

ZÁKLADY STATISTIKY V KLINICKÝCH STUDIÍCH

Adam Svobodník

Program kurzu

1. Proč bych se měl vzdělávat ve statistice?
2. Fáze klinického vývoje
3. Základní pojmy z designu klinických studií
4. Koncept P-hodnoty
5. Specifické designy klinických studií
6. Základy deskriptivní statistiky

1.

Proč bych se měl vzdělávat
ve statistice?

Proč je statistika v klinickém výzkumu „životně“ důležitá?

The image displays a Microsoft Excel spreadsheet containing a large table of clinical trial data. The spreadsheet is titled "prace upren 2011 doplneno 07092011-ed4A - Microsoft Excel". The data is organized into columns with headers such as "AU", "AV", "death(month)", "AC", "ATZ", "hypodiploemik", "antagonist", "sunkral", "kardiovaskulární sunkral", "acute IM(m)", "NAP(m)", "implantace KS", "RFA", "implantace ICD", "hosp pro", "masikulace(m)", and "Přičina smrti". The rows represent individual patients, with columns for patient ID, dates, and various clinical parameters. The data is color-coded, with green and yellow highlighting specific rows. The spreadsheet is displayed in a windowed environment, showing the Microsoft Office ribbon and taskbar.

Proč je biostatistika v klinickém výzkumu „životně“ důležitá?

The screenshot shows a web browser window displaying a New York Times article. The article title is partially visible as "How Bright Promise in Cancer-Fighting Cell Ap...". A large red semi-transparent box is overlaid on the article, containing the text "The Duke University scandal" in white. Below the article, there is a photo of two men, Keith Baggerly and Kevin Coombes, sitting in front of a presentation screen. The article text mentions "When Juliet Jacobs found out she had lung cancer, she was terrified, but realized that her hope lay in getting the best treatment medicine could offer. So she got a second opinion, then a third. In February of 2010, she ended up at Duke University, where she entered a research study whose promise seemed stunning." The article is by Gina Kolata and published on July 7, 2011. There are social media sharing options and a "Readers' Comments" section at the bottom.

The Annals of Applied Statistics
2009, Vol. 3, No. 4, 1309–1334
DOI: 10.1214/09-AOAS291
© Institute of Mathematical Statistics, 2009

DERIVING CHEMOSENSITIVITY FROM CELL LINES: FORENSIC BIOINFORMATICS AND REPRODUCIBLE RESEARCH IN HIGH-THROUGHPUT BIOLOGY

BY KEITH A. BAGGERLY¹ AND KEVIN R. COOMBES²

University of Texas

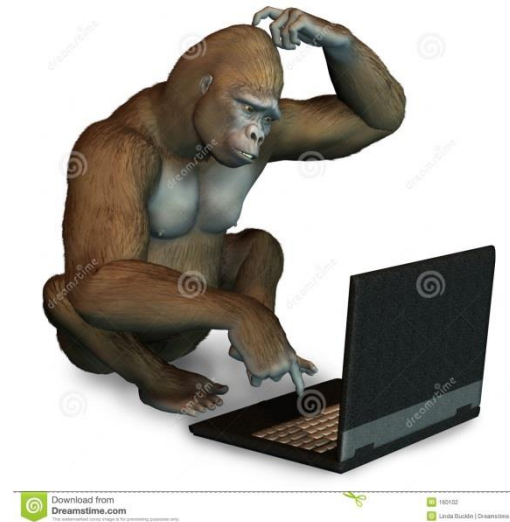
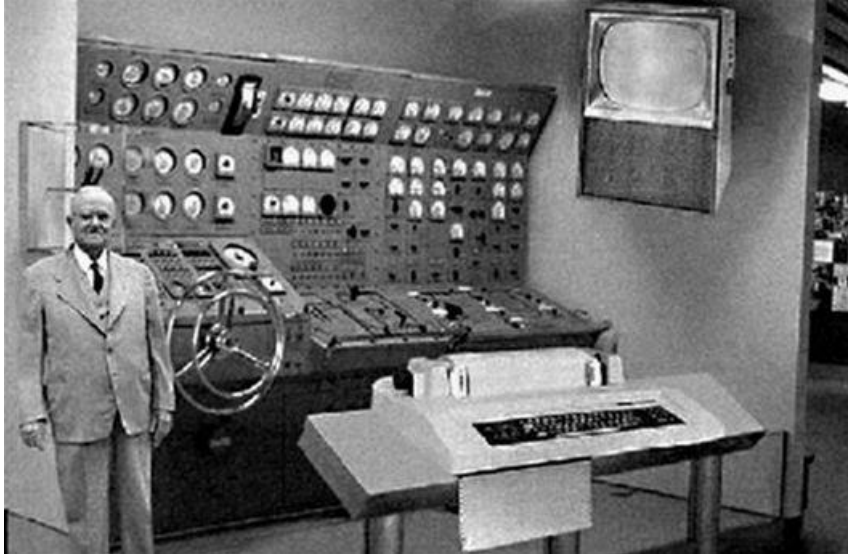
7. Discussion.

In the *Annals of Applied Statistics*, the case studies examined above, forensic reconstruction identifies errors that are hidden by poor documentation. Unfortunately, these case studies are illustrative, not exhaustive; further problems similar to the ones detailed above are described in the supplementary reports. The case studies also share other commonalities. In particular, they illustrate that *the most common problems are simple*: for example, confounding in the experimental design (all TET before all FEC), mixing up the gene labels (off-by-one errors), and mixing up the group labels (sensitive/resistant); most of these mixups involve simple switches or offsets. These mistakes are easy to make, particularly if working with Excel or if working with 0/1 labels instead of names (as with binreg). We have encountered these and like problems before. As part of the 2002 Competitive Analysis of Microarray Data (CAMDA) competition, Stivers et al. (2003) identified and corrected a mixup in annotation affecting roughly a third of the data which was driven by a simple one-cell

Paradox současnosti

Pokrok v technologiích umožňuje komukoliv provádět složité a sofistikované výpočty (např. data mining). Při absenci znalosti základních statistických technik tak roste riziko:

- chybných výsledků*
- nebo chybné interpretace správných výsledků*



Download from
Dreamstime.com

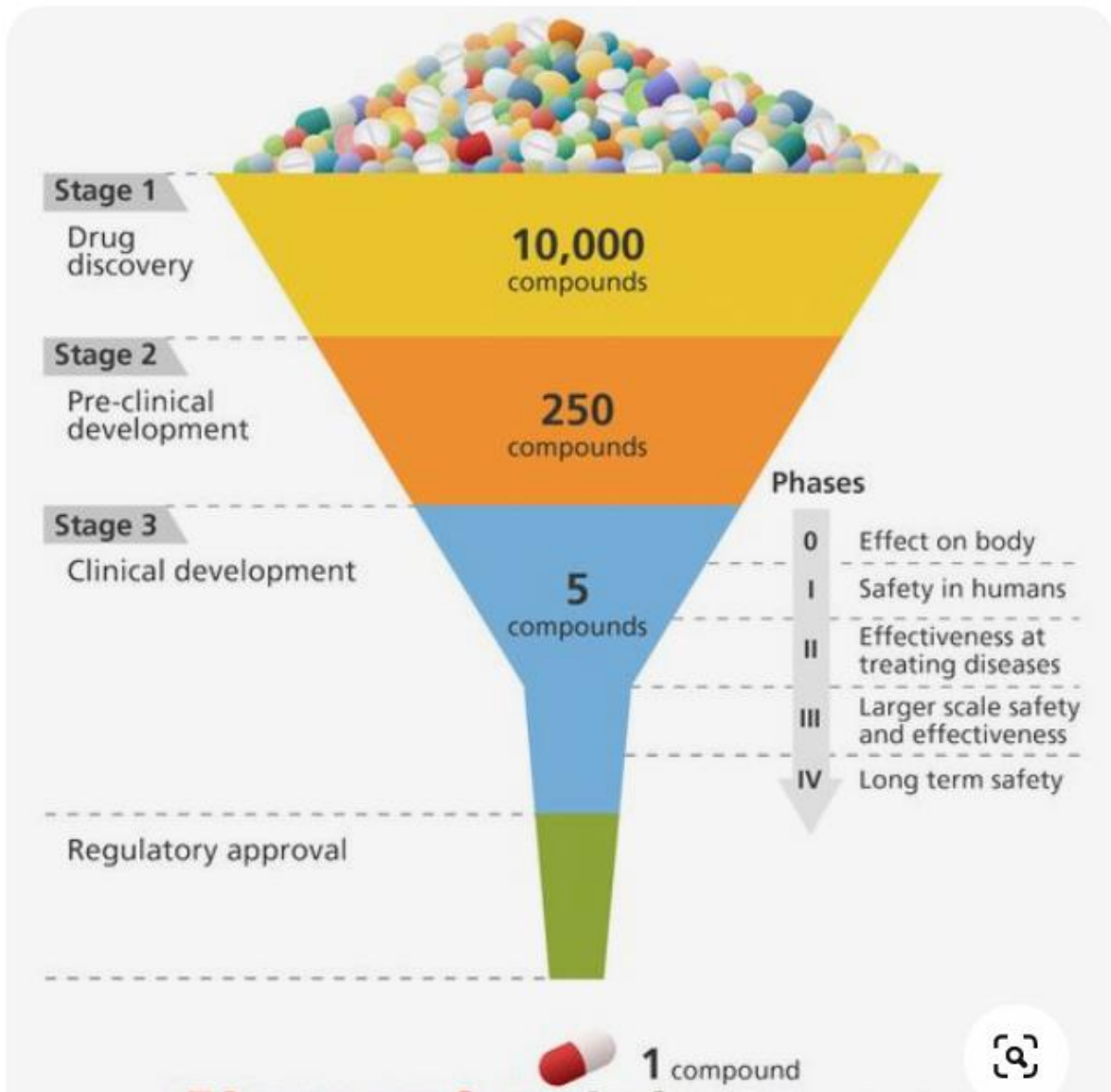
180102
Linda Stuklin | Dreamstime.com

Pro správnou interpretaci výsledků
klinických studií je potřeba „***klinické***“
i „***statistické***“ know-how ale také
nenahraditelný
zdravý „***selský rozum***“

2.

Fáze klinického vývoje

Realita vývoje nového přípravku



Úloha statistika v klinickém výzkumu



„Standardní“ proces vývoje nového přípravku



Laboratorní vývoj



Pre-klinický vývoj



**Klinické studie: Fáze I
Fáze II
Fáze III**



Registrace a zavedení do praxe

Post-marketingové studie





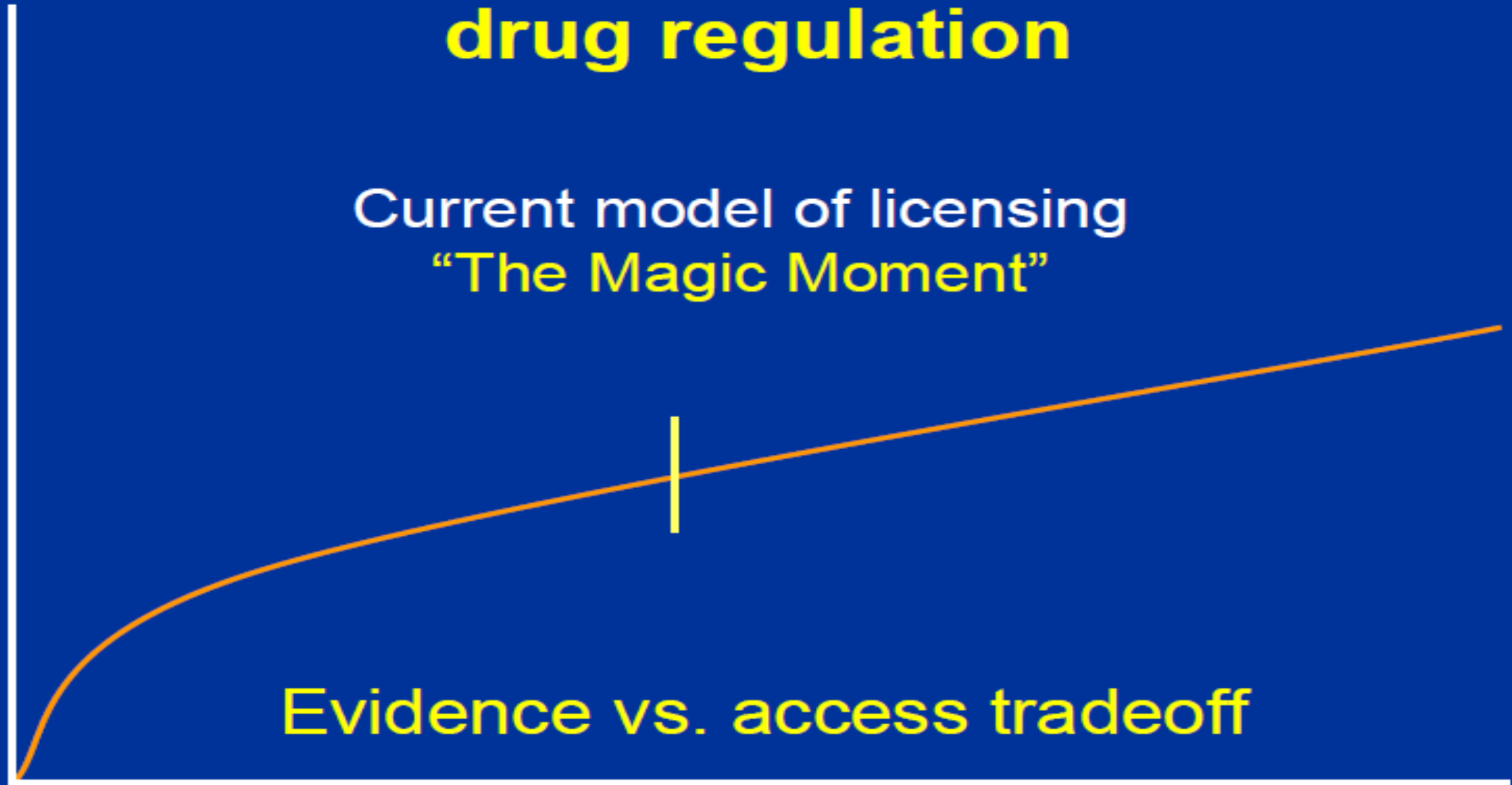
The binary nature of drug regulation

Knowledge, investment

Current model of licensing
“The Magic Moment”

Evidence vs. access tradeoff

Time (years)





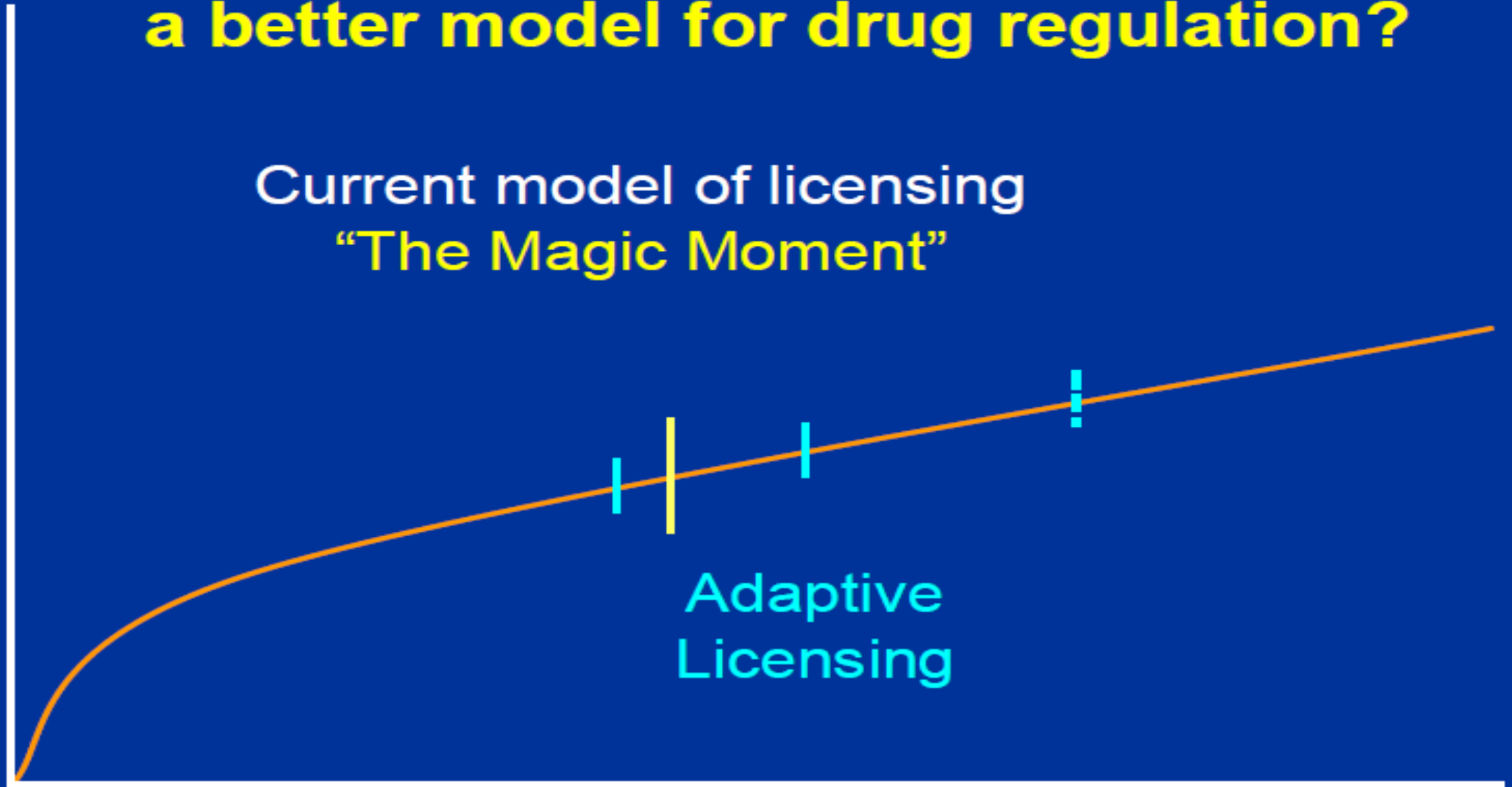
Going beyond 'binary'; a better model for drug regulation?

Knowledge, investment

Current model of licensing
"The Magic Moment"

Adaptive
Licensing

Time (years)



Fáze klinických studií z hlediska analýzy dat

Klinické hodnocení „FÁZE I“

- **Cíl:**

- Stanovení základních humánních farmakokinetických parametrů přípravku
- Stanovení maximální tolerovatelné dávky MTD (cytostatika apod.)
- Sledování nežádoucích účinků (bezpečnost)
- „Dose finding“ studie

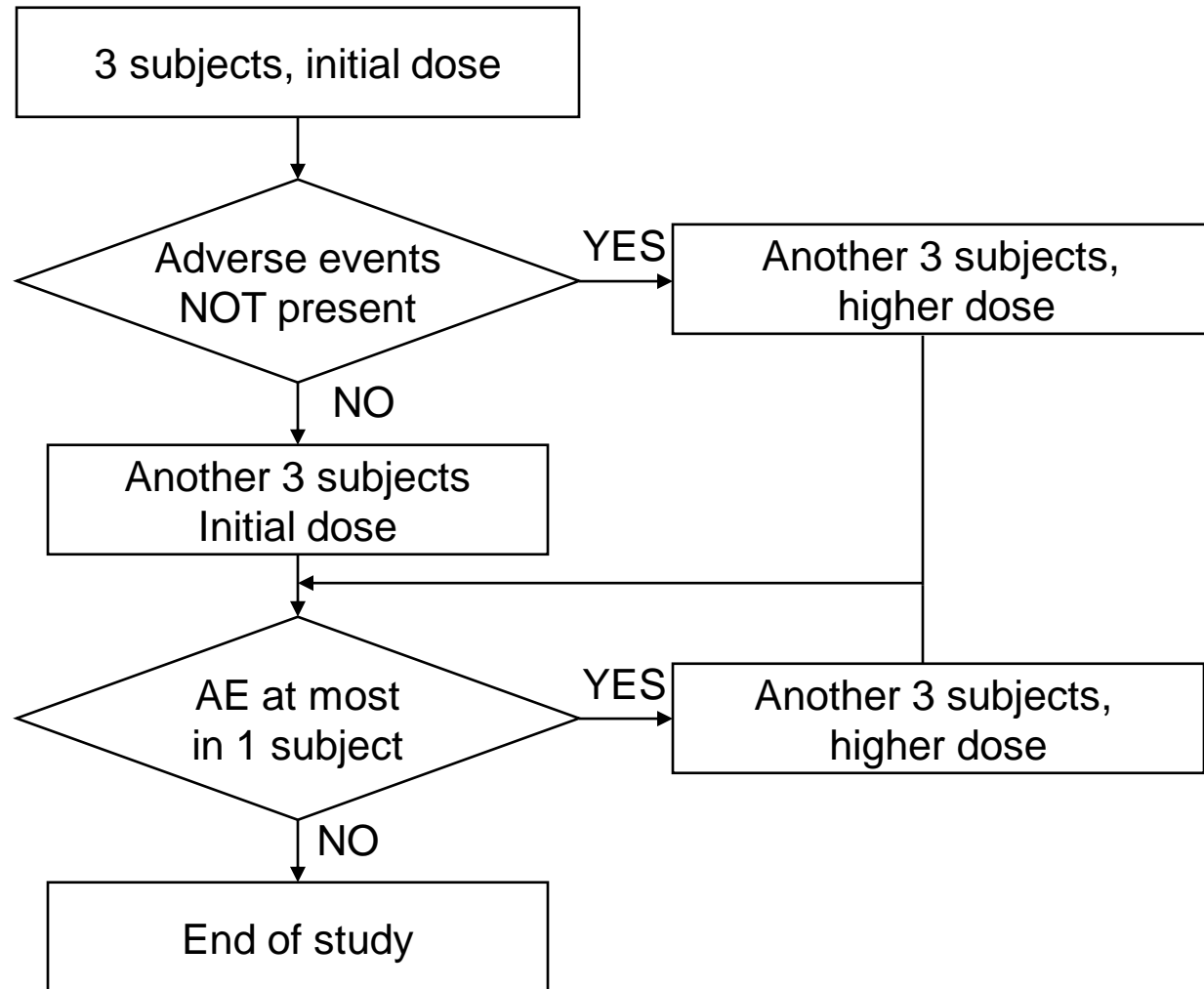
- **Subjekty hodnocení**

- 12-20
- Většinou zdraví dobrovolníci
- Nikdy ne *vulnerabilní* osoby

- **Design**

- Ideální uspořádání experimentu vede k odhadu základních parametrů křivky “dose – response”
 - Z etických důvodů využívány adaptivní designy:
následující dávka je stanovena na základě odpovědi subjektu hodnocení na dávku předchozí
 - První dávka je stanovena na základě preklinického hodnocení přípravku (animální testy)
-

