotézy.
- ☑ Parametry nezbytné pro výpočet optimální velikosti vzorku závisí na typu studovaného problému, pokud o nich nemáme apriorní informaci, lze je odhadnout na základě předchozích studií nebo expertní znalosti.

# X. Aspekty korektního zpracování dat KHL

# Na co si dávat při interpretaci pozor...

- I přes požadavky legislativních autorit na přípravu a průběh klinického hodnocení léčiv je možné v průběhu zpracování a prezentace výsledků studie postupovat nekorektně.
  
- Klíčová témata:
  1. Definice primárního cíle studie
  2. Zobecnění výsledků
  3. Adekvátní kontrolní skupina
  4. Absolutní vs. relativní hodnocení účinnosti
  5. Klinická významnost výsledků
  6. Analýzy podskupin

# 1. Definice primárního cíle studie

- **Primární cíl studie by měl být vždy stanoven jasně, tedy formou klinicky významného rozdílu v sledovaném klinickém parametru (úmrť, hospitalizace, % CMP, % IM) nebo parametru, o kterém není nejmenších pochyb, že s ním souvisí.**
- **Problematické jsou zejména následující:**
  - **Náhradní cílové parametry (*surrogate endpoints*) – souvislost s klinickým parametrem může být diskutabilní.**
  - **Souhrné cílové parametry (*composite endpoints*) – lze obtížně kvantifikovat ovlivnění jednotlivých parametrů, které mohou být důležité v případě individuálního pacienta.**



## 2. Zobecnění výsledku

- **Souvisí s externí validitou studie.**
- **Výsledky klinických studií nelze vztahovat na jinou populaci pacientů než na tu, která odpovídá souboru hodnocených pacientů.**
- **Př. Je-li sledovaný preparát účinný z hlediska snížení rizika celkové mortality u kardiologických pacientů s normální funkcí ledvin, NELZE účinnost daného preparátu předpokládat u skupiny pacientů se stejným kardiologickým problémem a dysfunkcí ledvin.**
- **Na druhou stranu, pokud se na selektované skupině pacientů ukáže preparát jako neúčinný, neznamená to, že je neúčinný také u skupiny, která ve studii nebyla uvažována (např. z hlediska bezpečnosti).**

# 3. Adekvátní kontrolní skupina

- **Srovnání účinnosti léčby by mělo být vždy prováděno vzhledem ke kontrolní skupině, která odpovídá klinické praxi nebo standardnímu léčebnému postupu.**
- **Je nepřijatelné srovnávání účinnosti experimentálního ramene s placebem (není-li standardem) nebo nestandardně nízkou dávkou jiného či stejného preparátu.**
- **Př.**  
KHL1: srovnání 80mg/den látky A vs. 40mg/den látky B.  
KHL2: srovnání 80mg/den látky A vs. 10mg/den látky A.  
**Chybí srovnání účinnosti 80mg/den látky A vs. 40mg/den látky A !**

# 4. Absolutní vs. relativní hodnocení účinnosti

- **Relativní vyjádření účinnosti preparátu (např. RRR – relativní redukce rizika) by měla být vždy doprovázena absolutním vyjádřením účinnosti (např. ARR – absolutní redukce rizika; NNT – „number needed to treat“).**
- **Př. Srovnání účinnosti ve smyslu prevence CMP u kardiaků.**
  - KHL1: výskyt CMP ve skupině A je 12 %, ve skupině B je 20 %.**
    - **RRR = 40 %; ARR = 8 %.**
  - KHL2: výskyt CMP ve skupině A je 0,9 %, ve skupině B je 1,5 %.**
    - **RRR = 40 %; ARR = 0,6 %.**

**Výsledkem je rozdílný přínos léčby při stejném RRR.**

# NNT a absolutní vs. relativní četnost

→ **Příklad:** Srovnání účinnosti léčiva ve smyslu prevence CMP u kardiaků.

Studie 1: Výskyt CMP ve skupině A je 12 %, ve skupině B je 20 %.  
Relativní změna v účinnosti = **40 %**; absolutní změna = **8 %**.

$$\longrightarrow \text{NNT} = \frac{1}{0,08} = \frac{100}{8} = 12,5$$

**NNT = Pro snížení počtu událostí o 1 je třeba léčit 13 pacientů.**

Studie 2: výskyt CMP ve skupině A je 0,9 %, ve skupině B je 1,5 %.  
Relativní změna v účinnosti = **40 %**; absolutní změna = **0,6 %**.

$$\longrightarrow \text{NNT} = \frac{1}{0,006} = \frac{100}{0,6} = 166,7$$

**NNT = Pro snížení počtu událostí o 1 je třeba léčit 167 pacientů.**

# 5. Klinická významnost výsledků

- **Statistická významnost výsledku nemusí odpovídat klinické významnosti výsledku, která je velmi často opomíjena.**
- **Př. Studie prokázala při léčbě preparátem A (proti B) statisticky významné snížení rizika mortality u starších pacientů s hypertenzí: RRR = 11 %, p = 0,05.**  
**ALE v absolutním měřítku jsou výsledky klinicky daleko méně zajímavé: ARR = 0,4 % → NNT = 250.**  
**Pro snížení počtu úmrtí o 1 by bylo nutné léčit 250 pacientů.**
- **Klinická významnost souvisí s optimalizací velikosti vzorku.**

# 6. Analýzy podskupin

- **V kontextu klinické studie je hlavním výsledkem studie vyhodnocení primárního cíle studie na souboru pacientů jako celku. Na základě analýz primárního cíle studie na podskupinách souboru pacientů nelze vytvářet klinická doporučení.**
- **Zásadním je zde problém násobného testování hypotéz: provádíme více testů zároveň (musíme výsledné p-hodnoty korigovat na počet provedených testů).**
- **Př. Pravděpodobnost falešně pozitivního výsledku při**
  - N = 2 testech zároveň je 9,8 %.**
  - N = 5 testech zároveň je 22,6 %.**
  - N = 10 testech zároveň je 40,1 %.**
  - N = 20 testech zároveň je 64,2 %.**

# XI. Příklady

## Randomized clinical trial evaluating elective laparoscopic appendicectomy for chronic right lower-quadrant pain

**Background:** It is questionable whether elective appendicectomy can effectively reduce persistent or recurrent right lower-quadrant abdominal pain due to chronic or recurrent appendicitis.

The primary outcome measure was pain scored by the patient at 6 months after operation in the presence of the surgical resident. At each follow-up appointment patients

### Vliv selektivní apendektomie na bolest břicha

dairy notes. Pain scores were: 1, pain unchanged (or even worse); 2, improvement with a remarkable reduction of pain, but not completely pain free; and 3, completely pain free, no more right lower abdominal complaints. After



# Odhad velikosti vzorku a randomizace

## Statistical analysis

Calculation of the required sample size was based on the assumption that the placebo effect of laparoscopy alone would result in improvement in pain in 30 per cent, whereas appendicectomy would result in improvement in 75 per cent. A two-sided test with an  $\alpha$  level of 0.05 and 90 per cent power required 20 patients per group to reach significance. The analysis of primary outcome was performed on an intention-to-treat basis. Proportions of patients in the two groups with no change in pain scores (score 1 at 6 months) were compared by means of Fisher's exact test. Two-tailed  $P < 0.050$  was considered statistically significant. Relative risks with 95 per cent confidence intervals (c.i.) were calculated; a relative risk of more than 1.0 indicated a beneficial effect of appendicectomy on pain level. Finally, the number

Předpoklad zlepšení bolesti u 30 %, respektive 75 % pacientů.

$\alpha=0,05$  (riziko falešné positivity)

a

$1-\beta=0,90$  (síla testu)



Odhad celkového počtu 40 randomizovaných pacientů = 20 v každé z obou skupin.