„Dose-finding“ studie



Klinické hodnocení „FÁZE II“

- **Cíl:**

- Ověření účinnosti testovaného přípravku
- Hodnocení tolerance nebo bezpečnosti
- Opodstatnění testování přípravku v rozsáhlejších KH

- **Subjekty hodnocení**

- 20 – 200
- počet hodnocených subjektů
 - fixní
 - nábor po skupinách
 - sekvenční design (průběžné hodnocení odpovědi každého SH)

- **Design**

- Ojediněle randomizace
 - Jednoramenná KH
 - Účinek a toxicita srovnávány se standardně používanou terapií nebo placebem
 - „Pilot trial“ (IIA)
 - „Pivotal trial“ (IIB)
-

Klinické hodnocení „FÁZE III“

- **Cíl:**

- Srovnání účinnosti a bezpečnosti testovaného přípravku se standardem (aktivní kontrolou) nebo placebem
- Vytvoření podkladů pro povolení k užívání
- “Cost – effectiveness” analýzy

- **Subjekty hodnocení**

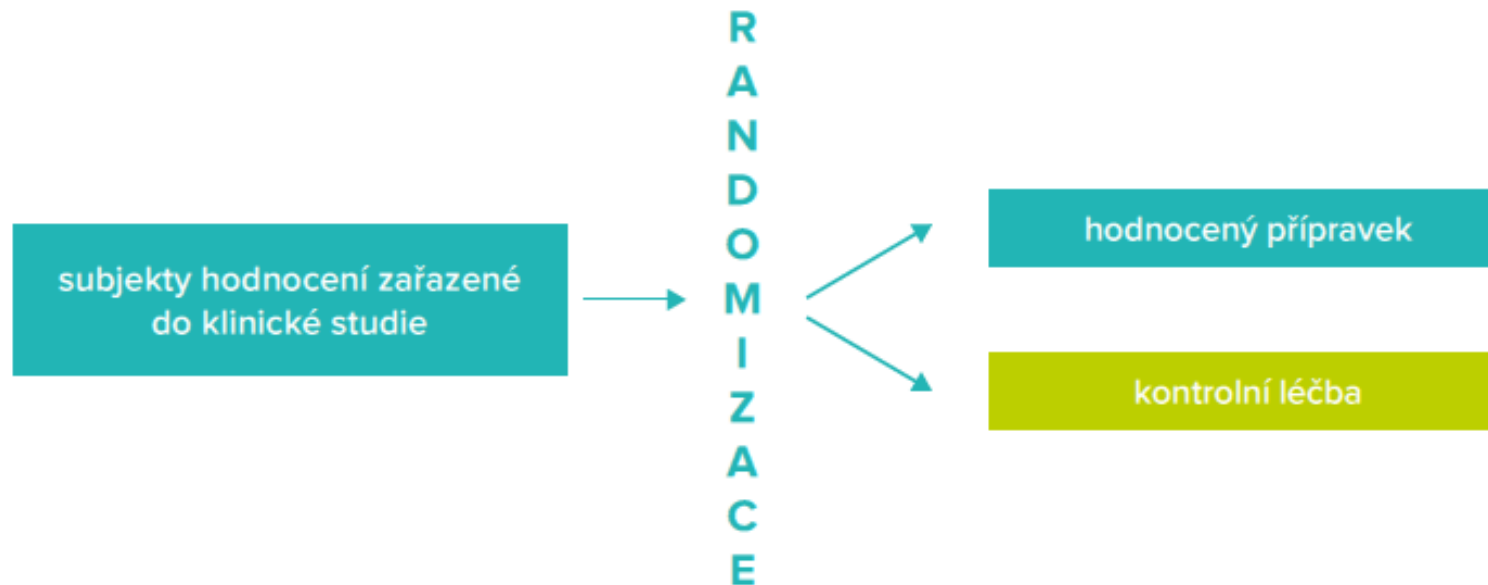
- 100 – 1 000
- počet hodnocených subjektů
 - fixní
 - nábor po skupinách
 - sekvenční design (průběžné hodnocení odpovědi každého SH)

- **Design**

- Paralelní
 - „Cross – over“
 - Faktoriální
 - Randomizace
-

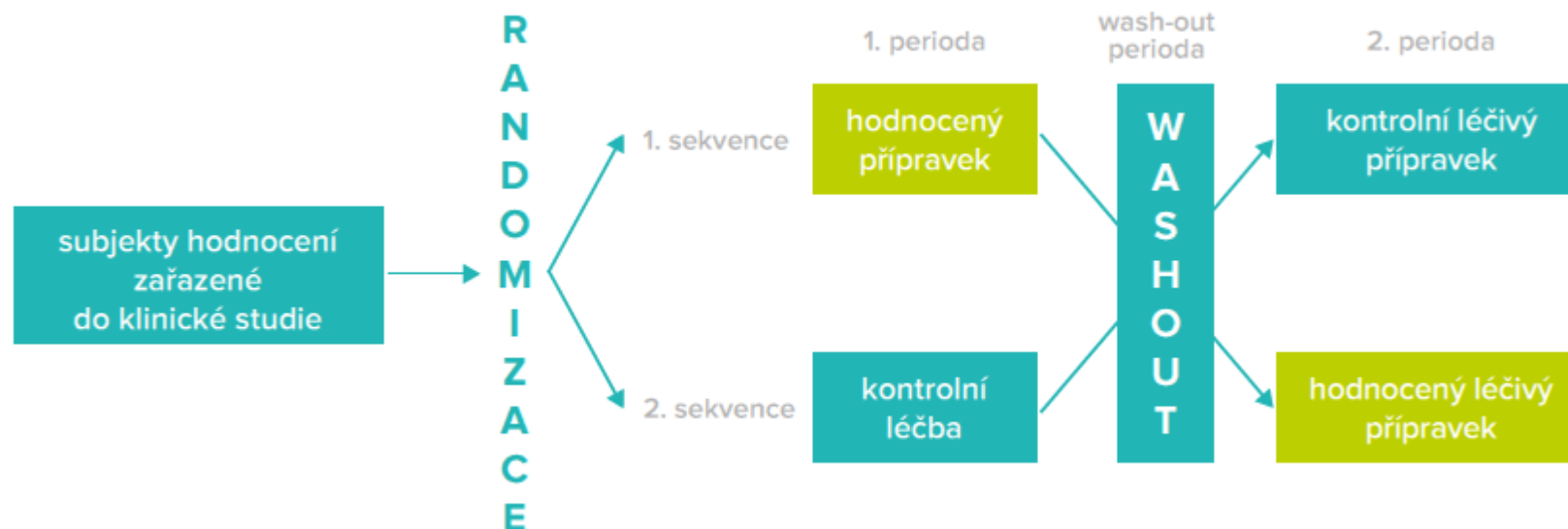
Klinické hodnocení „FÁZE III“: Paralelní design

Tento typ uspořádání je charakteristický pro KH fáze III kde je srovnávána účinnost a bezpečnost dvou nebo více terapeutických postupů. Základním prvkem designu je randomizace SH do jednotlivých léčebných skupin, v průběhu celého KH je zachováno jedinečné přiřazení SH do léčebných skupin.



Klinické hodnocení „FÁZE III“: Crossover uspořádání

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



Klinické hodnocení „FÁZE IV“

- **Cíl:**

- Potvrzení vlastností hodnoceného přípravku za „reálných“ podmínek
- Detailní analýza nežádoucích účinků
- Hodnocení kvality života
- Úpravy dávkovacího režimu přípravku
- „Cost-effectiveness“ studie

- **Design**

- **Deskriptivní studie** (analýza běžných informačních databází o pacientech)
 - „**Cross - sectional**“ studie (analýza voleného strukturovaného vzorku pacientů)
 - „**Case - control**“ studie (retrospektivní studie založené na párové volbě kontrolní skupiny)
 - **Kohortní studie** (retrospektivní nebo prospektivní srovnání kohorty s kontrolní skupinou)
-

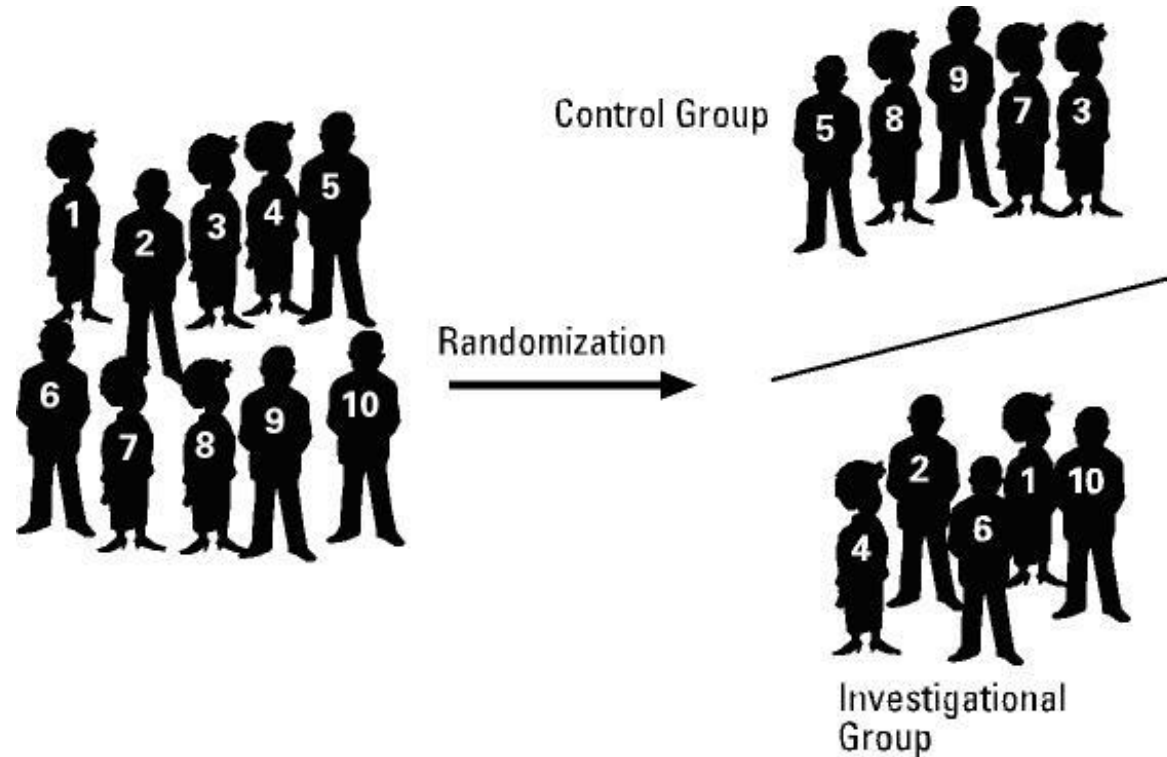
3.

Základní pojmy z designu klinických studií

Randomizace

Randomizace

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



Základní klasifikace randomizačních technik

- **Nepřípustné:**

Randomizace SH na základě:

- Pořadového čísla vstupu do KH
- Iniciál SH
- Data narození
- Data vstupu do KH

- **Méně vhodné:**

- Komplettní randomizace

- **Doporučené:**

- Permutační bloková randomizace
 - Stratifikovaná permutační bloková randomizace
 - Adaptivní randomizace
 - Minimalizace
 - Metoda “Play the winner”
 - Metoda “falešné mince” nebo “osudí”
-

Kompletní randomizace

– riziko nevyváženého počtu SH v léčebných skupinách

$N = 100$ SH

$P_{(A)} = 0,5$... pravděpodobnost randomizace SH do ramene A

$P_{(B)} = 0,5$.. pravděpodobnost randomizace SH do ramene B

$P_{(r)}$ pravděpodobnost ekvivalentního počtu SH v obou ramenech studie

$$P_{(r)[x=n/2]} = \binom{n}{\frac{n}{2}} \cdot \left(\frac{1}{2}\right)^n = 8\%$$

Závěr:

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

Kompletní randomizace

– riziko nevyváženého počtu SH v léčebných skupinách

Příklad nevyváženého počtu SH v multicentrickém KH
(Chaow and Liu 1998)

Př.: Multicentrické KH, 4 centra (1-4), 3 ramena (A,B,C), N = 96 SH
Randomizace provedená separátně pro každé centrum generátorem náhodných čísel

Centrum	Skupina			Σ
	A	B	C	
1	7	10	7	24
2	9	6	9	24
3	9	6	9	24
4	6	11	7	24
Σ	31	33	31	96

Závěr:

V multicentrických KH s limitovaným počtem SH roste riziko nerovnoměrné distribuce počtu SH v léčebných skupinách v rámci jednotlivých center.

Stratifikovaná permutační bloková randomizace

Princip:

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

1) Volba stratifikačních kriterií:

- pohlaví: muž x žena
- věk: ≤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	
	≤50	>50	≤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

Adaptivní randomizace: Minimalizace

- **Princip:** Každý SH je randomizován do léčebné skupiny na základě rozložení prognostických faktorů z dosavadního průběhu KH
- **Př.:** KH se sledovanými třemi prognostickými faktory. Dosavadní průběh KH:

		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

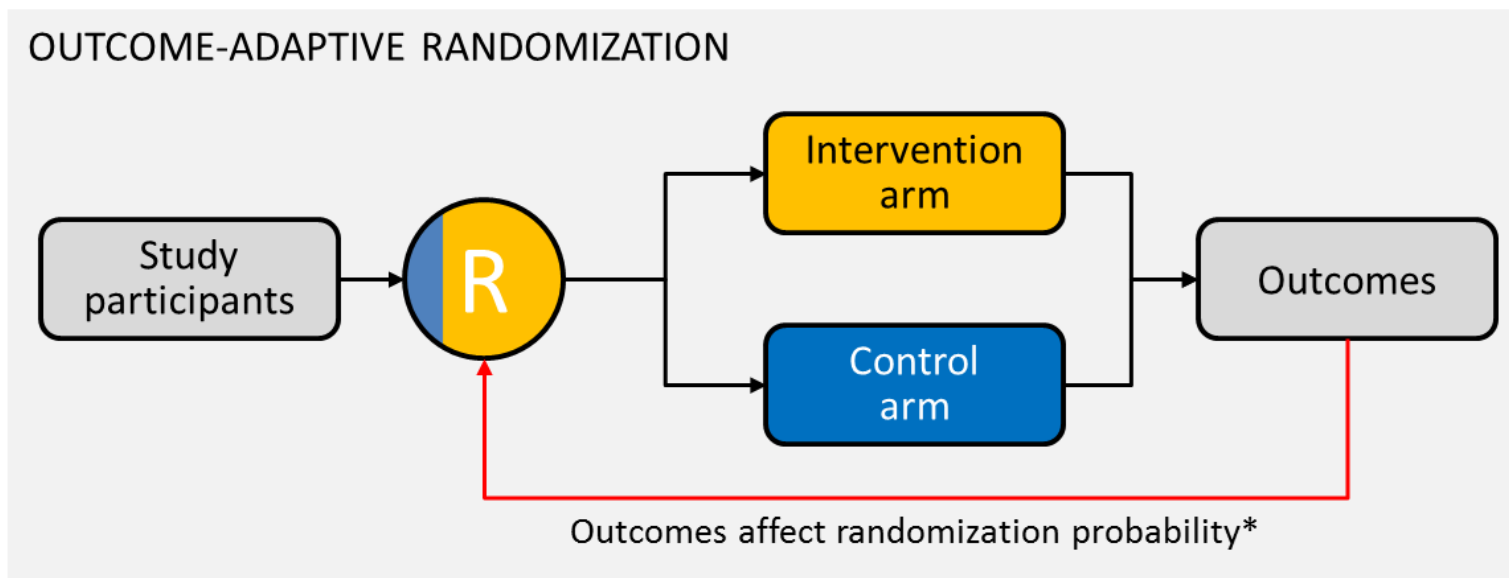
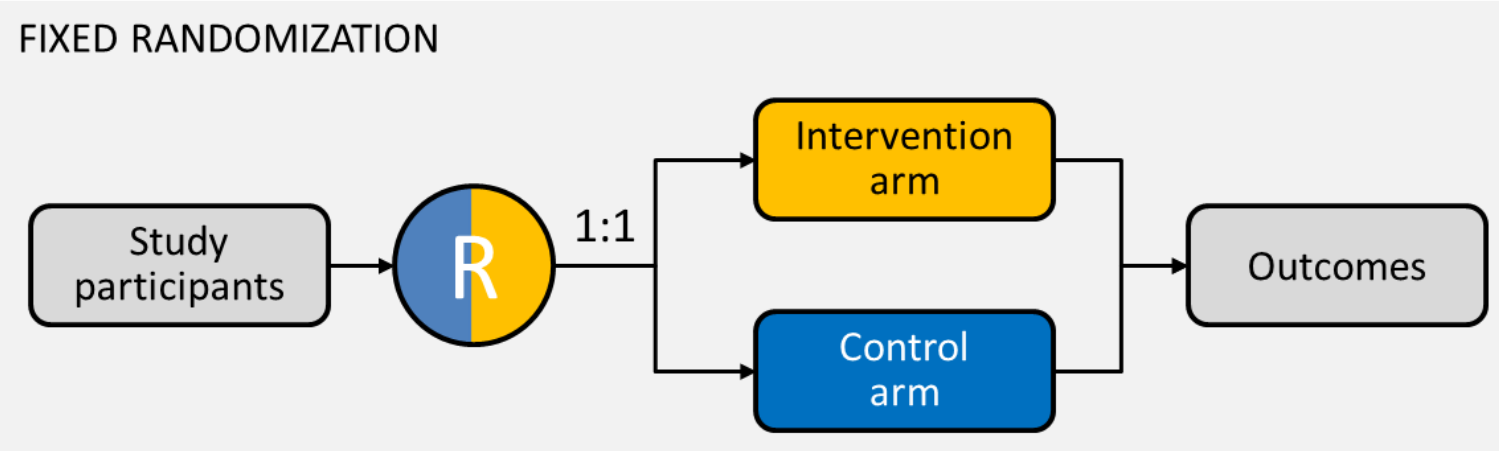
Počet SH se stejnou kombinací prognostických faktorů:

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.

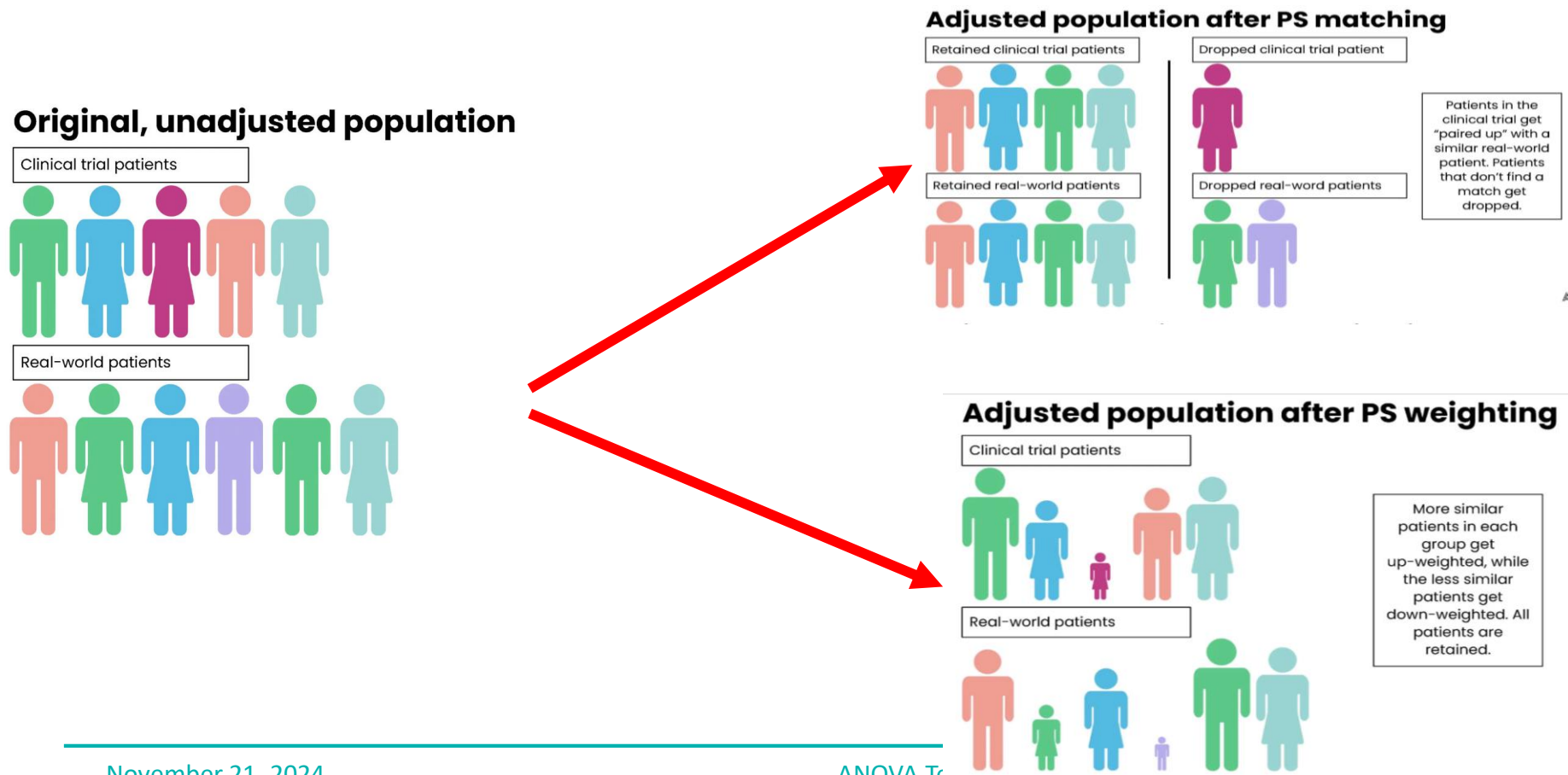
Adaptivní randomizace



Propensity score matching

Metoda IPD MAIC částečně nahrazuje randomizaci v metaanalýzách:

- disbalanci ve vybraných známých baseline charakteristikách pacientů srovnávaných skupin



ITT/PP populace

ITT a PP analýza/populace

Intention-to-treat (ITT) analytický přístup vychází z hodnocení dat všech SH randomizovaných do léčebných skupin bez ohledu na to zda

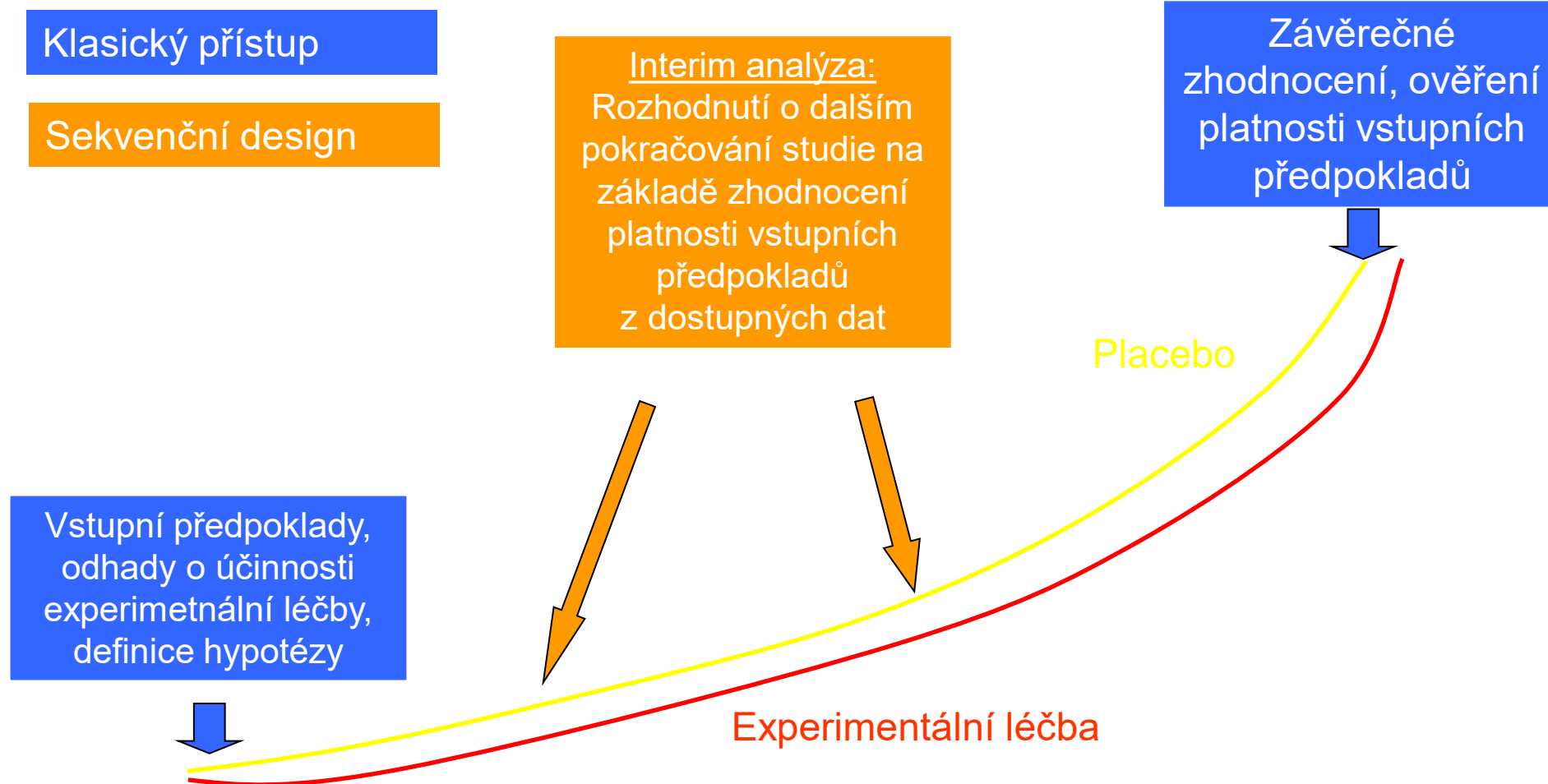
- splnily vstupní kritéria KH
- užívaly přípravek přidělený randomizačním kódem
- dodržely protokol KH
- předčasně ukončily účast v KH

Per-protocol (PP) analýza vychází pouze z dat SH u kterých byl v průběhu celého KH dodržen protokol.

Plán statistických analýz, blind review meeting

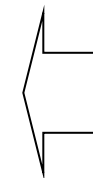
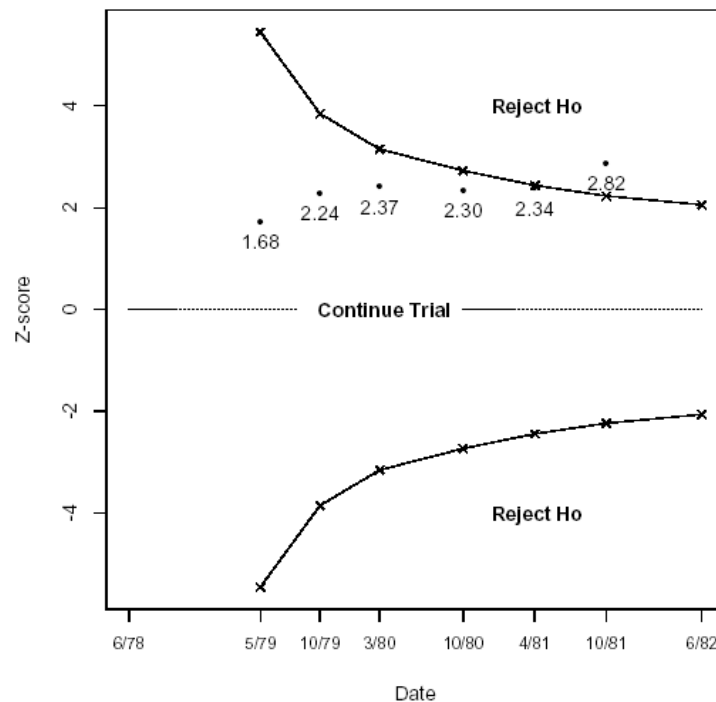
Interim analýza

Interim analýza



Interim analýza

Interim analýza (IA) je jednorázové nebo opakované statistické testování primární hypotézy před protokolárně plánovaným ukončením KH. Cílem IA je zvážit možnost ukončení KH v případě že další pokračování již nepřinese novou informaci.



O'Brien-Flemming interim monitoring boundaries for the primary endpoint are based on predetermined number of planned interim analysis with overall type error of $\alpha=0.05$.

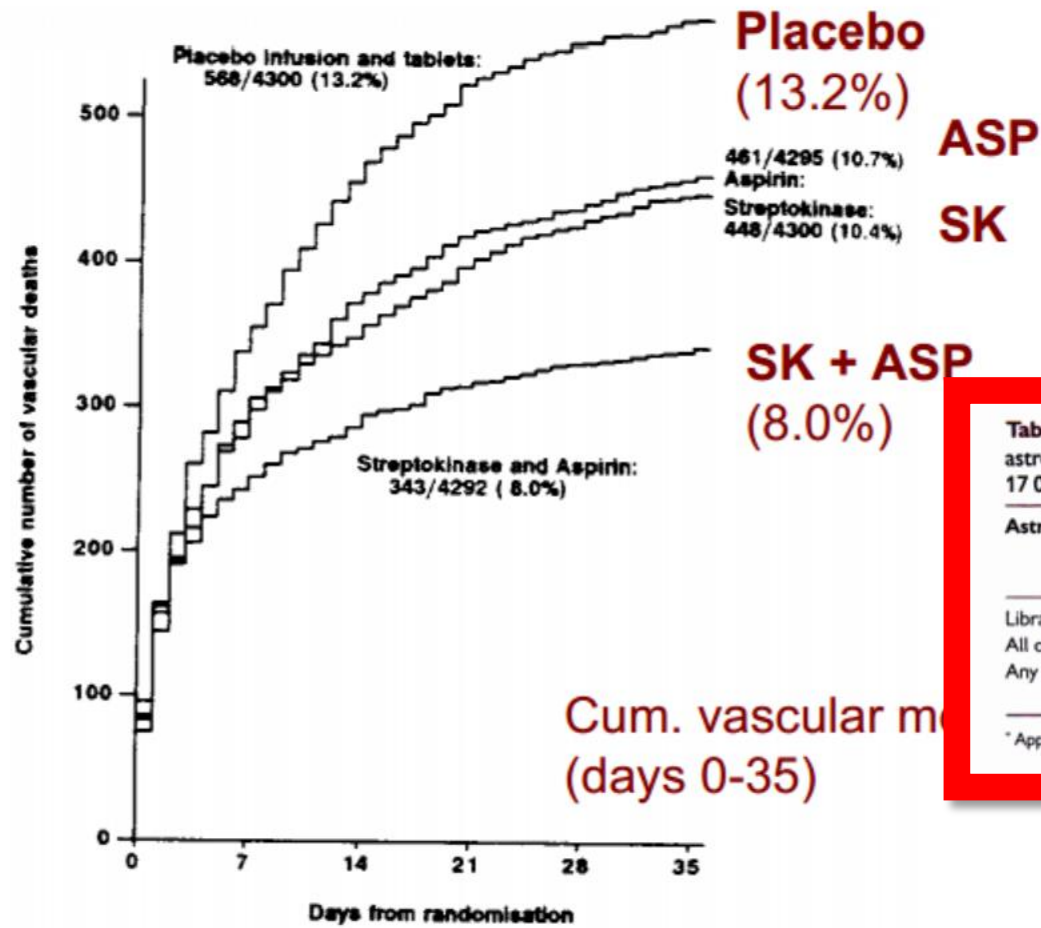
**Chybné provedení interim analýzy zvyšuje
pravděpodobnost falešně pozitivního výsledku
studie, proto se její provedení musí řídit
následujícími pravidly:**

- Počet a časování plánovaných interim analýz musí být uveden v protokolu studie (tedy před jejím zahájením)
 - Pravidla pro průkaz superiority experimentální léčby jsou přísnější při interim analýze než při závěrečném zhodnocení studie
-

Subgroup analýzy

Příkladová studie: ISIS-2

- 17 187 pacientů
- Význam aspirinu (ASP) a streptokinázy (SK) pro dlouhodobé přežití po podezření na IM
- Studie prokázala (na celém souboru) výrazné snížení mortality při užívání ASP, SK a jejich kombinace



Analýza podskupin podle dne a měsíce narození pacientů ukázala, že ASP a SK nemají vliv na přežití pacientů narozených ve znamení Váhy a Blíženci

Table 2 False-negative mortality effect in a subgroup defined only by the astrological birth sign: the ISIS-2 (1988) trial of aspirin among over 17 000 patients with acute myocardial infarction

Astrological birth sign	No. of 1-month deaths (aspirin versus placebo)	Statistical significance
Libra or Gemini	150 vs. 147	NS
All other signs	654 vs. 869	2p < 0.000001
Any birth sign*	804 vs. 1016 (9.4%) (11.8%)	2p < 0.000001

* Appropriate overall analysis for assessing the true effect in all subgroups.

Subgroup analysis

Subgroup		Median OS (95% CI)			Missing survival data (%)	
		PAL+LET	PBO+LET		PAL+LET	PBO+LET
All randomized patients		53.9 (49.8, 60.8)	51.2 (43.7, 58.9)		13	21
Age	<65 y	53.3 (47.0, 60.8)	54.4 (44.8, 60.2)		11	21
	≥65 y	58.6 (49.8, 66.7)	47.4 (36.2, 60.4)		17	21
Region	North America	53.8 (47.3, 61.3)	49.4 (37.0, 57.0)		16	23
	Europe	52.3 (46.0, 63.8)	53.8 (42.3, 78.7)		11	17
	Asia/Pacific	73.4 (47.3, NE)	55.1 (32.2, NE)		14	29
ECOG performance status	0	58.2 (52.1, 66.0)	85.9 (53.8, NE)		11	28
	1/2	47.3 (41.3, 60.8)	38.8 (32.2, 49.8)		16	16
Disease site	Visceral	48.1 (42.3, 53.8)	44.8 (32.2, 53.8)		13	23
	Non-visceral	60.8 (53.8, 72.3)	59.7 (47.4, 85.3)		14	20
Disease-free interval	De novo metastatic	54.6 (47.0, 69.1)	60.4 (49.8, 93.8)		10	24
	≤12 months	45.7 (36.1, 51.1)	37.7 (27.1, 56.4)		12	29
	>12 months	66.3 (52.1, 79.7)	47.4 (37.7, 57.0)		17	15
Prior endocrine therapy	Yes	53.3 (48.0, 62.9)	44.6 (34.3, 52.8)		16	18
	No	55.1 (47.3, 71.4)	60.4 (49.8, 93.8)		10	25
Prior chemotherapy	Yes	51.6 (45.6, 58.6)	44.6 (36.2, 54.4)		13	18
	No	58.4 (50.5, 71.7)	58.9 (47.7, 81.0)		14	24
Bone-only disease	Yes	63.5 (53.9, 79.7)	52.3 (42.3, 59.7)		16	15
	No	51.6 (46.9, 57.6)	49.8 (38.8, 60.4)		13	23
Number of disease sites	1	59.1 (53.8, 73.9)	54.4 (45.4, 70.3)		15	18
	2	60.8 (47.9, 87.2)	54.5 (33.5, NE)		15	37
	≥3	48.1 (41.3, 53.8)	45.8 (32.4, 57.0)		12	15



Příklady chybných závěrů subgroup analýz, které byly později vyvráceny

Observation	Refutation
Aspirin is ineffective in secondary prevention of stroke in women ^{29,30}	31
Antihypertensive treatment for primary prevention is ineffective in women ^{32,33}	34
Antihypertensive treatment is ineffective or harmful in elderly people ³⁵	36
Angiotensin-converting enzyme inhibitors do not reduce mortality and hospital admission in patients with heart failure who are also taking aspirin ³⁷	38
β blockers are ineffective after acute myocardial infarction in elderly people, ³⁹ and in patients with inferior myocardial infarction ⁴¹	40
Thrombolysis is ineffective >6 hours after acute myocardial infarction ⁴²	43
Thrombolysis for acute myocardial infarction is ineffective or harmful in patients with a previous myocardial infarction ⁴²	44
Tamoxifen citrate is ineffective in women with breast cancer aged <50 years ⁴⁵	46
Benefit from carotid endarterectomy for symptomatic stenosis is reduced in patients taking only low-dose aspirin due to an increased operative risk ⁴⁷	48
Amlodipine reduces mortality in patients with chronic heart failure due to non-ischaemic cardiomyopathy but not in patients with ischaemic cardiomyopathy ⁴⁹	50

Table 1: Examples of subgroup analyses that have shown apparently clinically important heterogeneity of treatment effect which has subsequently been shown to be false

Doporučení pro použití subgroup analýz

Panel 1: Rules of subgroup analysis: a proposed guideline for design, analysis, interpretation, and reporting

Trial design

- Subgroups analyses should be defined before starting the trial and should be limited to a small number of clinically important questions.
- Expert clinical input into the design of subgroup analyses is needed to ensure that all relevant baseline clinical and other data are recorded.
- The direction and magnitude of anticipated subgroup effects should be stated at the outset.
- The exact definitions and categories of the subgroup variables should be defined explicitly at the outset in order to avoid post hoc data-dependent variable or category definitions. For continuous or hierarchical variables the cut-off points for analysis should be predefined.
- Stratification of randomisation by important subgroup variables should be considered.
- If important subgroup-treatment effect interactions are anticipated, trials should ideally be powered to detect them reliably.
- Trial stopping rules should take into account anticipated subgroup-treatment effect interactions and not simply the overall effect of treatment.
- If relative treatment effect is likely to be related to baseline risk, the analysis plan should include a stratification of the results by predicted risk. The risk score or model should be selected in advance so that the relevant baseline data can be recorded.

Analysis and reporting

- The above design issues should be reported in the methods section along with details of how and why subgroups were selected.
- Significance of the effect of treatment in individual subgroups should not be reported; rates of false negative and false positive results are extremely high. The only reliable statistical approach is to test for a subgroup-treatment effect interaction.
- All subgroup analyses that were done should be reported—ie, not only the number of subgroup variables but also the number of different outcomes analysed by subgroup, different lengths of follow-up etc.

- Significance of pre hoc subgroup-treatment effect interactions should be adjusted when multiple subgroup analyses are done.
- Subgroup analyses should be reported as absolute risk reductions and relative risk reductions. Where relevant the statistical significance of differences in absolute risk reductions should be tested.
- Ideally, only one outcome should be studied and this should usually be the primary trial outcome, irrespective of whether this is one outcome or a clinically important composite outcome.
- Comparability of treatment groups for prognostic factors should be checked within subgroups.
- If multiple subgroup-treatment effect interactions are identified, further analysis is needed to check whether their effects are independent.

Interpretation

- Reports of the significance of the effect of treatment in individual subgroups should be ignored, especially reports of lack of benefit in a particular subgroup in a trial in which there is overall benefit, unless there is a significant subgroup treatment effect interaction
- Genuine unanticipated subgroup-treatment effect interactions are rare (assuming that expert clinical opinion was sought in order to pre-define potentially important subgroups) and so apparent interactions that are discovered post hoc should be interpreted with caution.
- No test of significance is reliable in this situation.
- Pre hoc subgroup analyses are not intrinsically valid and should still be interpreted with caution. The false positive rate for tests of subgroup-treatment effect interaction when no true interaction exists is 5% per subgroup.
- The best test of validity of subgroup-treatment effect interactions is their reproducibility in other trials.
- Few trials are powered to detect subgroup effects and so the false negative rate for tests of subgroup-treatment effect interaction when a true interaction exists will usually be high.

Panel 2: The four main clinical indications for subgroup analysis

Potential heterogeneity of treatment effect related to risk

- Differences in risks of treatment
- Differences in risk without treatment

Potential heterogeneity of treatment effect related to pathophysiology

- Multiple pathologies underlying a clinical syndrome
- Differences in the biological response to a single pathology
- Genetic variation

Clinically important questions related to the practical application of treatment

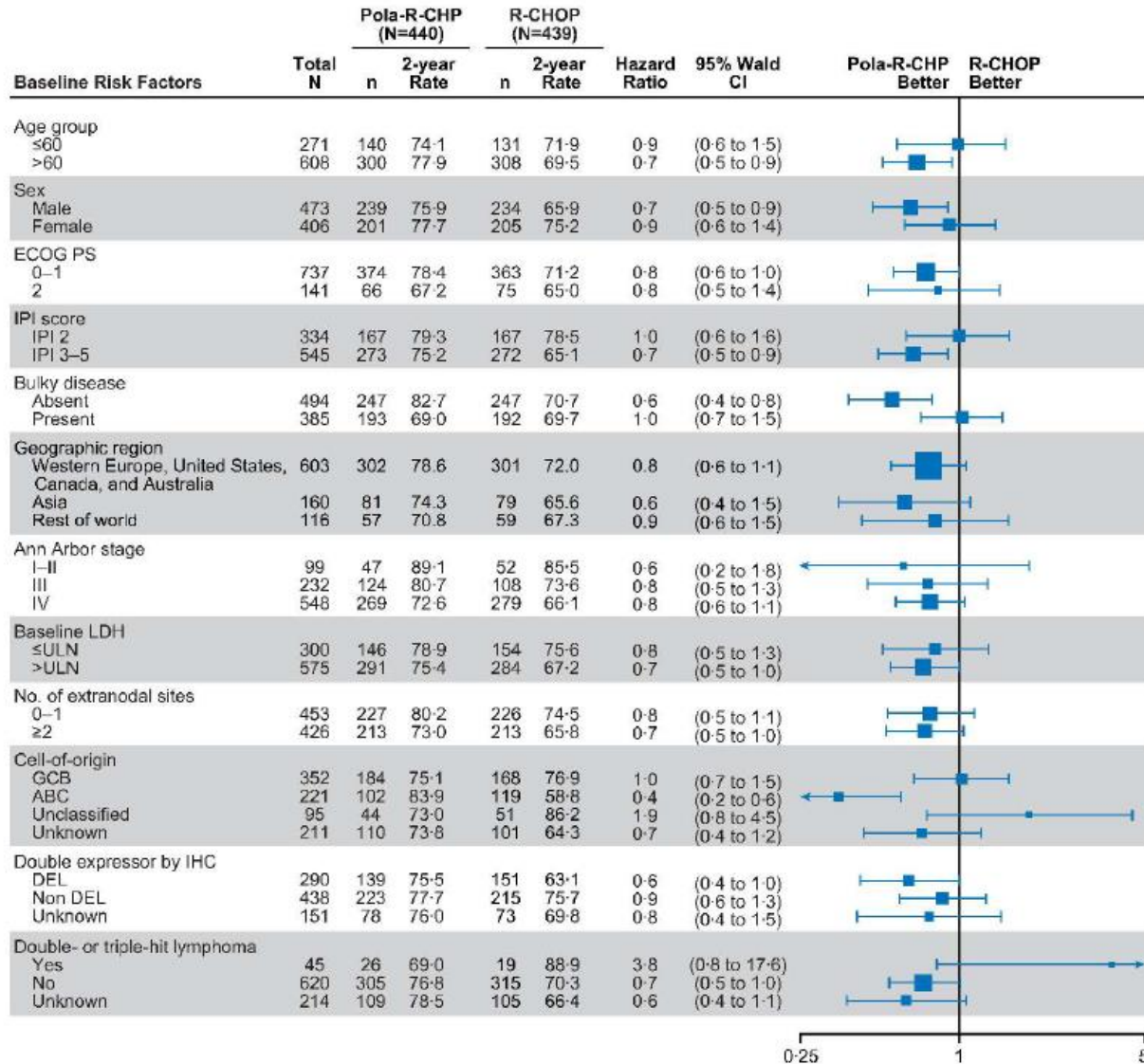
- Does benefit differ with severity of disease?
- Does benefit differ with stage in the natural history of disease?
- Is benefit related to the timing of treatment after a clinical event?
- Is benefit dependent on comorbidity?