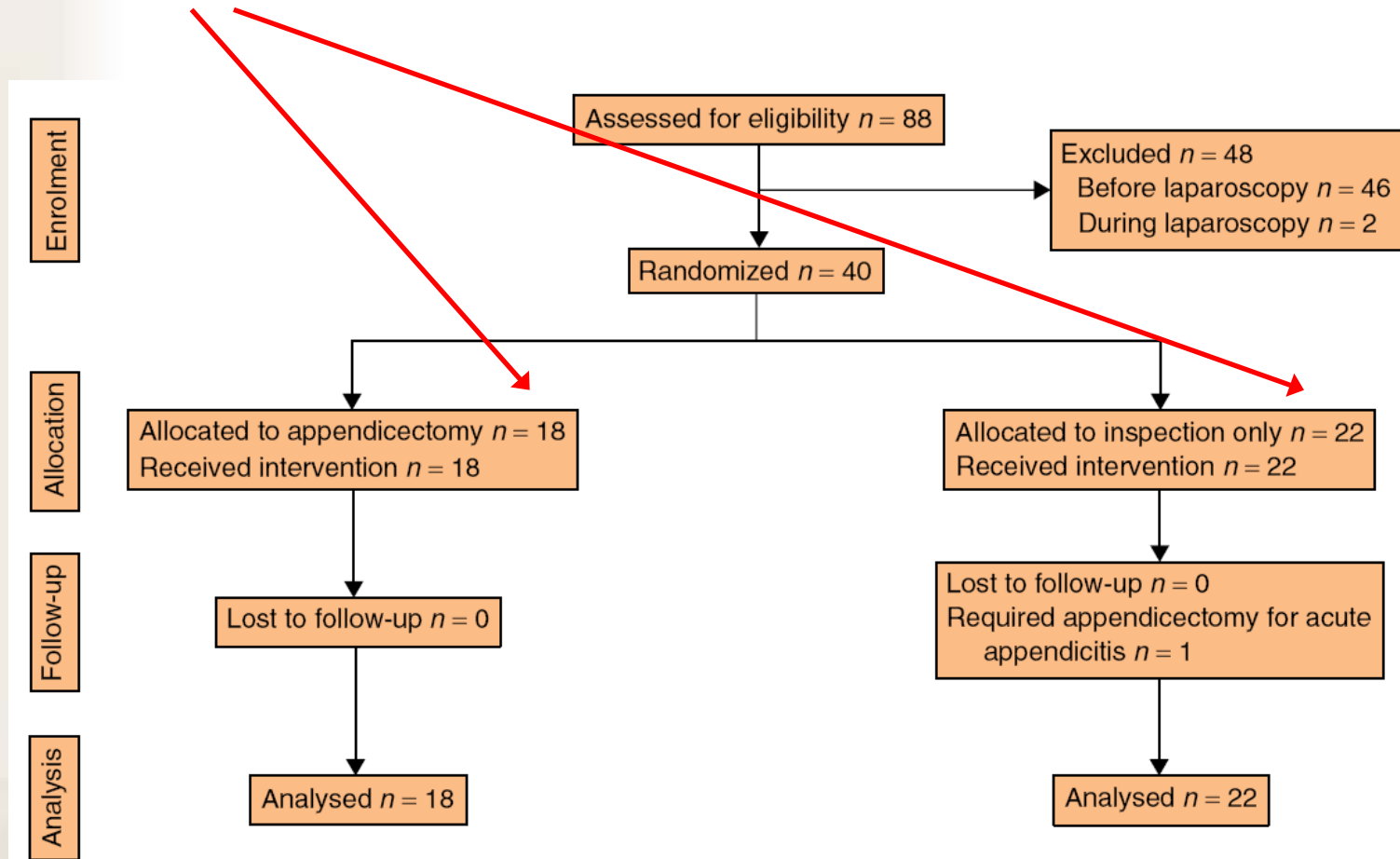
Provedena permutační bloková randomizace, která zajišťuje rovnoměrné rozdělení pacientů do skupin.

## Randomization procedure

Randomization was performed by computer and sequencing was based on permuted blocks of eight for a single centre without further stratification. Information on appendix

# Schéma studie

**! Nerovnoměrné rozdělení pacientů ve sledovaných skupinách !**



**+ Ve studii chybí charakteristika zařazených pacientů**

# Hodnocení výsledků

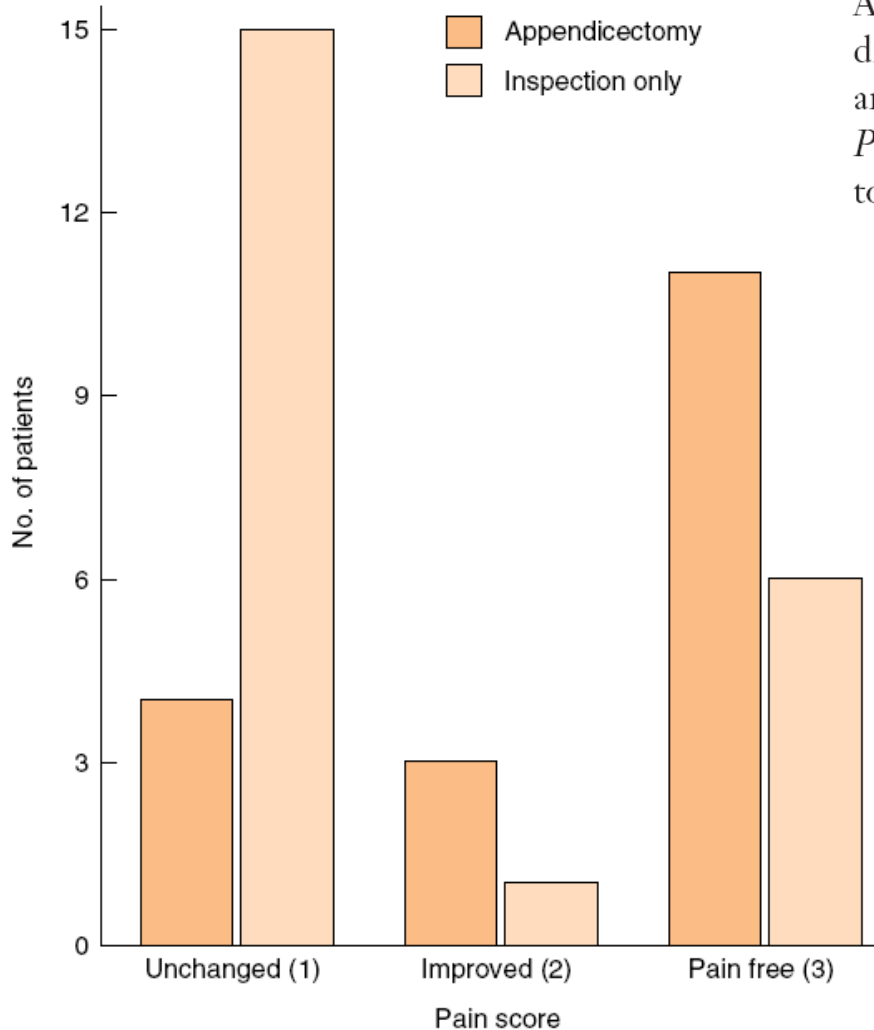


Fig. 2 Distribution of pain scores in patients whose appendix was removed during laparoscopy and those in whom it was left *in situ*

A significantly higher proportion of patients in the appendicectomy group than in the inspection-only group had an improvement in pain (14 of 18 *versus* seven of 22;  $P = 0.005$ ). The relative risk was 2.4 (95 per cent c.i. 1.3 to 4.0), indicating that patients who had an appendicectomy

- Graf je zavádějící protože zobrazuje pouze celkové počty a ne % srovnání – ve skupině bez apendektomie je ale více osob!
- Hodnocená proměnná je velmi subjektivní a může být ovlivněna charakteristikami pacientů!
- Výsledné hodnocení je založeno pouze na základě dvou kategorií „zlepšení“ – při použití tří kategorií, jako jsou v grafu, by mohlo dojít k posunu ve statistické významnosti!
- Nelze říci, že by výsledky nebyly relevantní, nicméně určité kroky bez vysvětlení vzbuzují podezření!

# Biventricular Versus Conventional Right Ventricular Stimulation for Patients With Standard Pacing Indication and Left Ventricular Dysfunction

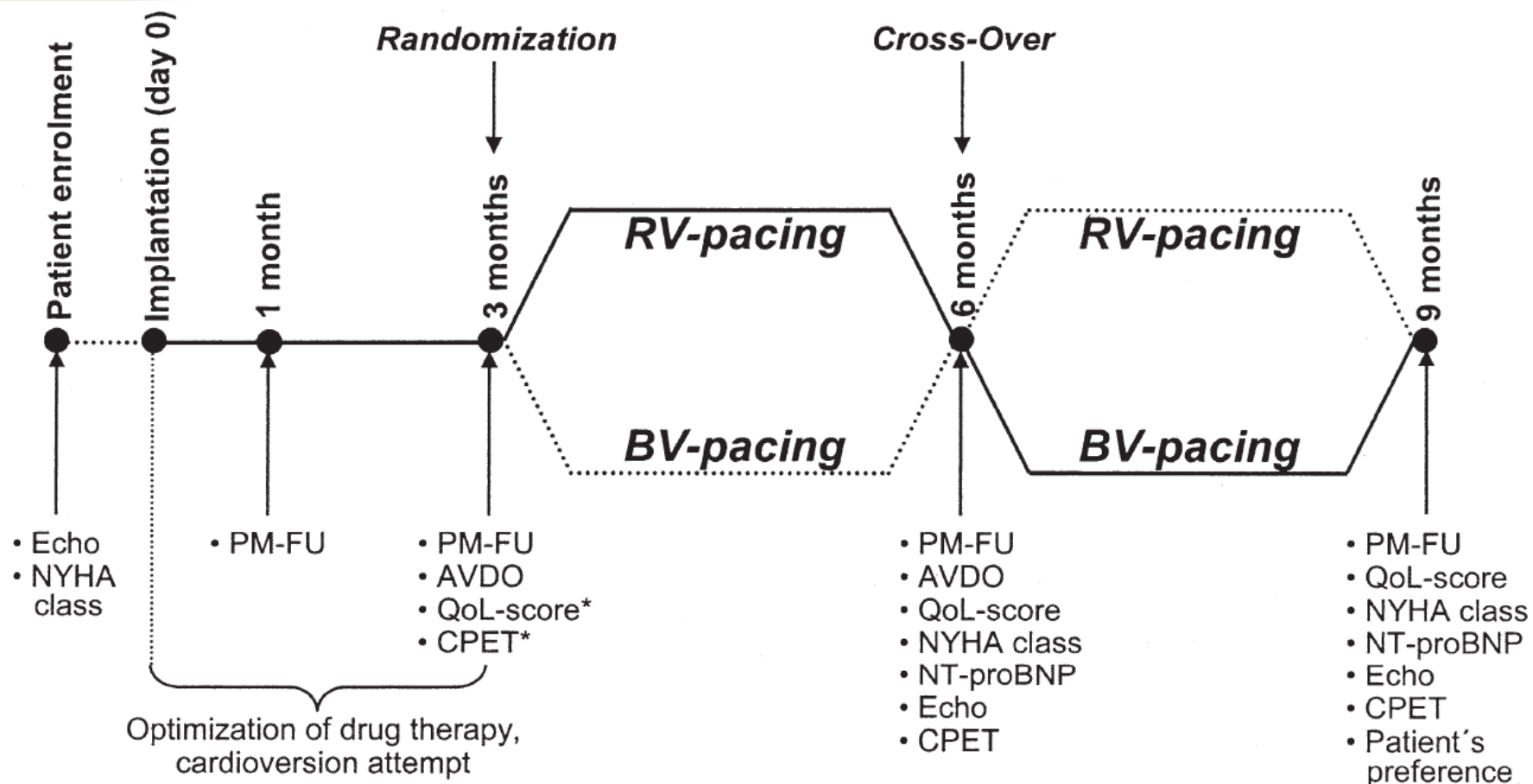
## The Homburg Biventricular Pacing Evaluation (HOBIPACE)