Underuse of treatment in routine clinical practice due to uncertainty about benefit

- Underuse of treatment in specific groups of patients eg, elderly people
- Confinement of treatment according a narrow range of values of a relevant physiological variable—eg, treatment thresholds for cholesterol level or blood pressure

Subgroup analysis: Fundamental principles

Subgroup analysis in POLARIX study (forest plot)

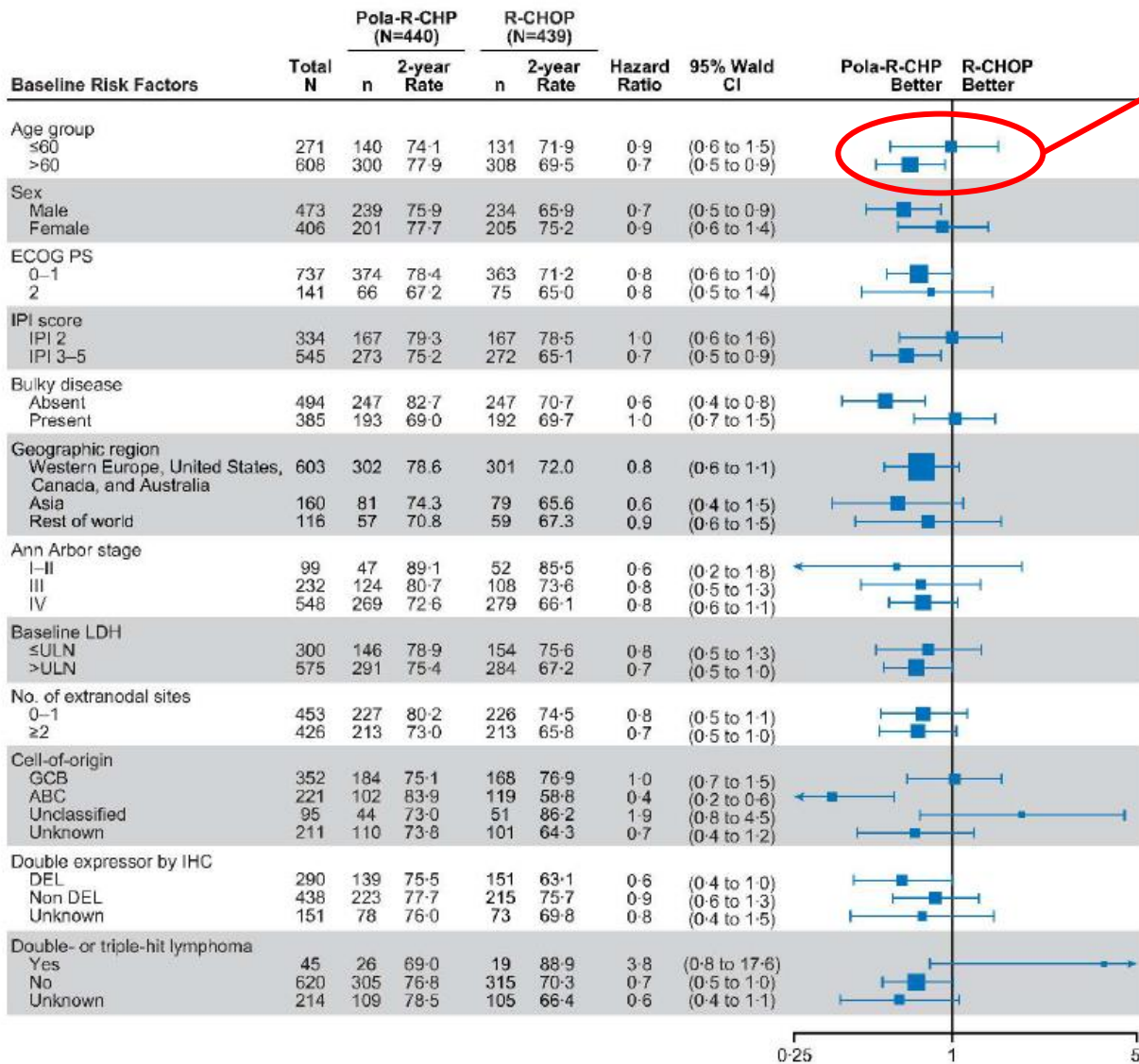


Subgroup analyses need to be interpreted carefully to avoid misinterpretation

The main mistake in interpretation of subgroup analyses:

- Restriction of the therapy use only to selected subgroups despite the positive effect in the total population

How to interpret correctly subgroup analyses

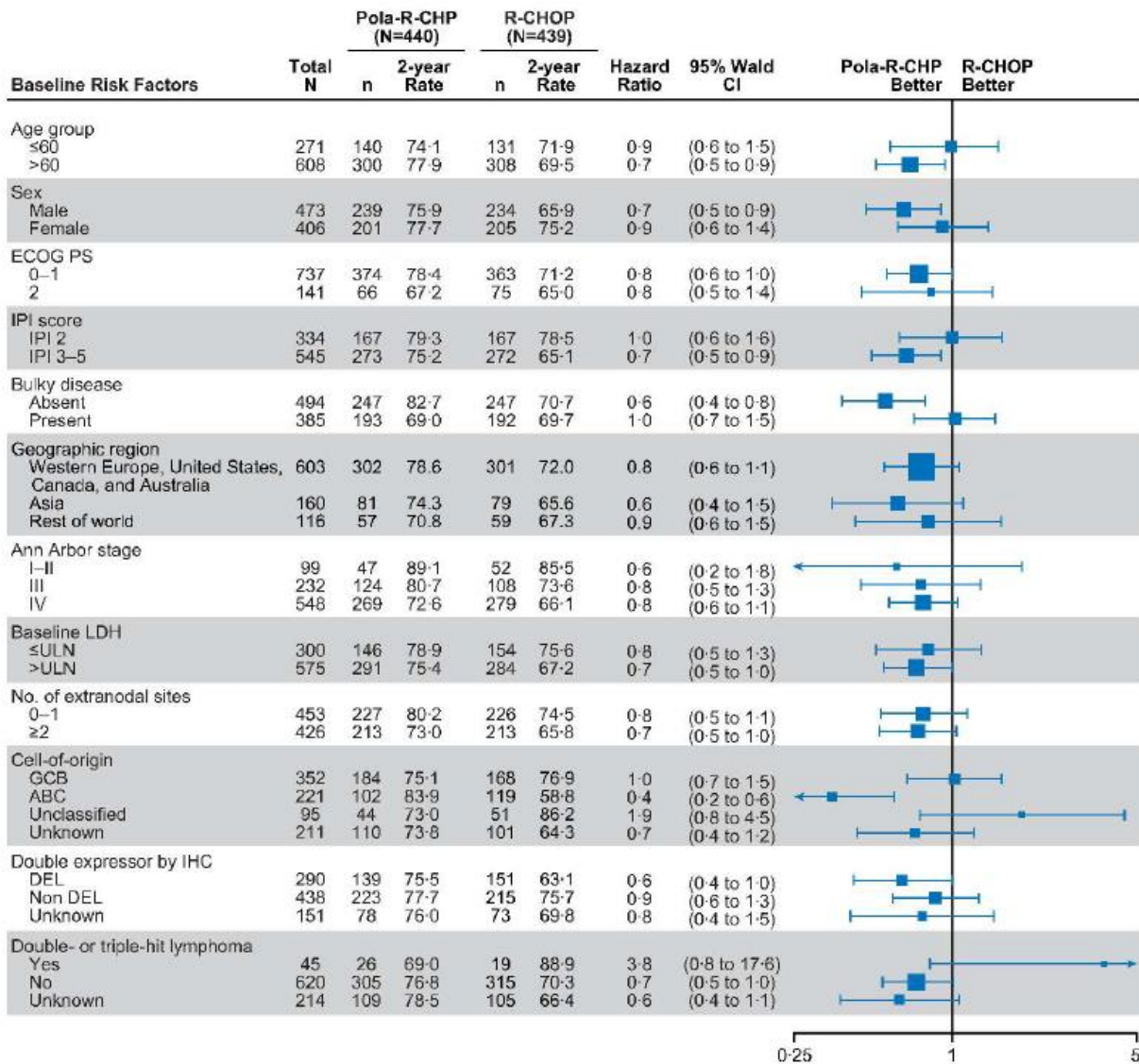


Outcome of one subgroup analysis itself is not sufficient to conclude that the treatment works better/worse in some subgroups.

To conclude, that the treatment works better/worse in particular subgroup we need to have (at least):

- Similar and consistent findings from several independent analyses
- Biological plausibility – justification based on disease biology, pathology etc.
- Statistical considerations (sample size, power and type I error control)

Interpretation of subgroup analyses results



The study was not designed and powered to enable conclusions on treatment effect in individual subgroups

Study results need to be interpreted only for the whole ITT population because we do not know the impact of other confounding factors (imbalance of prognostic factors) on subgroup analysis

No valid conclusions about significantly better or worse results for any of the investigated subgroups could be made

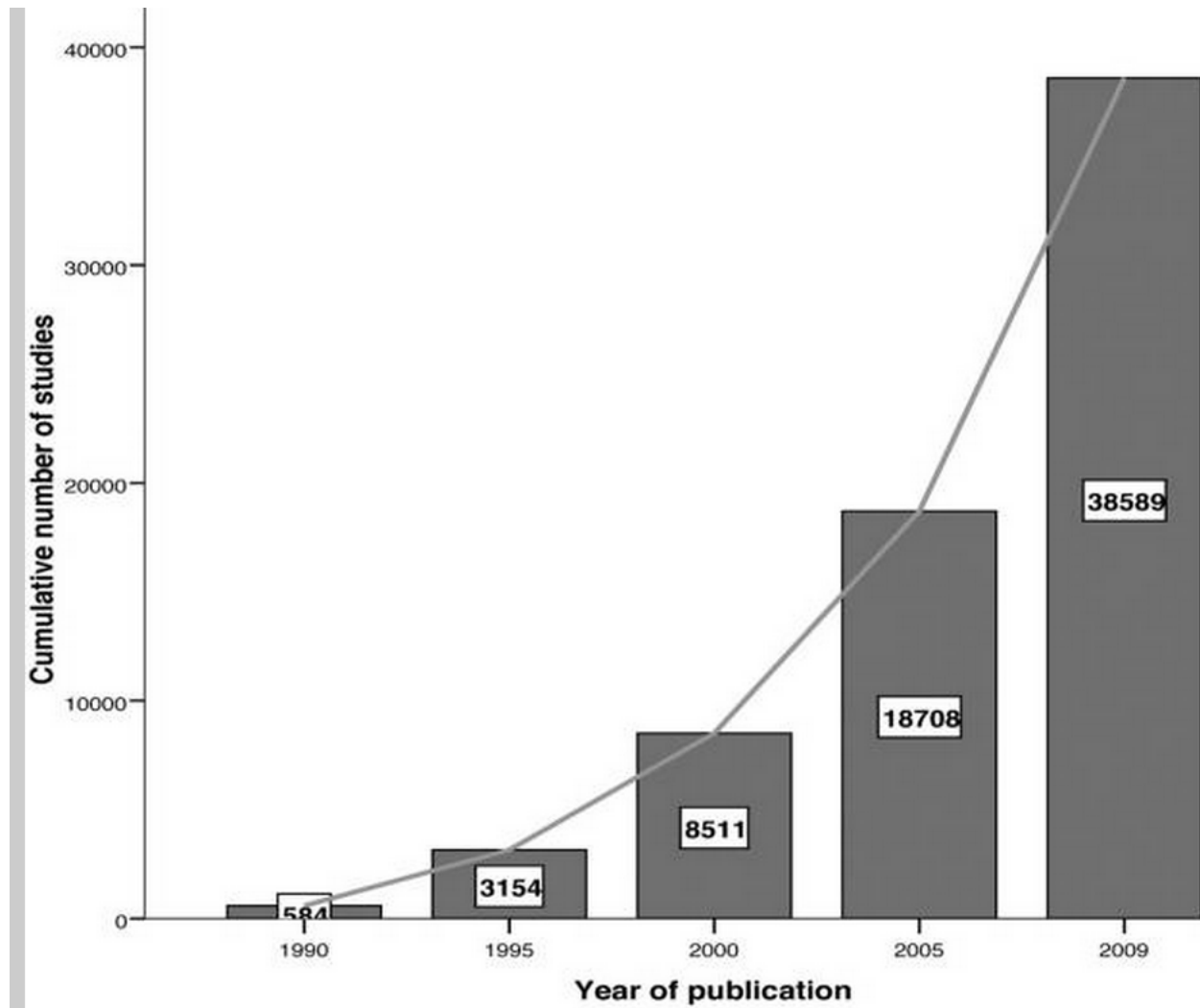
Metaanalýzy

Meta-analýza KH je technologie zobecnění výsledků několika nezávisle provedených KH. Tato technologie je založena na definování vstupních a vylučujících kritérií pro jednotlivá KH a na aplikaci speciálních statistických technik. Základem meta-analýz je hodnocení heterogenity výsledků jednotlivých KH. V případě průkazu výrazné heterogenity je tato interpretována, v případě průkazu homogenity jsou výsledky zobecňovány. Cílem meta-analýz je přinést ucelenou a klinicky relevantní informaci na základě výsledků jednotlivých KH, která mohou mít nejednotné závěry.

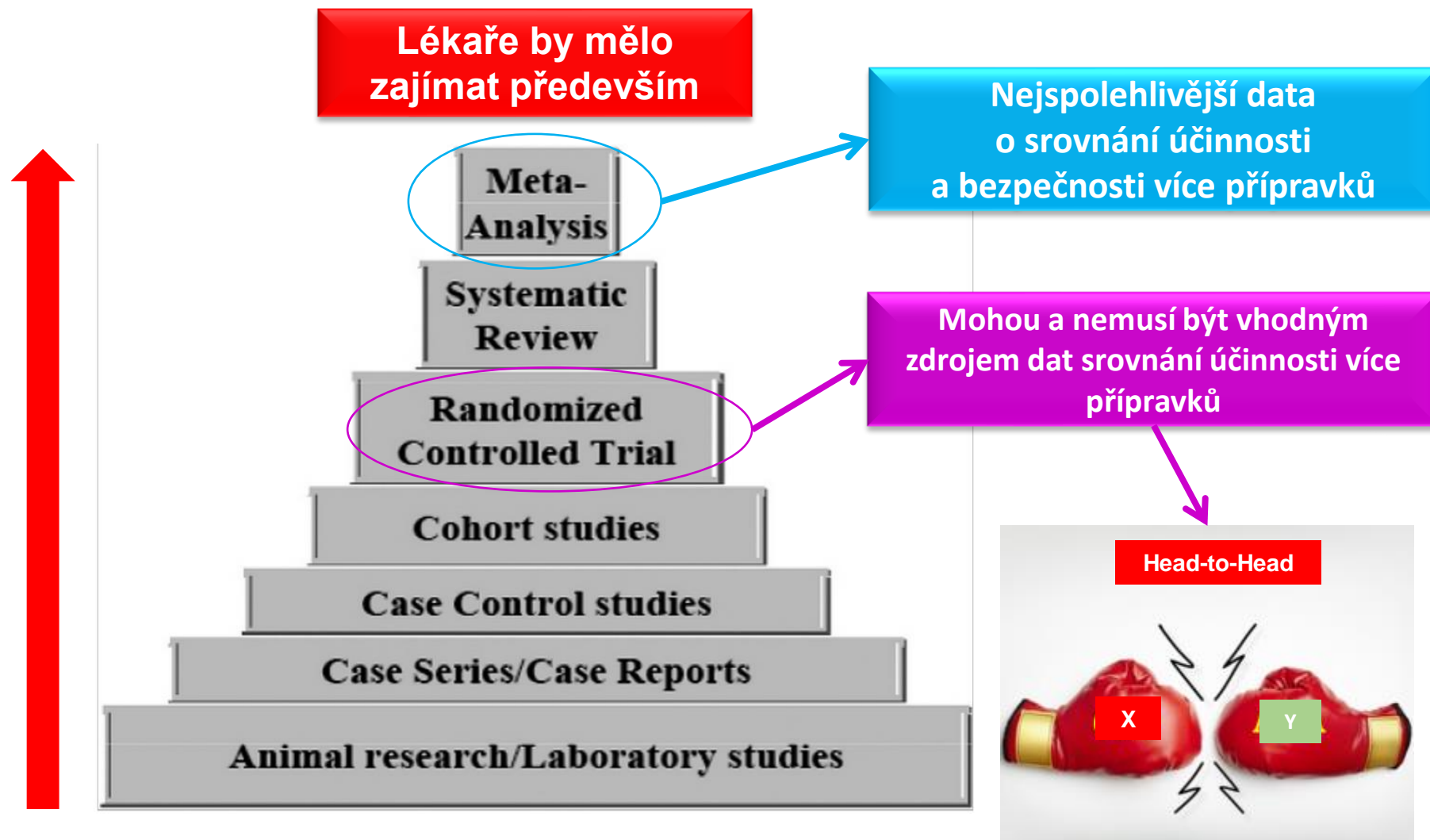
Vybrané problémy:

- „publication bias“ („funnel shape“)
 - vícenásobné publikace výsledků z jedné populace SH
 - nejednotnost hodnocení léčebné odpovědi a AE
-

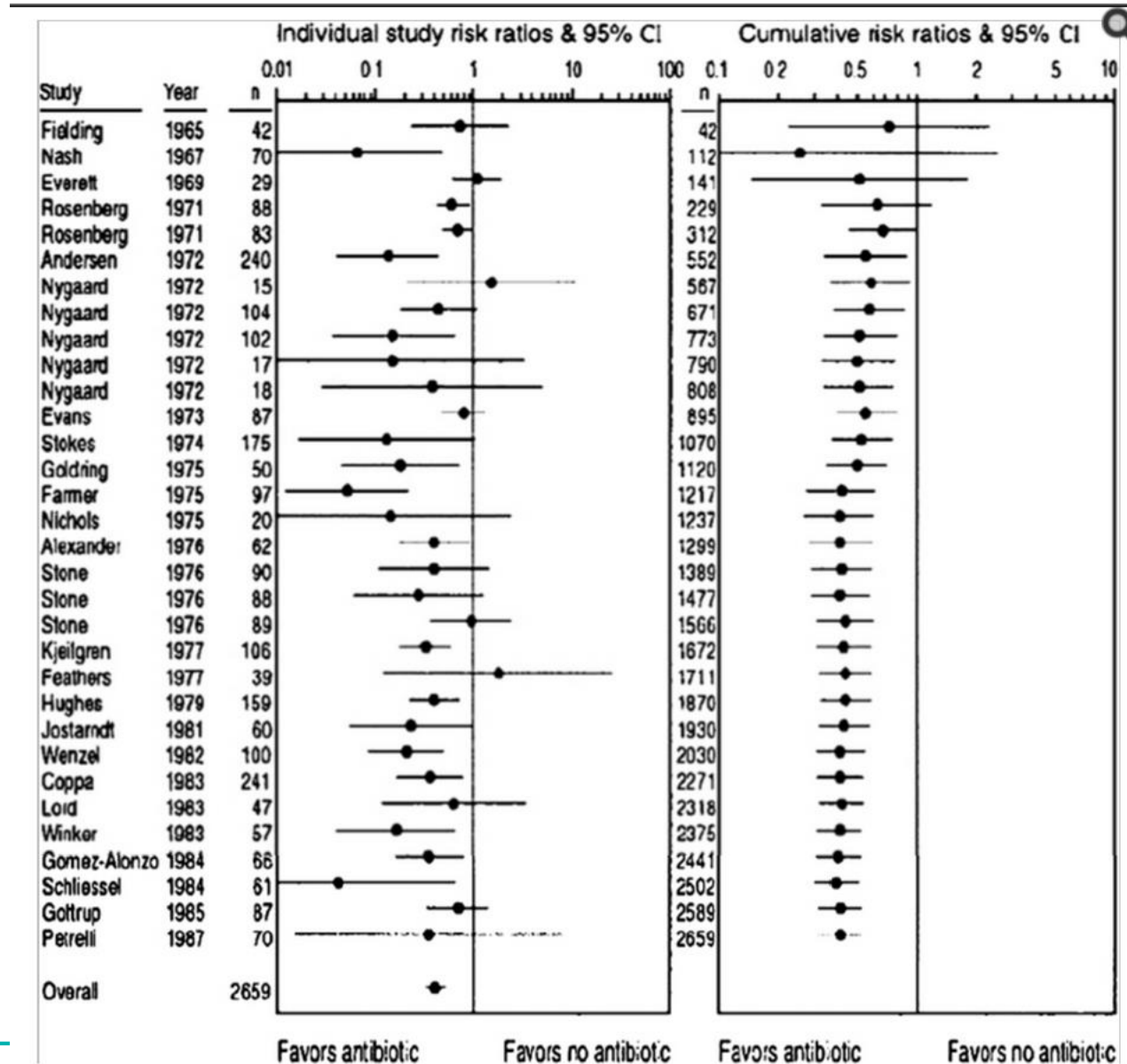
Vývoj počtu meta-analýz v čase



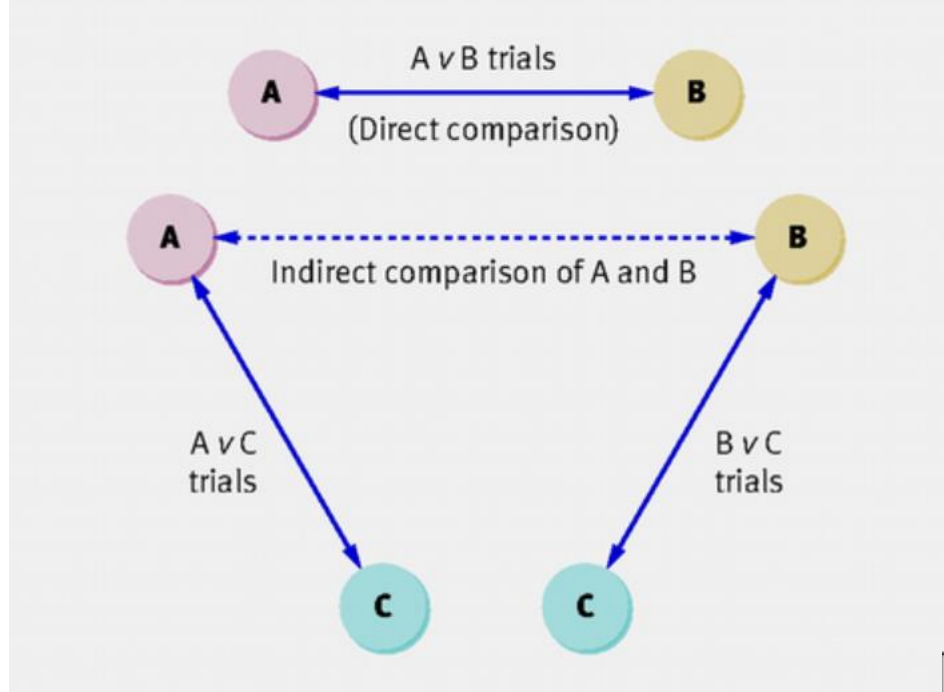
Levels of evidence



Grafické znázornění výsledků meta-analýz



„Mixed treatment comparisons“ a „Indirect comparisons“



Naive Approaches

Unadjusted indirect comparisons

- **Meta-analysis using data of single arms:**
 - MA using all data for intervention A,
 - MA using all data for intervention B,
 - MA using all data for intervention C, ...
- **Breaks randomisation**
- **Should never be used!**
- **A correct analysis has to be based on the estimated effects of each RCT**