- OBJECTIVES** The Homburg Biventricular Pacing Evaluation (HOBIPACE) is the first randomized controlled study that compares the biventricular (BV) pacing approach with conventional right ventricular (RV) pacing in patients with left ventricular (LV) dysfunction and a standard indication for antibradycardia pacing in the ventricle.
- BACKGROUND** In patients with LV dysfunction and atrioventricular block, conventional RV pacing may yield a detrimental effect on LV function.
- METHODS** Thirty patients with standard indication for permanent ventricular pacing and LV dysfunction defined by an LV end-diastolic diameter  $\geq 60$  mm and an ejection fraction  $\leq 40\%$  were included. Using a prospective, randomized crossover design, three months of RV pacing were compared with three months of BV pacing with regard to LV function, N-terminal pro-B-type natriuretic peptide (NT-proBNP) serum concentration, exercise capacity, and quality of life.

three-month period. The study had three primary end points: LV end-systolic volume, LV ejection fraction, and peak oxygen consumption. Functional class according to the

# Design studie

**Study protocol.** HOBIPACE was a monocentric, prospective, randomized, single-blinded crossover comparison between RV and BV pacing. It was an investigator-driven trial



# Odhad velikosti vzorku a statistický test

Sample size calculations were focused on primary end points assuming an absolute 4% increment in LV ejection fraction, a relative decrease in LV end-systolic volume of 15%, and an increase in peak oxygen uptake of 1.5 ml/min/kg with BV compared with RV pacing. Considering an estimated coronary sinus lead failure rate of 10% and a lost-to-follow-up rate of an additional 10%, the minimum number of patients required to detect a highly significant ( $p < 0.01$ ) difference between RV and BV pacing with a power of 90% was 21.

**Statistics.** A two-tailed  $p$  value of  $<0.05$  was considered significant. Differences between pre-operative data, RV and BV pacing periods were assessed using a two-way univariate analysis of variance (ANOVA) for repeated measurements with post-hoc Student-Newman-Keuls test. The stimula-

# Hodnocení cílových parametrů

**Impact of BV pacing. PRIMARY END POINTS.** All three primary end points for the comparison between RV and BV pacing were met (Figs. 4A, 4B, and 5). When compared with RV pacing, BV stimulation reduced LV end-systolic volume by 17% ( $p < 0.001$ ), increased LV ejection fraction by 22% ( $p < 0.0002$ ), and peak oxygen consumption by 12% ( $p < 0.0003$ ).

**LV function and structure.** In comparison with RV pacing, BV stimulation improved all parameters of LV dimension and function as well as left atrial diameters (Table 3).

# Dodatečné hodnocení klinických parametrů

**Table 3.** Comparison of Functional Class According to the NYHA Classification, Echocardiographic, and Electrocardiographic Parameters Before Device Implantation, After Three Months of RVP, and After Three Months of BVP

Parameters	Pre-Operative	RVP	BVP
NYHA functional class	3.0 ± 0.6**	2.5 ± 0.7†	1.9 ± 0.6†††
LV end-diastolic diameter (mm)	69.9 ± 8.5	68.7 ± 9.1†††	64.9 ± 8.9†††
LV end-systolic diameter (mm)	59.0 ± 9.6	58.1 ± 11.1††	53.6 ± 11.1†††
LV end-diastolic volume (ml)	236.1 ± 74.8*	215.6 ± 76.2†	196.3 ± 77.3†††
LV end-systolic volume (ml)	177.3 ± 68.7*	160.2 ± 73.4†††	133.1 ± 66.5†††
LV ejection fraction (%)	26.1 ± 7.8*	28.5 ± 11.2†††	34.8 ± 8.9†††
Left atrial diameter (mm)	55.0 ± 7.4	53.9 ± 9.3†	51.7 ± 8.1††
Grade of mitral regurgitation	1.5 ± 0.6	1.4 ± 0.7	1.1 ± 0.5†
Cardiac index (l/min/m <sup>2</sup> )	2.15 ± 0.65	2.00 ± 0.40††	2.31 ± 0.42
LV dp/dt (mm Hg/s)	509 ± 216	619 ± 138†	789 ± 251†
LV muscle mass index (g/m <sup>2</sup> )	185.1 ± 45.5	178.8 ± 41.2	168.2 ± 28.5††
LV hypertrophy index (%)	30.8 ± 5.3	31.6 ± 6.3††	34.8 ± 7.5†††
Duration of the QRS interval (ms)	174 ± 42**	193 ± 25†††	151 ± 21†††
LV pre-ejection time (ms)	174 ± 37**	204 ± 35†	184 ± 33
RV pre-ejection time (ms)	125 ± 34**	157 ± 36†	177 ± 34†††
Pre-ejection time difference (ms)	49 ± 31	47 ± 26†††	8 ± 24†††
LV Z-ratio (%)§	67.6 ± 9.7**	73.3 ± 9.2†	77.6 ± 6.8†††
LV Tei index§	1.11 ± 0.39**	0.91 ± 0.38††	0.72 ± 0.24†††
RV Z-ratio (%)§	75.1 ± 10.1**	83.1 ± 10.2	84.2 ± 7.8††
RV Tei index§	0.85 ± 0.48***	0.55 ± 0.39	0.49 ± 0.26†††
LV synchrony (%)	58.6 ± 14.9	61.1 ± 15.1†††	79.2 ± 8.8†††
LVS-SD (%)	20.6 ± 7.4	17.8 ± 6.2††	13.5 ± 5.2†††

\*/\*\*/\*\*/\*\*\*\* p < 0.05/p < 0.01/p < 0.001 for preoperative data vs. RVP; †/††/††† p < 0.05/p < 0.01/p < 0.001 for RVP vs. BVP; ‡/‡†/‡†† p < 0.05/p < 0.01/p < 0.001 for preoperative data vs. BVP; §Measured only in patients without atrial fibrillation.

BVP = biventricular pacing; LV = left ventricular; LVS-SD = standard deviation of LV synchrony; NYHA = New York Heart Association; RV = right ventricular; RVP = right ventricular pacing.