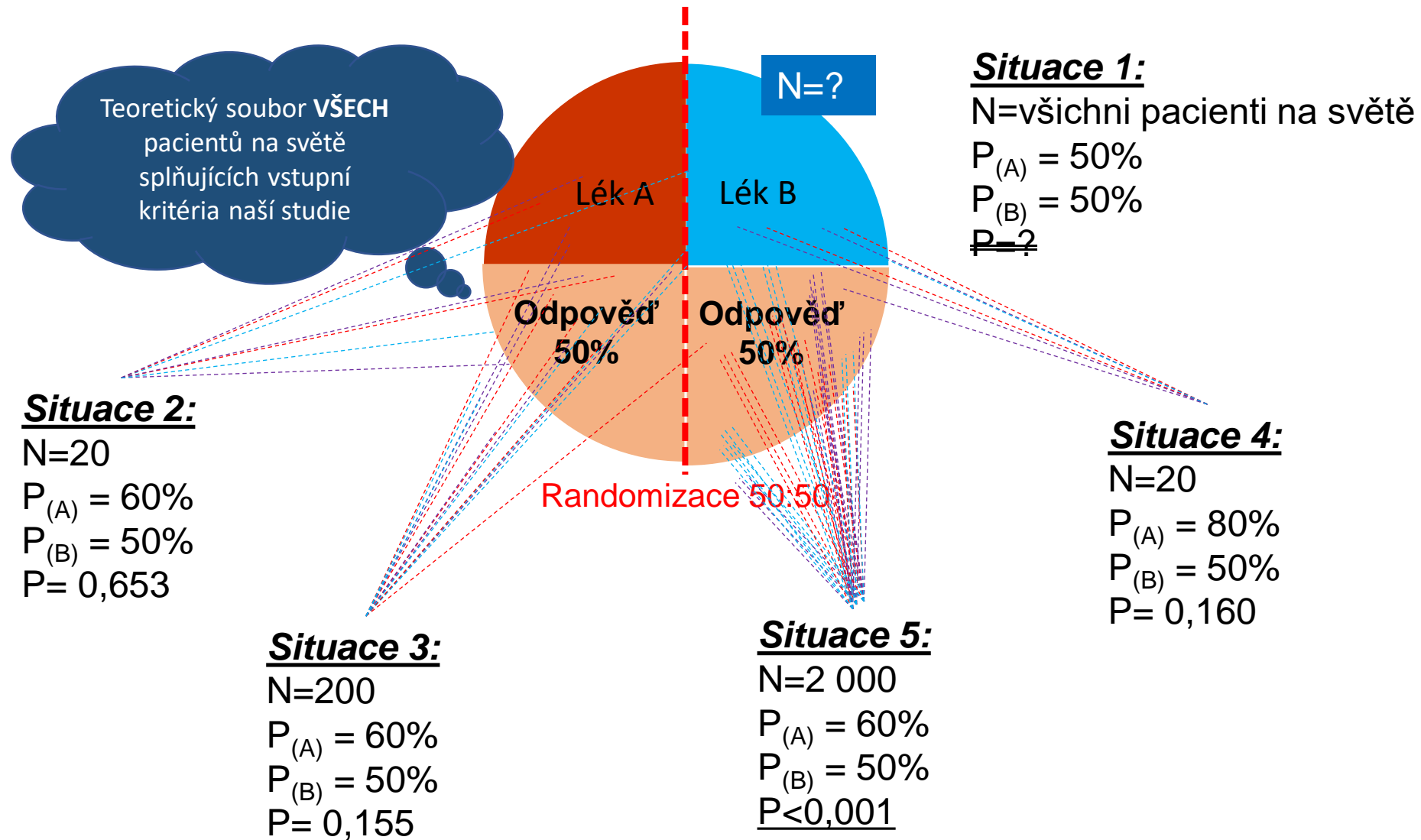
4.

Koncept P-hodnoty

Koncept P-value

- Co přesně a „laicky“ znamená hladina významnosti „P“, např. $(p < 0,0001)$?
 - Jak je zvolen počet pacientů potřebných pro zařazení do konkrétní studie?
 - Je důležitější hladina významnosti, nebo zjištěný rozdíl v účinnosti mezi rameny?
-

Modelová klinická studie ověřující účinnost nové léčby



Klinicky vs. statisticky významný rozdíl

Situace A:

N=20

$P_{(A)} = 80\%$

$P_{(B)} = 50\%$

P= 0,160

X

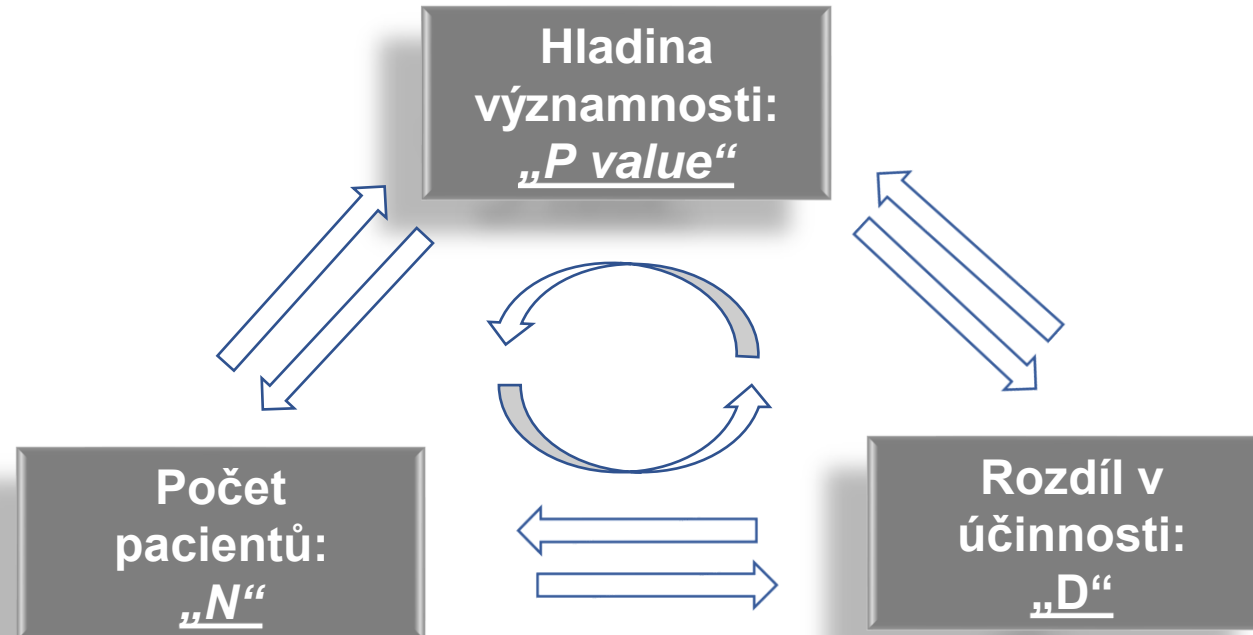
Situace B:

N=20 000

$P_{(A)} = 80\%$

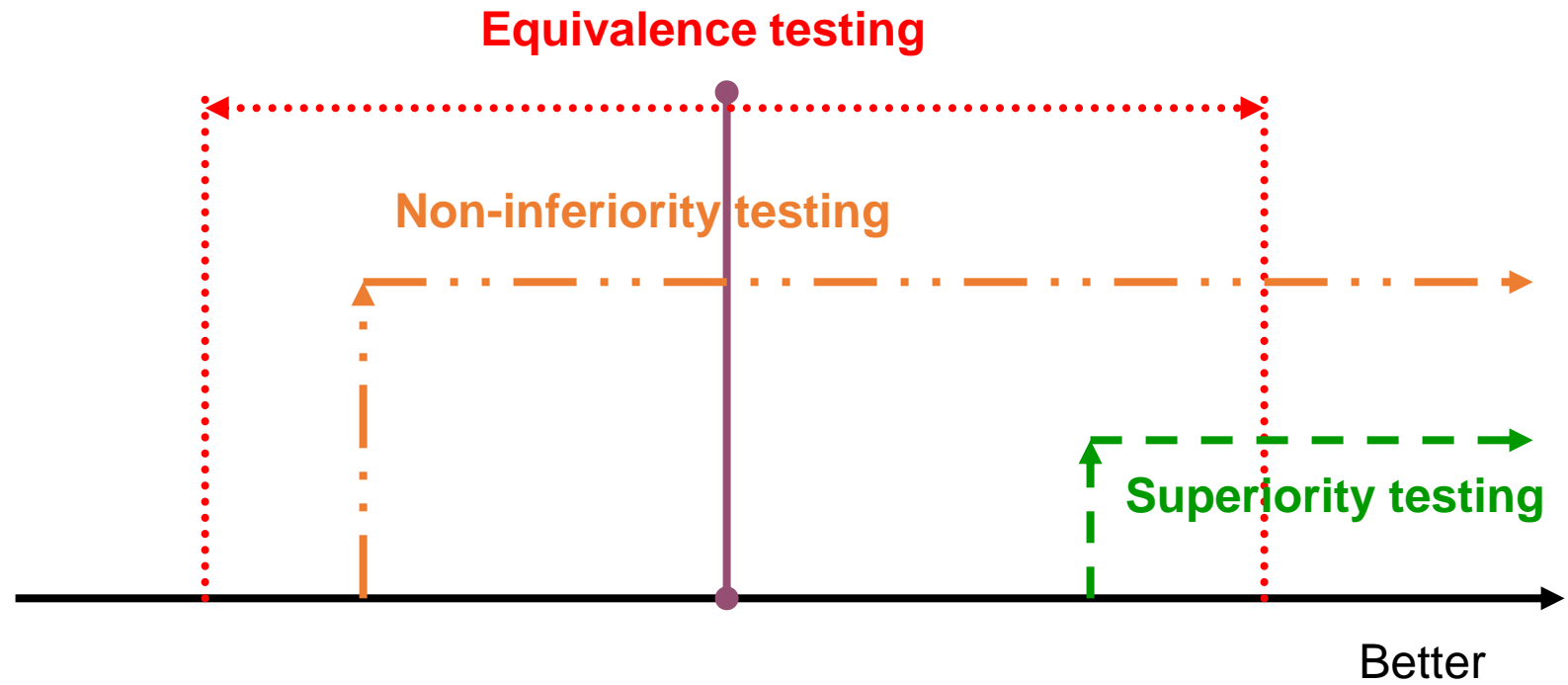
$P_{(B)} = 81\%$

P<0,05

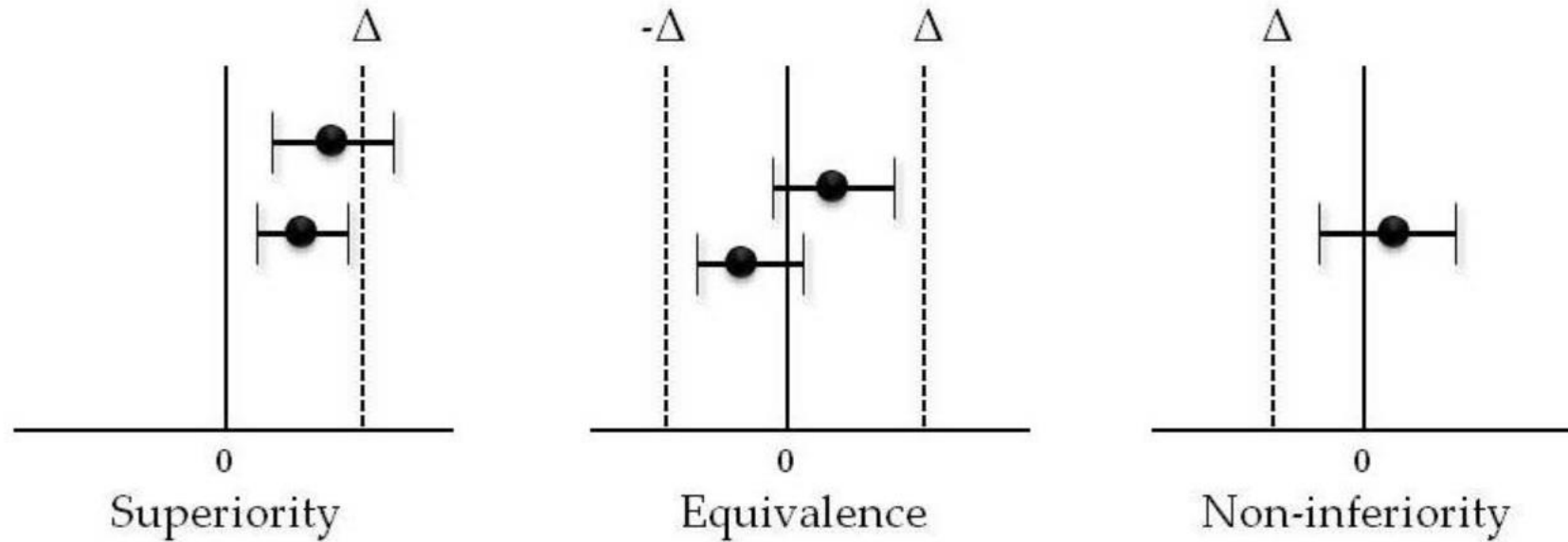


Typy hypotéz v klinických studiích

Primární hypotéza

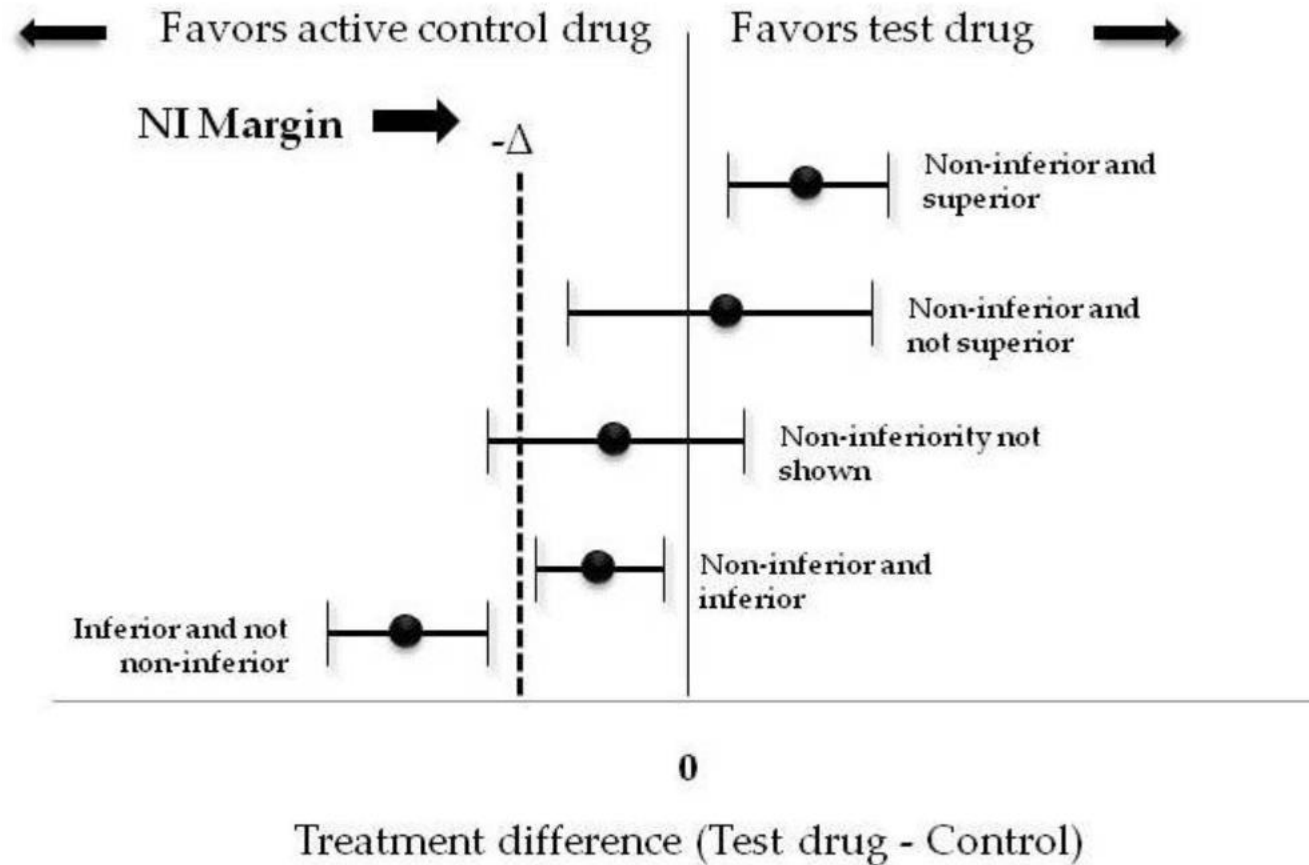


Typy hypotéz v klinických studiích



- „assay sensitivity“
- „bio-creep“ and „techno-creep“

Typy hypotéz v klinických studiích

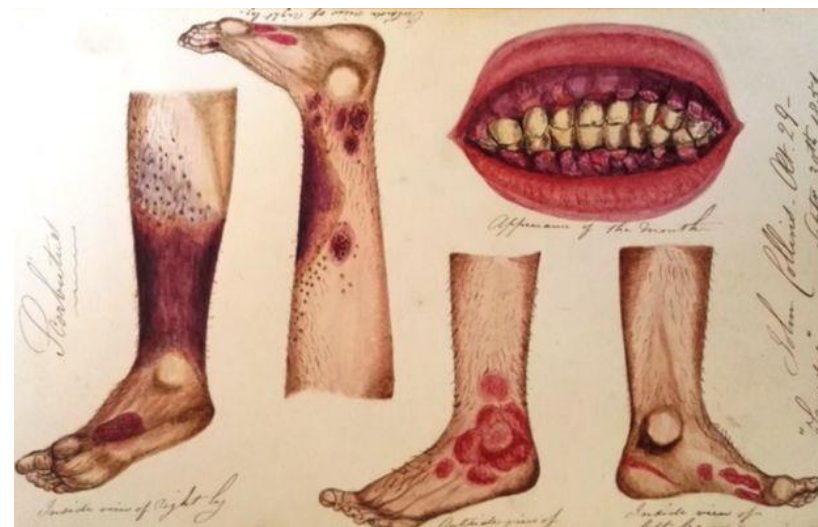


Kolik pacientů je potřeba
na studii?

James Lind, 1747,.....



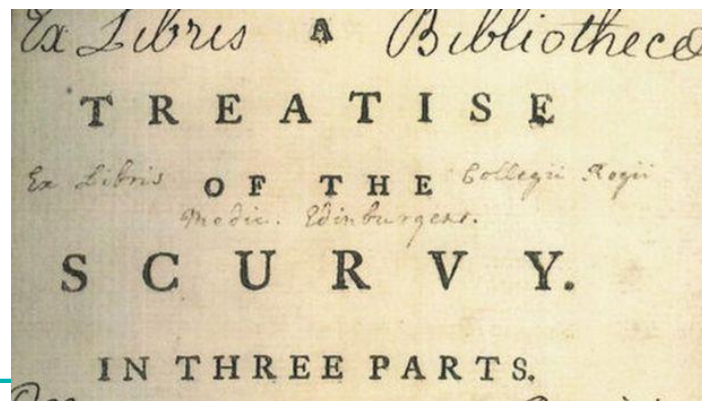
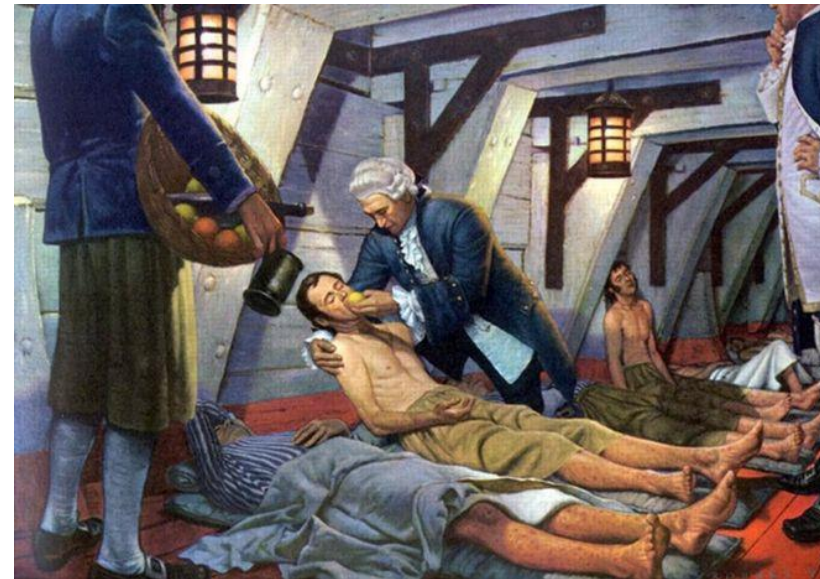
...hledání vhodné léčby na kurděje



Kolik pacientů je zapotřebí pro studii?