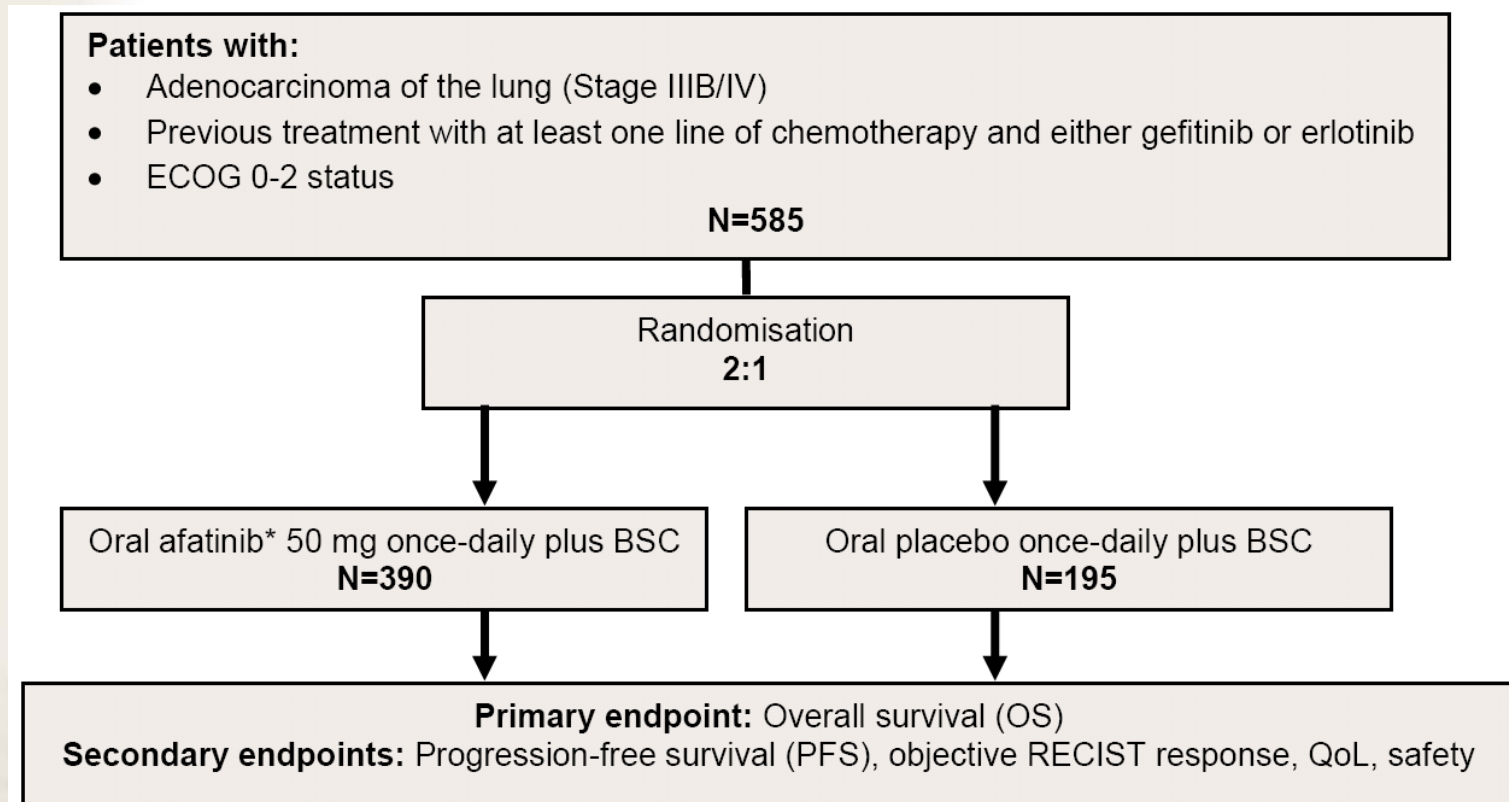


## **Hodnocení studie k článku**

***„Afatinib benefits lung cancer patients whose cancer progressed after treatment with EGFR inhibitors“***

# Cílová populace a adekvátní kontrolní skupina

- Studie LUX-Lung 1 zaměřená na pacienty s NSCLC, kteří zprogredovali na léčbě erlotinibem nebo gefitinibem.



# Primární a sekundární cíle studie

## → Primární cíl:

- Celkové přežití (overall survival, OS)

## → Sekundární cíle:

- Přežití bez známek progrese (progression-free survival, PFS)
- Odpověď na léčbu dle RECIST
- Kvalita života
- Bezpečnost

# Design studie

- **Selekce pacientů** – jasně daná předpokládaným účinkem afatinibu, je třeba posoudit zastoupení nově diagnostikovaných pacientů a pacientů s rekurentním onemocněním v obou skupinách.
- **Paralelní design** – u onkologických onemocnění není jiná volba.
- **Adekvátní kontrolní skupina** – kontrola se zdá vybraná logicky, nicméně hodnocení primárního cíle vznáší otázku:  
**reprezentativnost vs. průkaznost.**
- **Randomizace** – poměrem 2:1, specifikovány balancované proměnné: věk, ECOG PS, etnikum, historie kouření, stadium, historie léčby (souvisí ).

# Výpočet velikosti vzorku

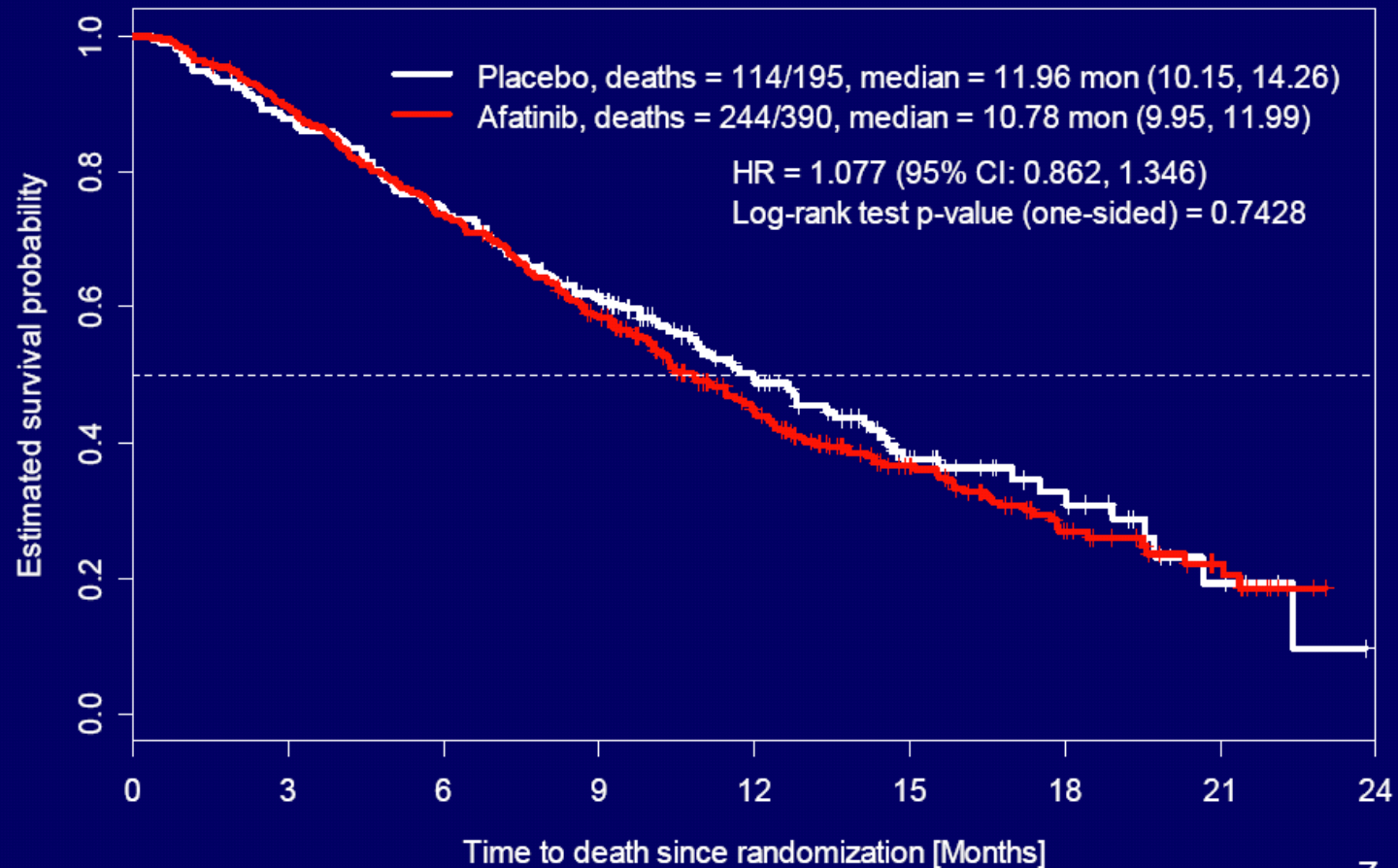
- Předpoklad mediánu celkového přežití u kontrolní skupiny 4.7 měsíců
- **Předpoklad snížení rizika úmrtí v čase o 30% (tomu odpovídá  $HR=0.7$ )**  
→ **zvýšení mediánu celkového přežití u skupiny s afatinibem na 6.7 měsíců.**
- Tomu odpovídá 359 událostí (úmrtí), což lze vzhledem k předpokládané délce studie, podílu cenzorovaných a drop-out rate přepočítat na celkový počet pacientů pro zařazení do studie: **585 randomizovaných.**

# Výsledky

- Statisticky významně více léčebných odpovědí (PR + SD)
- Statisticky významně delší přežití bez známek progresu (PFS) a to i v podskupinách
- Standardní profil nežádoucích účinků (vzhledem k anti-EGFR TKI)
- Zlepšení symptomů spojených s tumorem
- **ALE ...**

# ALE ... Celkové přežití

## Overall survival



7

# Výsledky

- Statisticky významně více léčebných odpovědí (PR + SD)
- Statisticky významně delší přežití bez známek progresu (PFS) a to i v podskupinách
- Standardní profil nežádoucích účinků (vzhledem k anti-EGFR TKI)
- Zlepšení symptomů spojených s tumorem
- **ALE nevýznamné zlepšení celkového přežití**
  - **Reprezentativnost?**
  - **Srovnatelnost?**
- **Hlavním a zásadním výsledkem pro tvorbu klinických doporučení je hodnocení primárního cíle studie!**
- **Na ten je totiž optimalizována velikost vzorku!**
- **Další výsledky mohou sloužit jako podklad (hypotézy) pro další výzkum!**



- **Je otázkou, zda se dal při přípravě studie předjímat problém s celkovým přežitím**
- **Možná zbytečné souběžné použití logrank testu a Coxova modelu (model asi neuvažoval další proměnné)**
- **Je třeba brát s rezervou všechna statistická hodnocení kromě hodnocení primárního cíle studie**
- **Zatím pouze předběžné hodnocení – je třeba doplnit kvalitní diskuzi**