Klinická studie s 12 pacienty, léčebná ramena:

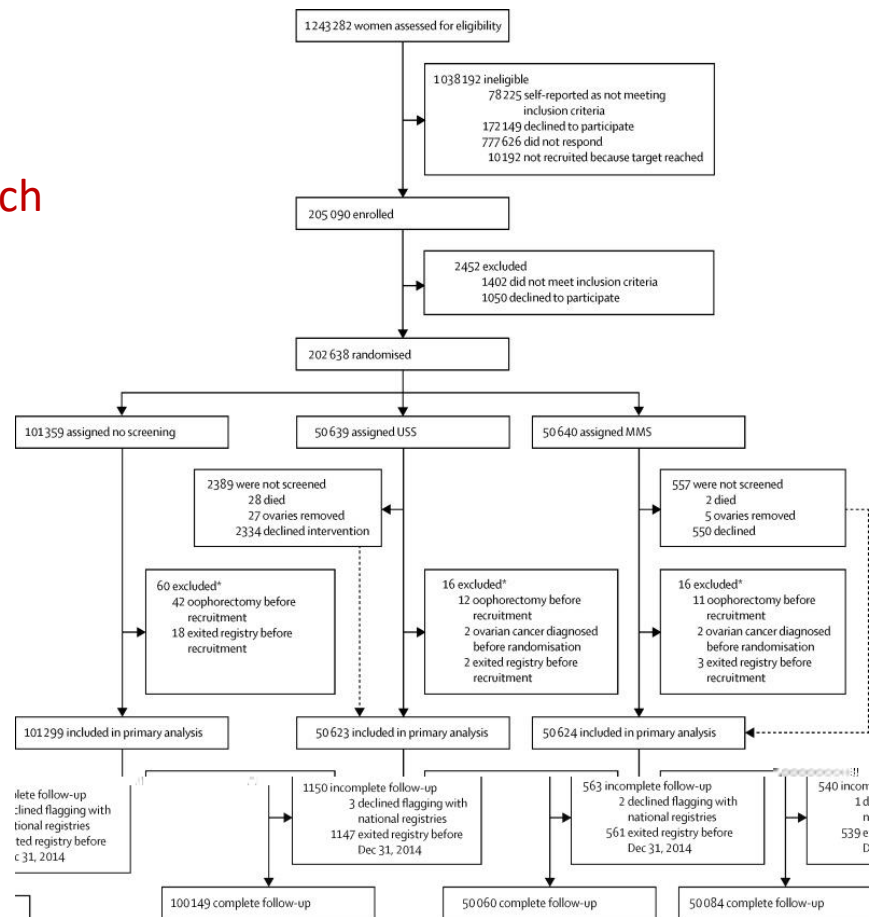
- cider
- vitriol 3x denně
- mořská voda
- bylinná směs a balsám z Peru
- vinný ocet
- dva pomeranče a jeden citrón denně



Kolik pacientů je zapotřebí pro studii?

Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomized controlled trial

Klinická studie:
1 243 282 pacientů,
202 638 randomizovaných



Proč je důležité optimalizovat prospektivně velikost vzorku?

- Etické aspekty
 - Ekonomické aspekty
 - Požadavek regulačních autorit (SÚKL, FDA)

 - **Riziko neplánovaného počtu subjektů hodnocení**
 - **Malý vzorek** – ztráta času, nemožnost prokázat rozdíl mezi srovnávanými rameny
 - **Velký vzorek** – ztráta času a prostředků, průkaz klinicky nevýznamného rozdílu mezi srovnávanými rameny
-

Jak je možné prakticky „ušetřit“ počet pacientů potřebných pro studii?

DESIGN

- Vhodně zvolená primární hypotéza
- Použití vhodných randomizačních technik
- Specifické designy jako např. N-of-1, apod.
- Pokud je to možné, využít cross-over design (možná úspora až 60-80% pacientů)
- Group-sequential design (možná úspora cca 30% pacientů) ale má svoje rizika, adaptivní design bez ovlivnění hladiny významnosti

ENDPOINTY

- Vhodná volba typu endpointu (spojitá vs. diskrétní data)
 - Využití kompozitních endpointů
 - U Time-to-event dat co nejdelší sledování (možná úspora cca 30% pacientů)
-

Příklad výpočtu velikosti vzorku (binární endopint)

Rameno P (placebo): Pokles bolesti u 30% SH

Rameno E (akupunktura): Pokles bolesti u 40% SH

Zvýšení podílu SH s poklesem symptomů o 10% je považováno za klinicky signifikantní

$$\alpha = 0,05$$

$$\beta = 0,20 \dots\dots \text{síla testu } (1-\beta) \dots\dots 0,80$$

$$p_1 = 0,30$$

$$p_2 = 0,40$$

N = 376 SH v jednom rameni KH

Příklad výpočtu velikosti vzorku (spojitý endpoint)

Rameno P (placebo): Pokles bolesti na 4,0 (stupnice 0-10)

Rameno E (akupunktura): Pokles bolesti na 3,0 (stupnice 0-10)

Snížení bolesti o 1,0 je považováno za klinicky signifikantní

$$\alpha = 0,05$$

$$\beta = 0,20 \dots\dots \text{síla testu } (1-\beta) \dots\dots 0,80$$

$$u_1 = 3$$

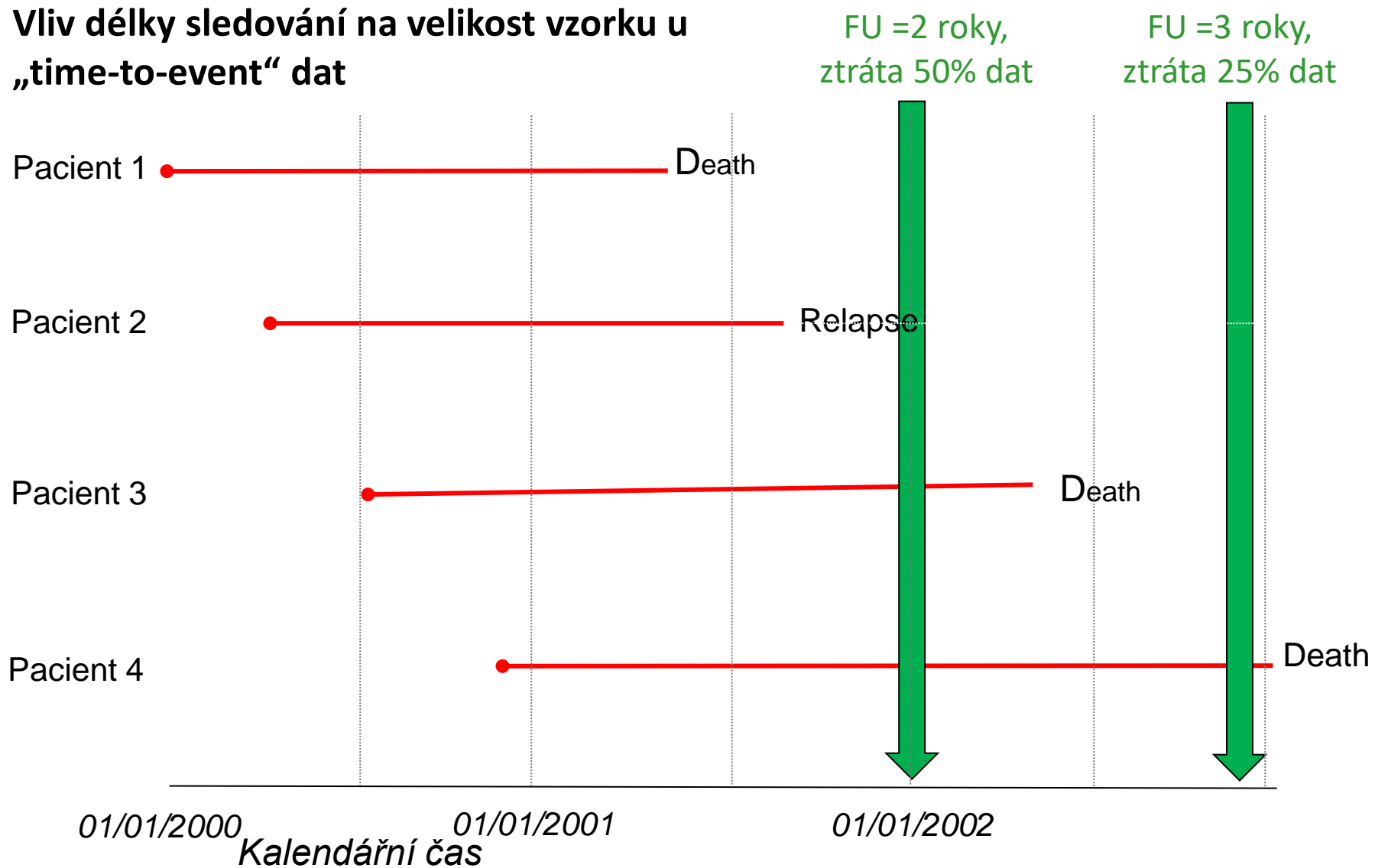
$$u_2 = 4$$

$$SD = 2$$

N = 64 SH v jednom rameni KH

Vliv endpointu na velikost vzorku

Vliv délky sledování na velikost vzorku u „time-to-event“ dat



Multiplicita

Představme si, že hodíme 100x mincí a padne nám 100x orel....

Jaká je pravděpodobnost, že mince není „cinknutá“?

Pravděpodobnost nastání této situace je $1/2^{100}$

Představme si, že máme 2^{100} mincí a každou hodíme 100x

Jaká je pravděpodobnost, že alespoň u jedné mince padne 100x orel?

63,2 %.....

Počet primárních endpointů


INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL
REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN
USE

ICH HARMONISED TRIPARTITE GUIDELINE

STATISTICAL PRINCIPLES FOR CLINICAL TRIALS

2.2.2 Primary and Secondary Variables

The primary variable ('target' variable, primary endpoint) should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial. There should generally be only one primary variable. This will usually be an efficacy variable, because the primary objective of most confirmatory trials is to provide strong scientific evidence regarding efficacy. Safety/tolerability may sometimes be the primary variable, and will always be an important consideration. Measurements relating to quality of life and health economics are further potential primary variables. The selection of the primary variable should reflect the accepted norms and standards in the relevant field of research. The use of a reliable and validated variable with which experience has been gained either in earlier studies or in published literature is recommended. There should be sufficient evidence that the primary variable can provide a valid and reliable measure of some clinically relevant and important treatment benefit in the patient population described by the inclusion and exclusion criteria. The primary variable should generally be the one used when estimating the sample size (see section 3.5).



*This Guideline has been
has been subject to consult
Process. At Step 4 of the
regulatory bodies of the E*

Počet primárních endpointů

Co-Primary Endpoints

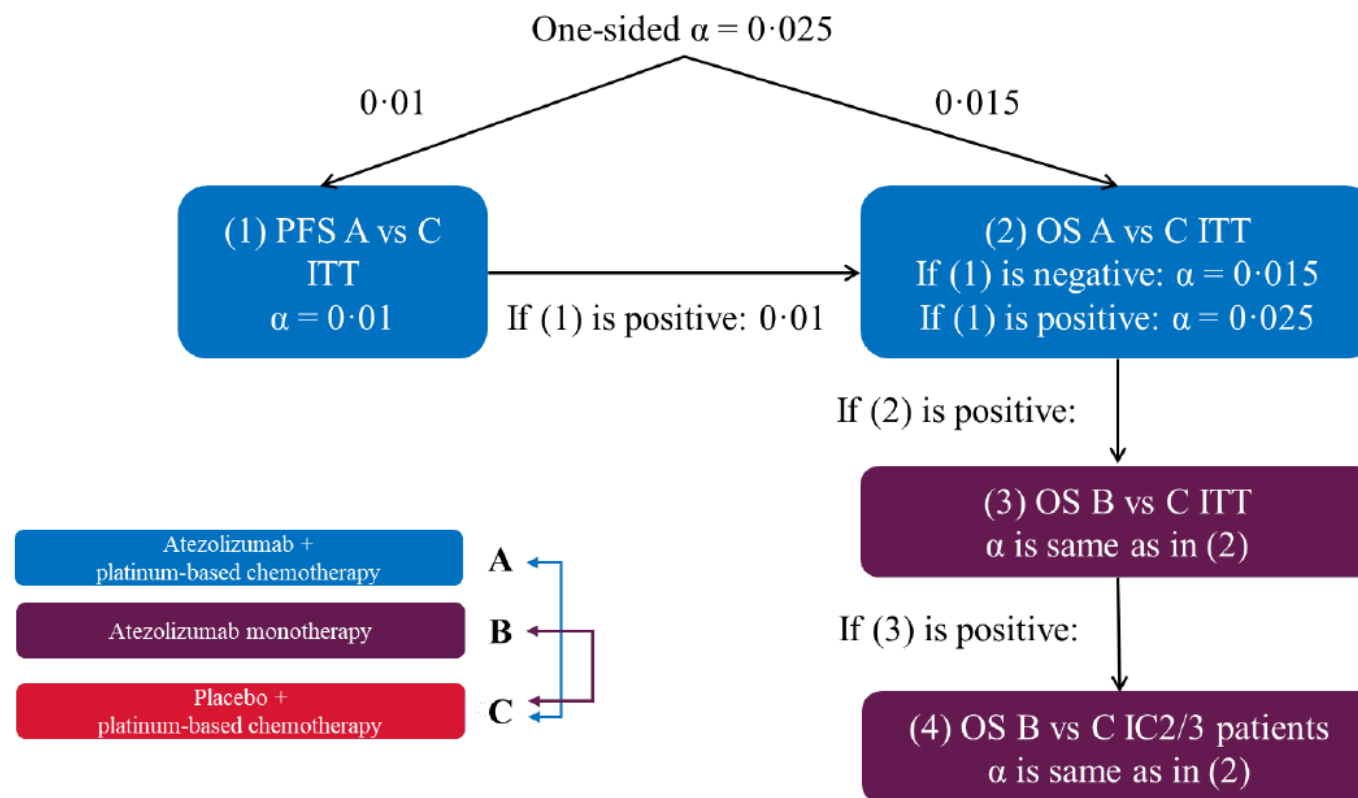
- In some cases, benefit on one endpoint may not be sufficient
 - In acute cardiac failure, immediate (e.g., within two hours) relief of symptoms may be necessary
 - Mid-term (e.g., 14 days) mortality may also be necessary
- In such cases, we refer to co-primary endpoints
- Both are necessary; not either/or

Composite Endpoints

- Different to co-primary
- Here, either of the separate endpoints may be sufficient
 - Common example in stroke . . . 28-day mortality or myocardial infarction or recurrent stroke
 - Having any one (or more) of these endpoints would be considered a treatment failure

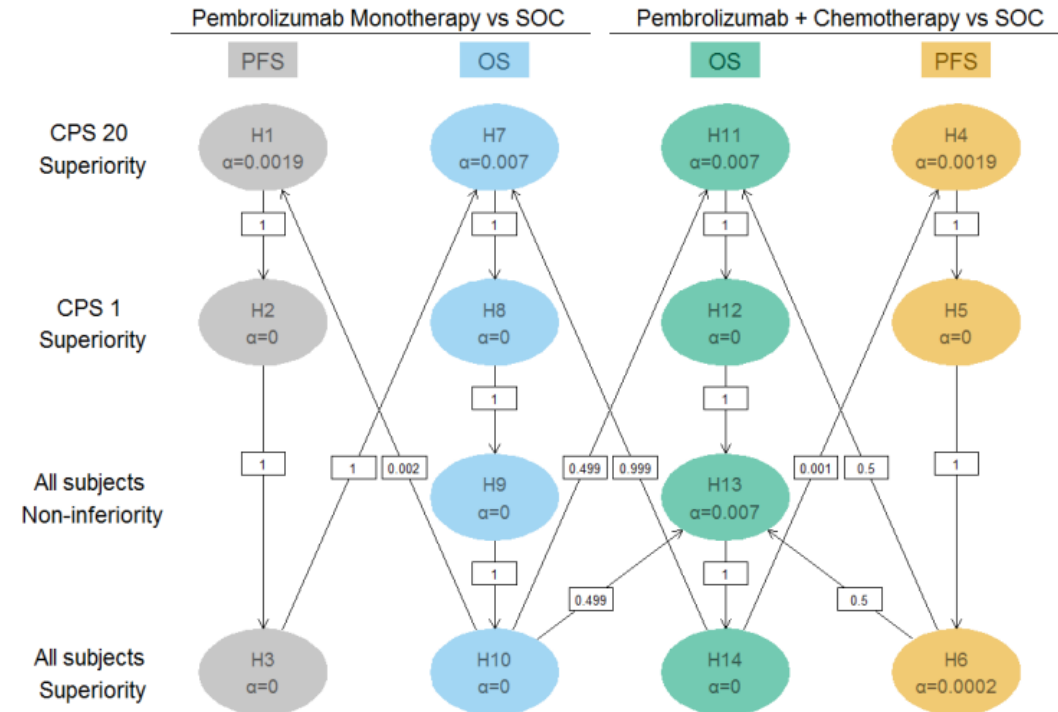
HIERARCHICKÉ TESTOVÁNÍ

Figure S2: Type I Error Control. Depiction of type I error control for co-primary endpoints PFS and OS, with alpha split, alpha recycling, and statistical testing hierarchy.



Příkladová studie: Alokace hladiny významnosti

Figure 3 displays the multiplicity strategy diagram for the study.



Endpointy

Examples of Outcome Types

Qualitative:

- **Binary:** Complete remission (yes vs no), Patient status (death vs alive)
- **Nominal:** Adverse event type
- **Ordinal:** Disease state (improved vs stable vs worsened), Neutropenia grade (1-4)

Quantitative:

- **Discrete:** pain scale (0=no pain, 10=worst pain imaginable)
- **Continuous:** Gene expression, Tumor size in mm, Visual analog scale in mm, Quality of Life score

Time-to-event: Time to death, Time to progression, Time to response

Typy endpointů v klinických studiích

Common Endpoints in Cancer Clinical Trials

Serial No	Endpoint	Definition	Unique feature
1.	Overall survival	Time from randomization to death.	The 'gold standard' primary clinical endpoint.
2.	Progression-free survival	Time from randomization to disease progression or death, whichever comes first.	Used to assess therapies targeting advanced or metastatic malignancies.
3.	Time to progression	Time from randomization to disease progression.	Only uses time to progression and does not include time to death.
4.	Event-Free Survival	Time from randomization to disease progression, discontinuation of treatment for any reason, or death.	Used to evaluate highly toxic treatments.
5.	Disease-free survival	Time from randomization to disease recurrence.	Used to assess adjunctive and curative therapies.
6.	Time to Treatment Failure	Time from initiation of chemotherapy to premature discontinuation of treatment.	Used with other endpoints to assess reasons for discontinuing treatment.
7.	Time to Next Treatment	Time from initiation of treatment to beginning the next line of therapy.	Used as a meaningful endpoint for patients with low grade, incurable malignancies.
8.	Duration of Clinical Benefit	Time from randomization to progression or death in patients who had a complete or partial response or a stable disease for over 24 weeks.	Used in settings where disease stabilization is meaningful.
9.	Duration of Response	Time from randomization to progression or death in patients who had a complete or partial response.	Used to assess therapies for durable response.
10.	Objective Response Rate	Proportion of patients with partial or complete response to therapy.	Used to assess neoadjuvant therapies.
11.	Complete Response	Lack of detectable evidence of tumor.	Included as a major goal of multiple myeloma treatment.
12.	Pathological Complete Response	Lack of residual invasive cancer in resected breast tissue or regional lymph nodes.	Used in accelerated approval for neoadjuvant therapies targeting breast cancer.
13.	Disease Control Rate	Percentage of patients with complete response, partial response, or stable disease as a result of their therapy.	Used to assess the tumorstatic efficacy of a therapy.
14.	Clinical Benefit Rate	Percentage of patients with complete response, partial response, or at least months of stable disease as a result of their therapy.	Used to capture tumorstatic efficacy of a therapy and stable disease.
15.	Health-Related Quality of Life	Assessment of patient quality of life with respect to health status.	Used to directly measure patient quality of life.
16.	Milestone survival	Survival probability at a prespecified time point.	Used to evaluate a cross-section of OS data.

Delgado A, Guddati AK. Clinical endpoints in oncology - a primer. Am J Cancer Res. 2021

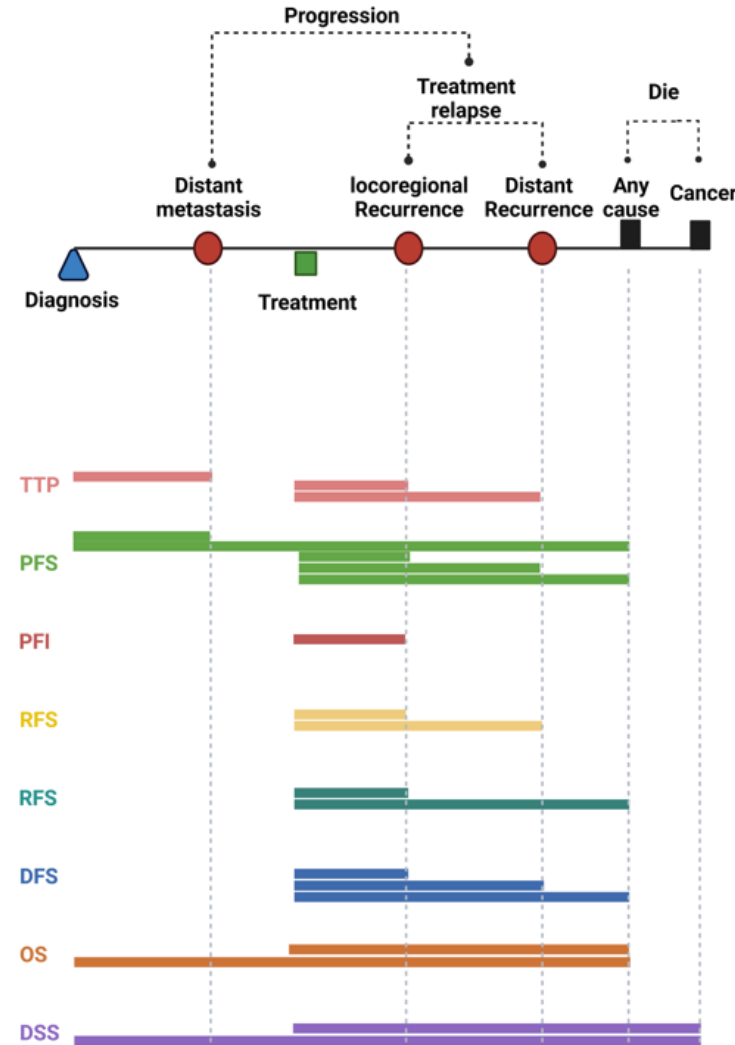
Definice vybraných time-to-event endpointů

Clinical Endpoints

Persistent disease	Disease left following initial surgery: residual tumor in the area of the primary tumor and/or regional lymph node
Recurrent disease	Biochemical and/or new structural disease in patient rendered disease-free
Local recurrence	Evidence of thyroid cancer in the thyroid bed
Regional recurrence	Nodal disease in central, retropharyngeal, or lateral compartments of lymph node
Distant recurrence	Disease recurrence beyond the regional nodal basins

Time-to-event Endpoints in Cancer

Time to progress	The time from diagnosis/treatment to the progression of tumor (in any aspect).
Progression-free survival	The time from diagnosis/treatment to the progression of tumor (in any aspect) or death (for any cause).
Progression-free interval	The time from treatment to locoregional recurrence.
Relapse-free survival	The time from treatment to the relapse (local, regional, distant).
Recurrence-free survival	The time from treatment to locoregional recurrence or death (for any cause)
Disease-free survival	The time from treatment to locoregional recurrence, metastasis or death (for any reason)
Overall survival	The time from diagnosis/treatment to death (for any reason).
Disease-specific survival	The time from diagnosis/treatment to death (cancer-specific).



Regulatory perspective: FDA

FDA guideline defines endpoints to be used in cancer clinical trials

Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Oncology Center of Excellence
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

December 2018
Clinical/Medical

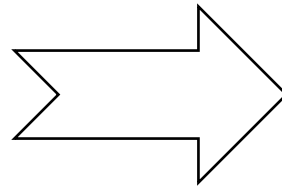


Table 2. Advantages and Disadvantages of Important Cancer Approval Endpoints

Endpoint	Advantages	Disadvantages
Overall Survival	<ul style="list-style-type: none"> Easily and precisely measured Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> May be affected by switch-over of control to treatment or subsequent therapies Needs longer follow-up Includes noncancer deaths
Symptom Endpoints (patient-reported outcomes)	<ul style="list-style-type: none"> Generally assessed earlier and with smaller sample size compared with survival studies 	<ul style="list-style-type: none"> Blinding is important for assessing the endpoint Potentially subject to assessment bias, particularly in open-label studies Lack of validated instruments in many disease areas Definitions vary among studies Balanced timing of assessments among treatment arms is critical
Disease-Free Survival or Event-Free Survival	<ul style="list-style-type: none"> Generally assessed earlier and with smaller sample size compared with survival studies Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> Potentially subject to assessment bias, particularly in open-label studies Definitions vary among studies Balanced timing of assessments among treatment arms is critical Includes noncancer deaths
Objective Response Rate	<ul style="list-style-type: none"> Generally assessed earlier and with smaller sample size compared with survival studies Effect on tumor attributable to drug(s), not natural history Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> Definitions vary among studies Frequent radiological or other assessments May not always correlate with survival
Complete Response	<ul style="list-style-type: none"> Generally assessed earlier and with smaller sample size compared with survival studies Effect on tumor attributable to drug(s), not natural history Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> Definitions vary among studies Frequent radiological or other assessments May not always correlate with survival
Progression-Free Survival or Time to Progression	<ul style="list-style-type: none"> Generally assessed earlier and with smaller sample size compared with survival studies Measurement of stable disease included Generally based on objective and quantitative assessment 	<ul style="list-style-type: none"> Potentially subject to assessment bias, particularly in open-label studies Definitions vary among studies Frequent radiological or other assessments Balanced timing of assessments among treatment arms is critical May not always correlate with survival

Rationale for the use of PFS as primary study endpoint

Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry


U.S. Department of Health and Human Services
Food and Drug Administration
Oncology Center of Excellence
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

December 2018
Clinical/Medical

- PFS is not confounded by subsequent therapy
- For a given sample size, the magnitude of effect on PFS can be larger than the effect on overall survival.
- Treatment effect measured by PFS can be a surrogate endpoint to support accelerated approval, a surrogate endpoint to support traditional approval, or it can represent direct clinical benefit based on the specific disease, context of use, magnitude of the effect, the disease setting, location of metastatic sites, available therapy, the risk-benefit relationship, and the clinical consequences of delaying or preventing progression

Regulatory perspective: EMA

EMA guideline defines the rules for the use of PFS and OS in cancer clinical trials



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

18 November 2023
EMA/CHMP/205/95 Rev.6
Committee for Medicinal Products for Human Use (CHMP)

Guideline on the clinical evaluation of anticancer medicinal products

Draft agreed by Oncology Working Party	24 March 2020
Adopted by CHMP for release for consultation	5 October 2020
Start of public consultation	13 November 2020
End of consultation (deadline for comments)	15 February 2021
Adoption at the ONCWP	19 May 2023
Adoption at the CHMP	18 November 2023

Keywords: Cancer, malignancy, biomarker, targeted drugs, pharmacogenomics

Official address: Domenico Scarlattilaan 6 • 1083 HS Amsterdam • The Netherlands
Address for visits and deliveries: Refer to www.ema.europa.eu/how-to-find-us
Send us a question: Go to www.ema.europa.eu/contact Telephone +31 (0)88 781 6000

© European Medicines Agency, 2024. Reproduction is authorised provided the source is acknowledged.

- If PFS is the selected primary endpoint, OS should be reported as a secondary and vice versa.
- When OS is reported as secondary endpoint, the estimated treatment effect on OS should ensure that there are no relevant negative effects on this endpoint, in most cases by showing trends towards superiority.
- In situations where there is a large effect on PFS, or if there is a long expected survival after progression, and/or a clearly favourable safety profile, precise estimates of OS may not be needed for approval.

Number needed to treat

Number needed to treat (NNT)

Tento ukazatel je jen jinou formou vyjádření ARR a počítá se jako převrácená hodnota ARR, tedy $1/ARR$. Při použití předchozího uvedeného příkladu by tedy hodnota NNT byla $1 / 0.1=10$. Tato hodnota se interpretuje jako počet pacientů, které je nutné léčit úspěšnější léčbou abychom zabránili jednomu negativnímu výstupu léčby ve srovnání s méně efektivní léčbou. V našem případě to tedy znamená, že na 10 pacientů léčených léčbou A bude jeden pacient u kterého zabráníme úmrtí do pěti let ve srovnání se situací, kdy jsou všichni pacienti léčeni léčbou B.

To gain a rapid picture of the benefit (B) or harm of a given treatment, it is sometimes useful to calculate the number of people that would need to be treated to save or lose one life. That number is simply the reciprocal of the difference in absolute (not relative) risk (R). Considering again the example of thrombolytic therapy for an acute MI, the absolute risk difference is $12\% - 9\%$, or 3% (B), so the number needed to treat (NNT) is $1/0.03$ or 33. In the setting of an acute MI, one would need to treat 33 patients with thrombolytic therapy to save one life. On the other hand, the absolute risk difference for an intracranial hemorrhage is 1% (R), so the number needed to harm (NNH) is $1/0.01$ or 100. One would need to give 100 patients thrombolytic therapy to cause one fatal intracranial hemorrhage. These simple calculations do not weigh the relative value of the benefit and the harm, but the NNT and the NNH can provide a useful snapshot of benefits and risks.

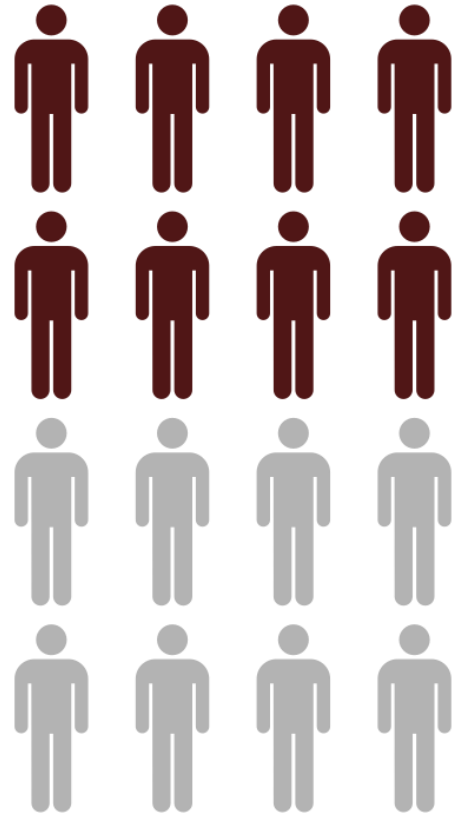
NNT: How is it calculated?

Let's have a clinical trial comparing interventional treatment A with control treatment B:

Treatment A:
Progression in 25% pts



Treatment B:
Progression in 50% pts



Calculation of NNT:

Risk of progression on treatment A: 0,25

Risk of progression on treatment B: 0,50

$$\text{NNT} = 1 / (0,50 - 0,25) = \underline{4}$$

Interpretation of NNT:

Four patients need to be treated by treatment A to avoid progression in one patient (in comparison to treatment B).

Relativní riziko = Relative risk

- ➔ Výpočet relativního rizika (RR) umožňuje srovnat pravděpodobnosti výskytu sledovaného jevu ve dvou různých skupinách.
- ➔ 1. skupina – **experimentální nebo skupina s expozicí určitému faktoru**
- ➔ 2. skupina – **kontrolní nebo skupina bez expozice**

$$RR = \frac{\text{Pravděpodobnost výskytu jevu v 1. skupině (experimentální)}}{\text{Pravděpodobnost výskytu jevu ve 2. skupině (kontrolní)}} = \frac{P_1}{P_0}$$

Sledovaný jev	Skupina		Celkem
	Experimentální	Kontrolní	
Ano	a	b	$a + b$
Ne	c	d	$c + d$
Celkem	$a + c$	$b + d$	n




$$RR = \frac{P_1}{P_0} = \frac{\frac{a}{a+c}}{\frac{b}{b+d}}$$

Poměr šancí = Odds ratio

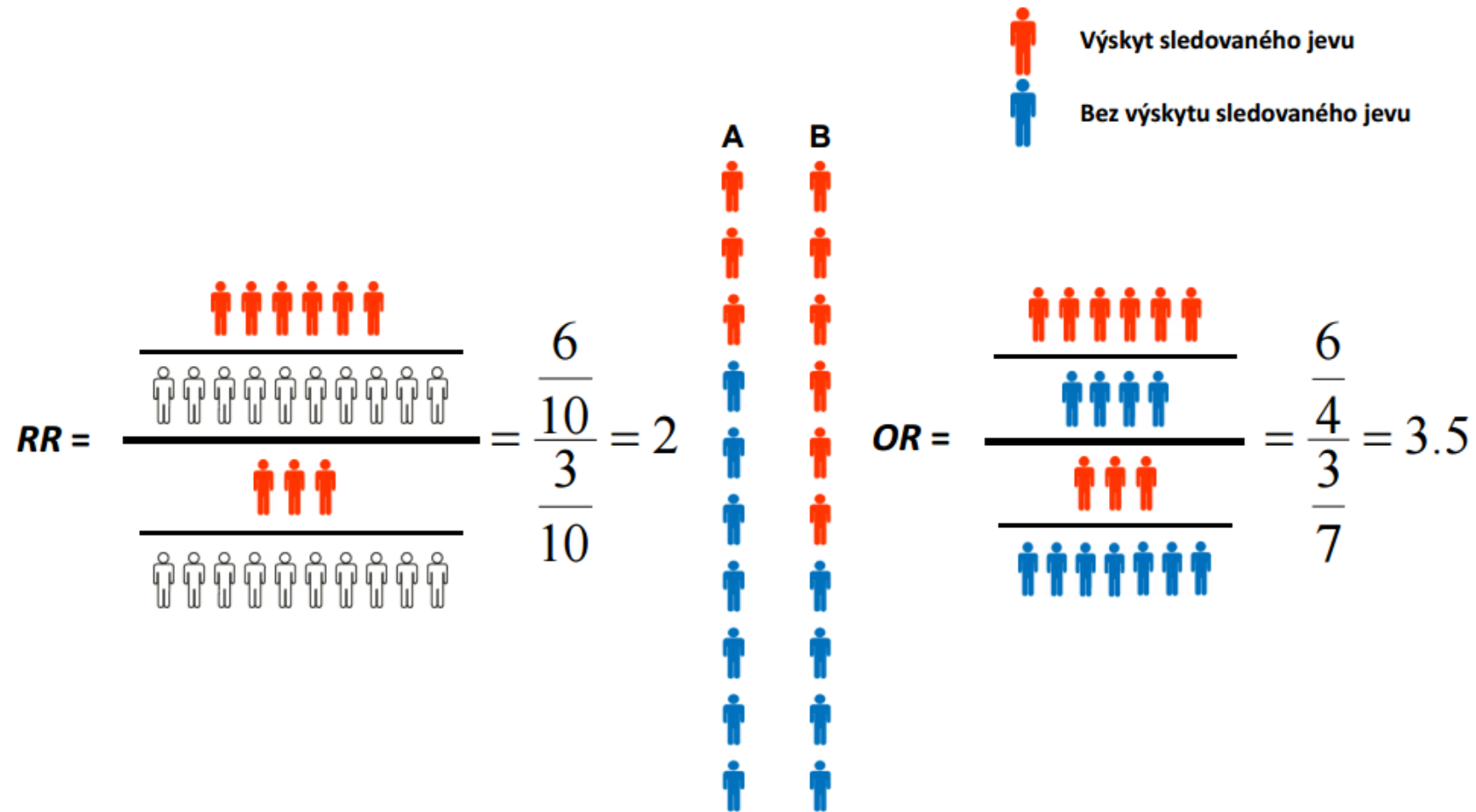
- Poměr šancí (OR) je další charakteristikou, která umožňuje srovnat výskyt sledovaného jevu ve dvou různých skupinách.
- 1. skupina – **experimentální nebo skupina s expozicí určitému faktoru**
- 2. skupina – **kontrolní nebo skupina bez expozice**

$$OR = \frac{\frac{\text{Pravděpodobnost výskytu jevu v 1. skupině (experimentální)}}{1 - \text{Pravděpodobnost výskytu jevu v 1. skupině (experimentální)}}}{\frac{\text{Pravděpodobnost výskytu jevu ve 2. skupině (kontrolní)}}{1 - \text{Pravděpodobnost výskytu jevu ve 2. skupině (kontrolní)}}} = \frac{O_1}{O_0} = \frac{\frac{P_1}{1-P_1}}{\frac{P_0}{1-P_0}}$$

Sledovaný jev	Skupina		Celkem
	Experimentální	Kontrolní	
Ano	a	b	$a + b$
Ne	c	d	$c + d$
Celkem	$a + c$	$b + d$	n


$$OR = \frac{\frac{P_1}{1-P_1}}{\frac{P_0}{1-P_0}} = \frac{a}{b} \cdot \frac{d}{c}$$

Grafické srovnání *RR* a *OR*



Hazard Ratio (HR)

Hazard ratio is a measure of how often a particular event happens in one group compared to how often it happens in another group, over time

Calculation of Hazard ratio

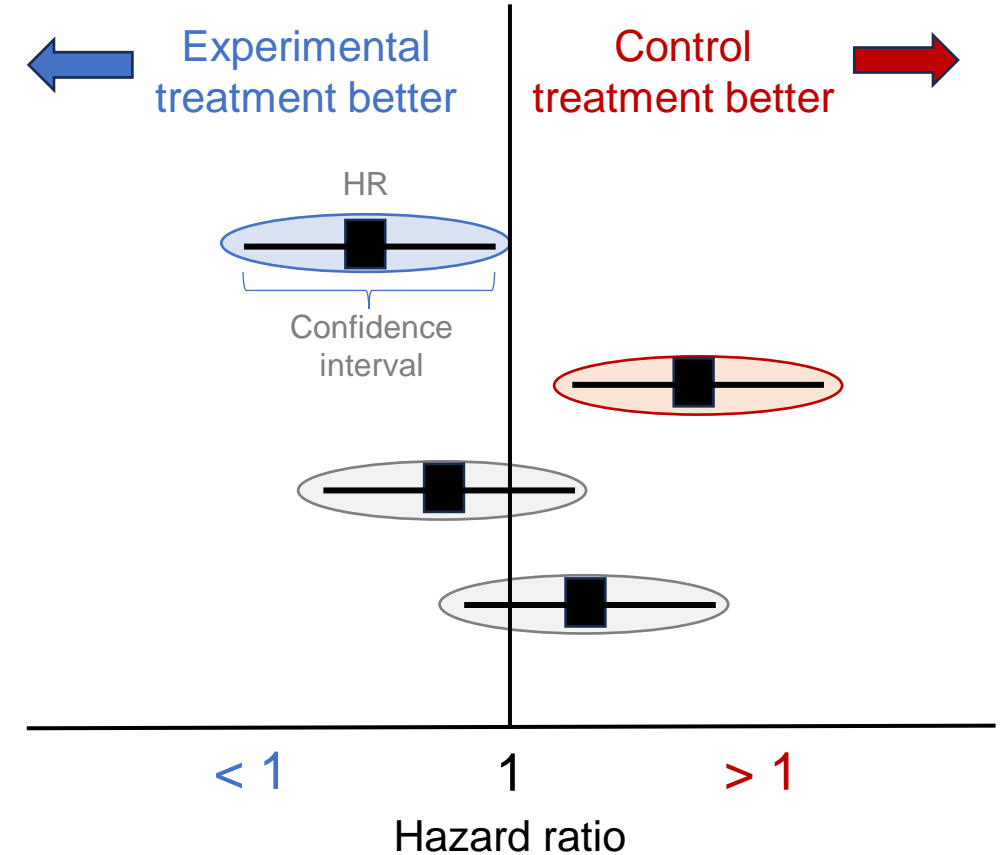
$$\text{HR} = \frac{\text{Risk of event on experimental treatment}}{\text{Risk of event on control treatment}}$$

< 1 Experimental treatment is better

$= 1$ No difference

> 1 Control treatment is better

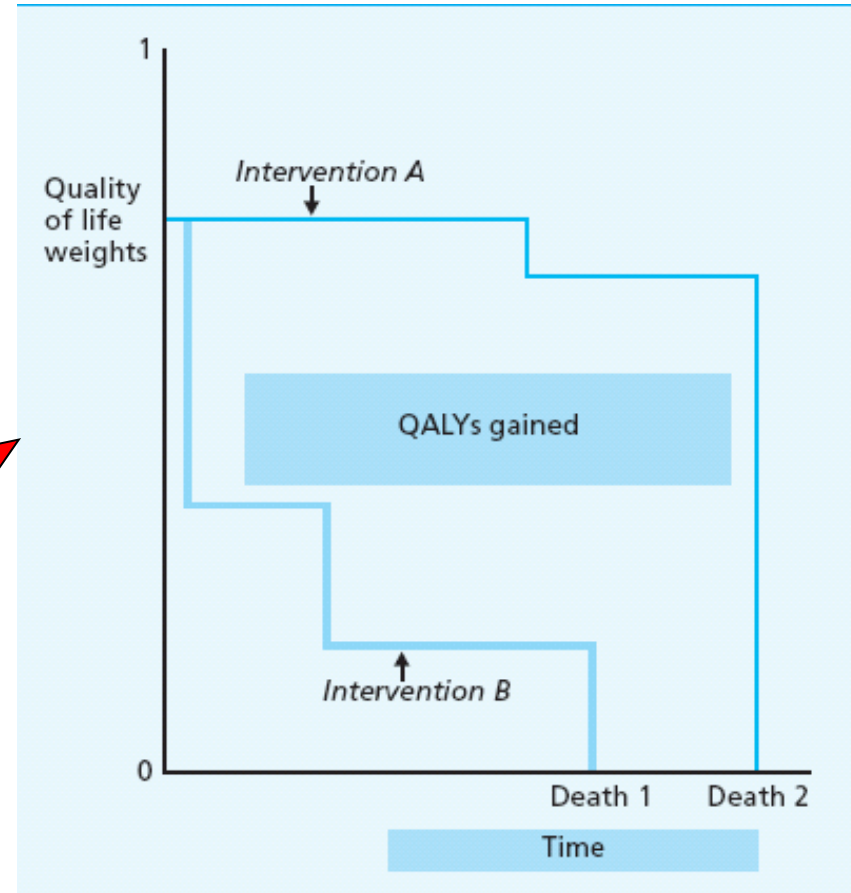
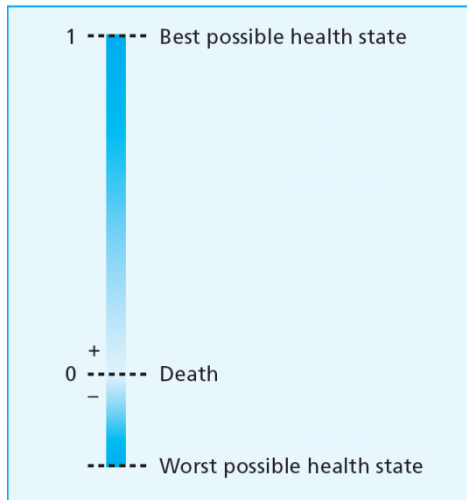
Hazard ratio in forest plot



95% confidence interval for Hazard Ratio (e.g. HR=0.68: 0.42-0.88) must not include 1 to be statistically significant.

Kvalita života

Parametr QALY (quality-adjusted life years) umožňuje indexové posouzení délky a kvality života jako benefitu použití léčebného postupu



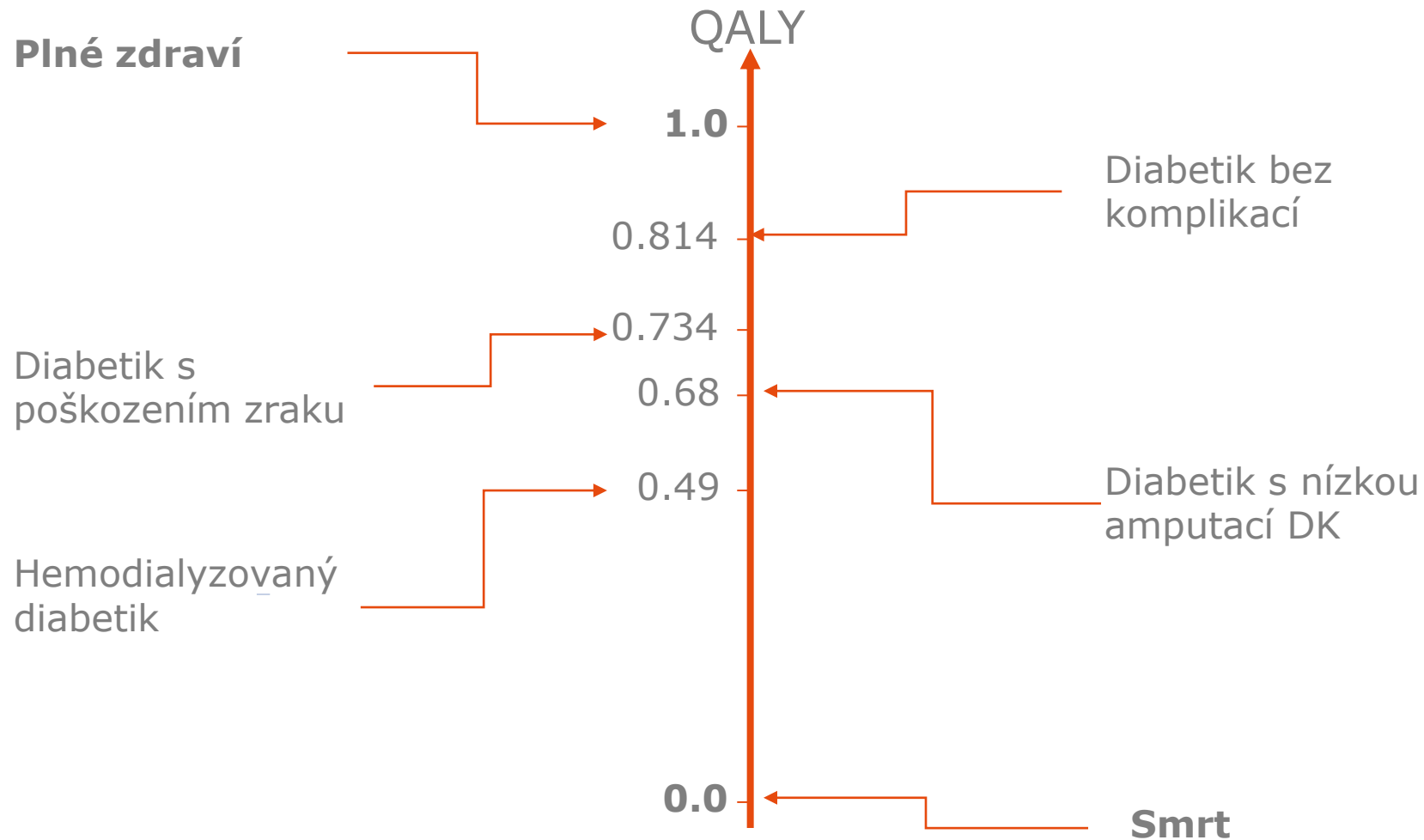
Calculating QALYs: an example

Intervention A: Four years in health state 0.75	3 QALYs
Intervention B: Four years in health state 0.5	2 QALYs
<u>Additional number of QALYs generated by A</u>	<u>1 QALY</u>

Intervention	£/QALY at 1990 prices
Cholesterol testing and diet therapy (all adults aged 40–69)	220
Neurosurgical intervention for head injury	240
GP advice to stop smoking	270
Neurosurgical intervention for subarachnoid haemorrhage	490
Antihypertensive treatment to prevent stroke (ages 45–64)	940
Pacemaker implantation	1,100
Hip replacement	1,180
Valve replacement for aortic stenosis	1,410
Cholesterol testing and treatment (all adults aged 40–69)	1,480
Docetaxel (as opposed to paclitaxel) in treatment of recurrent metastatic breast cancer	1,890*
CABG (left main-vessel disease, severe angina)	2,090
Kidney transplantation	4,710
Breast cancer screening	5,780
Heart transplantation	7,840
Cholesterol testing and treatment incrementally (all adults aged 25–39)	14,150
Home haemodialysis	17,260
CABG (one-vessel disease, moderate angina)	18,830
Hospital haemodialysis	21,970
Erythropoietin treatment for anaemia in dialysis patients (assuming 10% reduction in mortality)	54,380
Addition of interferon- α 2b to conventional treatment in newly diagnosed multiple myeloma	55,060 ⁵
Neurosurgical intervention for malignant intracranial tumours	107,780
Erythropoietin treatment for anaemia in dialysis patients (assuming no increase in survival)	126,290

* Adjusted to 1990 prices using *Hospital and Community Health Service Pay and Prices Index, Unit Costs of Health and Social Care*. PPSRU, 1996. $(2,431 \div 200.7 \times 155.6 = 1,890)$. ⁵ Translated into 1990 prices, as above

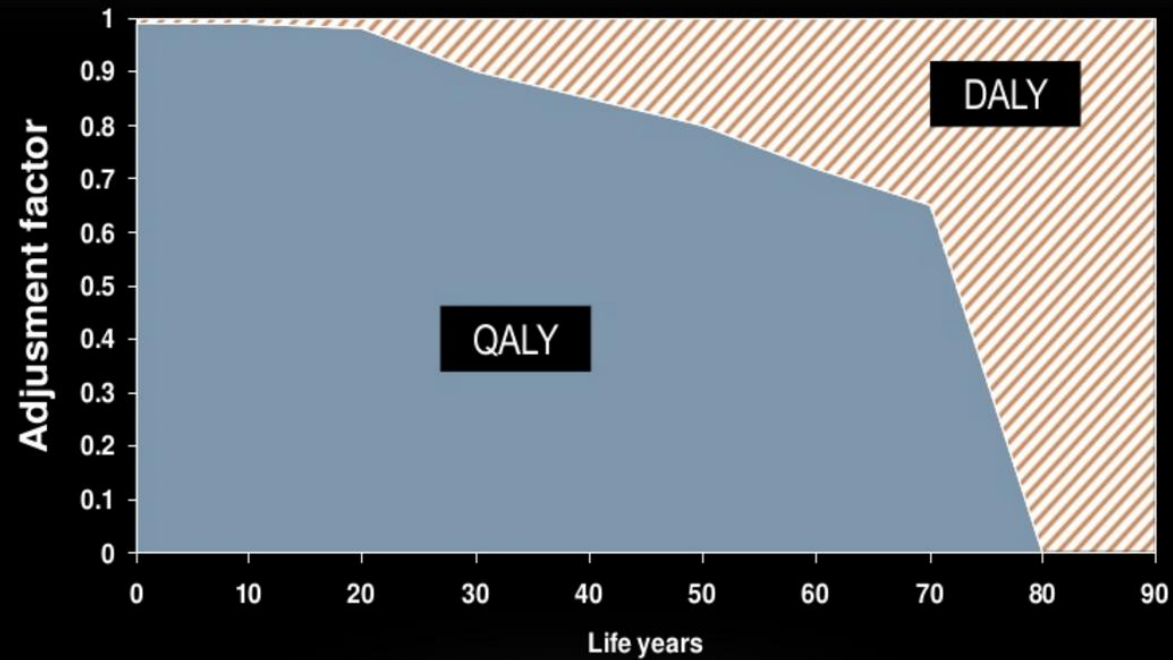
QALY: příklad u diabetiků



Ilavská, Tomek 2006: Farmakoekonomické posouzení léčby diabetologií, podle:

Clarke P et al (2002) Med Decision Making, 22 (4) (UKDPS 62), Teng TO et al (2000) Med Care, 38 (6)

QALYs and DALYs



Inkrementální náklady: ICER

Inkrementální náklady jsou definovány jako rozdíl mezi náklady dvou technologií, které jsou neslučitelné.

Pro srovnání dvou technologií z hlediska jejich nákladů a přínosů se používá tzv. *inkrementální poměr*, který vyjadřuje náklady na jednotku účinnosti nové technologie ve srovnání se stávající:

$$\text{ICER} = (\text{NÁKLADY HI} - \text{NÁKLADY SI}) / (\text{PŘÍNOSY HI} - \text{PŘÍNOSY SI})$$

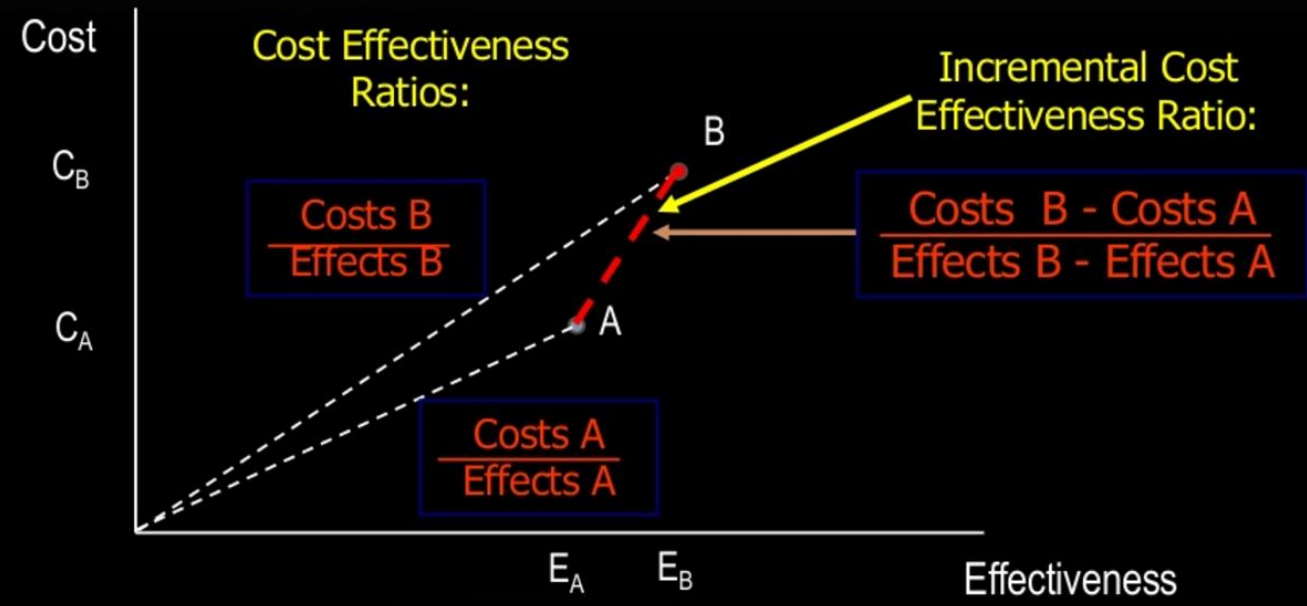
HI – hodnocená intervence *SI* – srovnávaná intervence

Incremental Cost-Effectiveness Ratio (ICER)

$$\text{ICER} = \frac{(C_1 - C_0)}{(E_1 - E_0)}$$

C_1 = cost in intervention group
 C_0 = cost in control group
 E_1 = effect in intervention group
 E_0 = effect in control group

INCREMENTAL COST – EFFECTIVENESS RATE (ICER)



Redukce rizika

Absolutní redukce rizika (ARR)

Tento parametr vyjadřuje pouze rozdíl v absolutní hodnotě rizika negativního výstupu léčby mezi srovnávanými skupinami. Pokud je tedy např. primárním cílem studie prodloužit pětileté celkové přežití pacientů s určitou onkologickou diagnózou a ve skupině A bude absolutní riziko úmrtí do pěti let 10% a ve skupině B 20%, tak absolutní redukce rizika při použití léčby A je 20%-10%, tedy 10%.

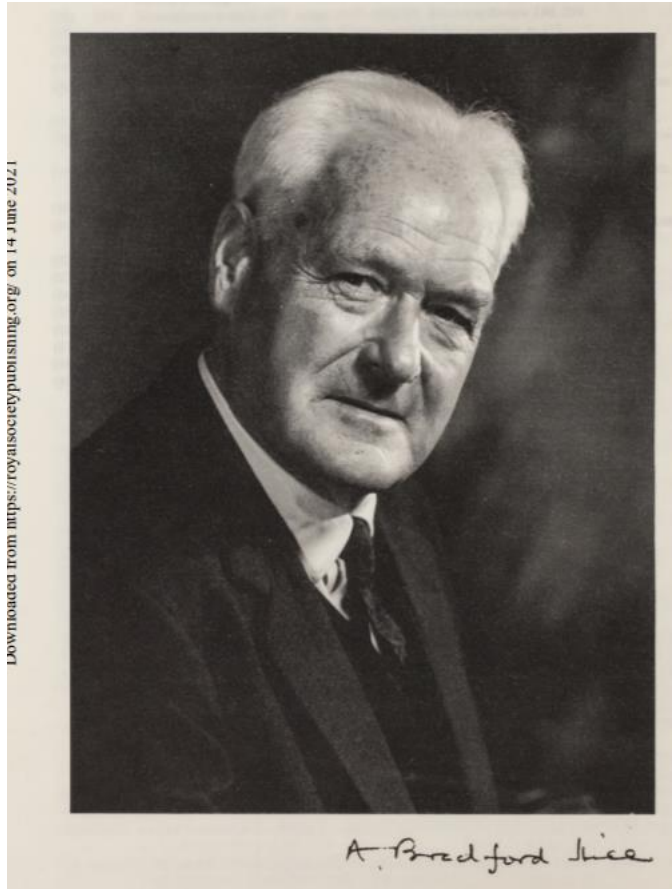
Relativní redukce rizika (RRR)

Při výpočtu relativní redukce rizika se absolutní hodnoty negativního výstupu léčby neodečítají ale dělí, tedy při použití výše uvedeného příkladu je relativní redukce rizika při použití léčby A 50% ($10\% / 20\% = 0.5 = 50\%$).

5.

Specifické designy klinických studií

CO NÁM (NE)ŘÍKAJÍ KLINICKÉ STUDIE?

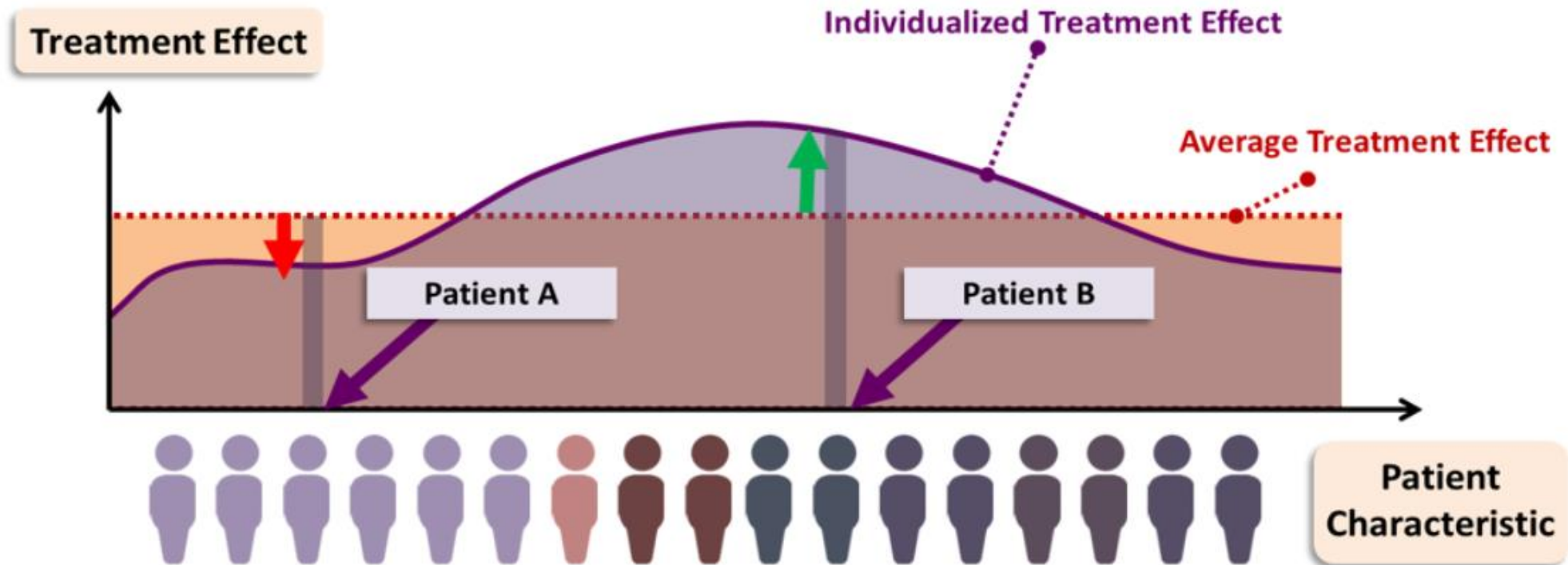


Austin Bradford Hill: „.....*RCTs do not answer the practicing doctor's question: what is the most likely outcome when this particular drug is given to a particular patient?*”

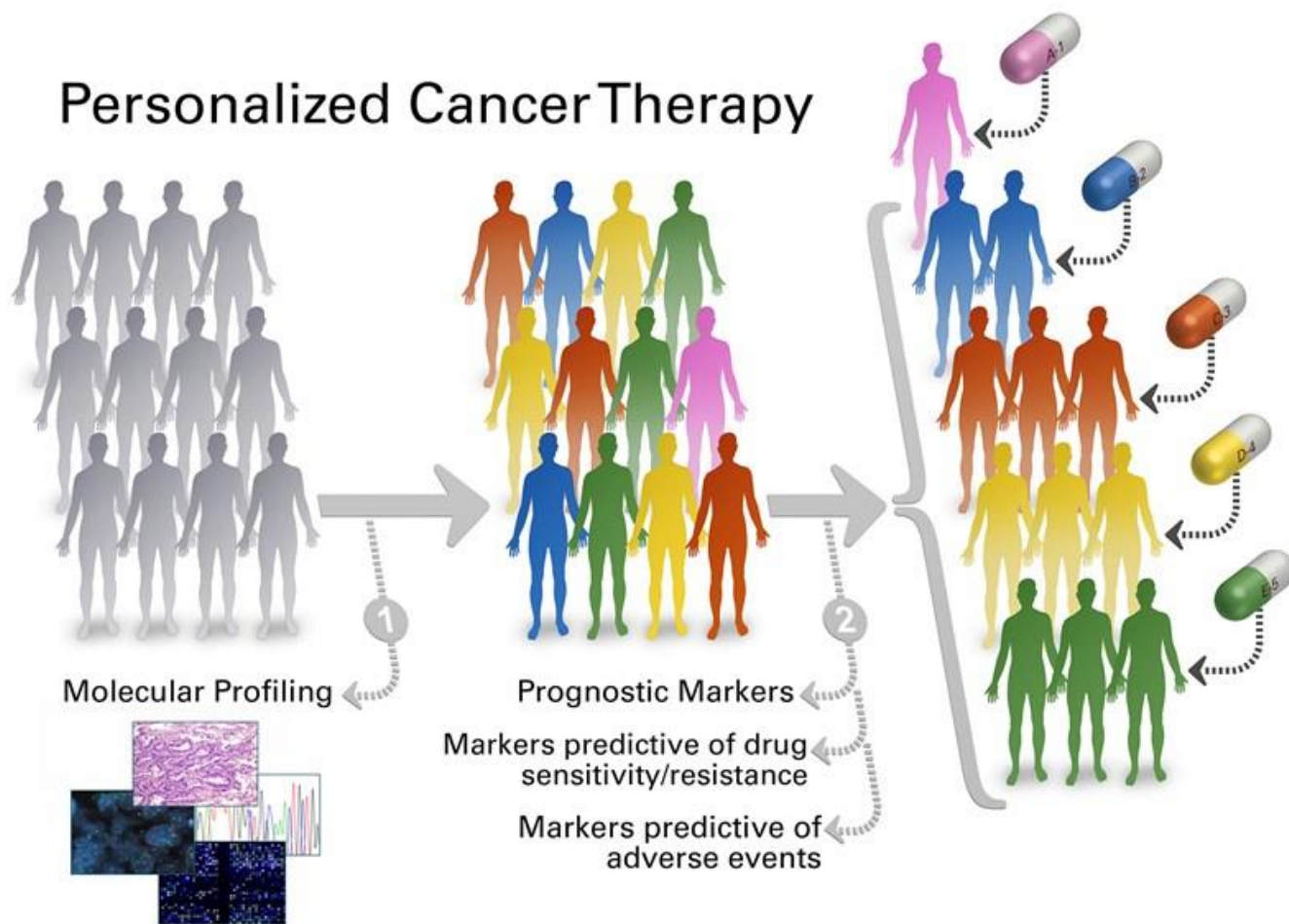


Klinická studie nám „pouze“ říká, jaká bude průměrná odpověď na léčbu u skupiny pacientů splňujících všechna vstupní a vylučující kritéria dané klinické studie.

PROBLÉM TRADIČNÍCH KLINICKÝCH STUDIÍ: Heterogeneity of Treatment Effect

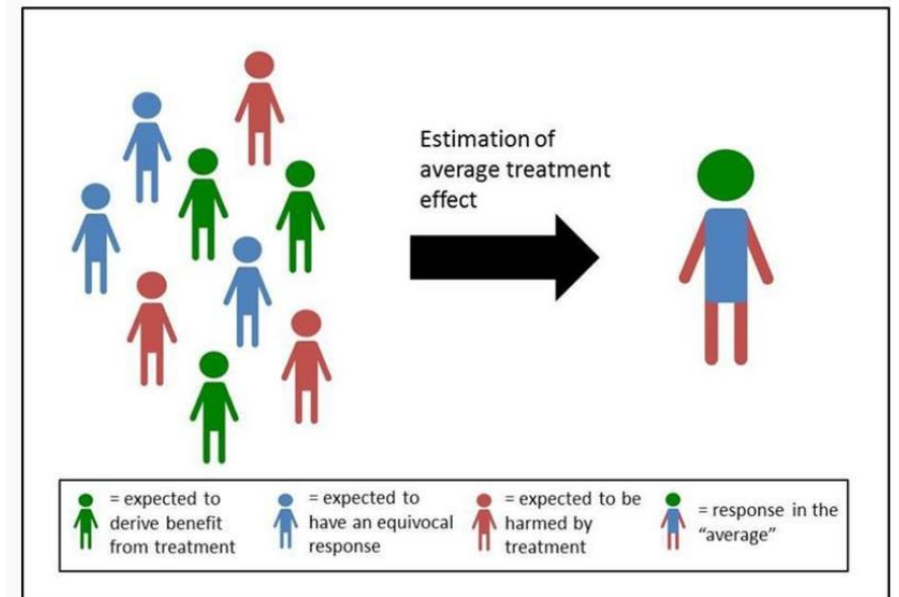


PŘÍKLAD ROLE PERSONALIZOVANÉ MEDICÍNY V REÁLNÉ KLINICKÉ PRAAXI



KLINICKÁ STUDIE X INDIVIDUÁLNÍ PACIENT

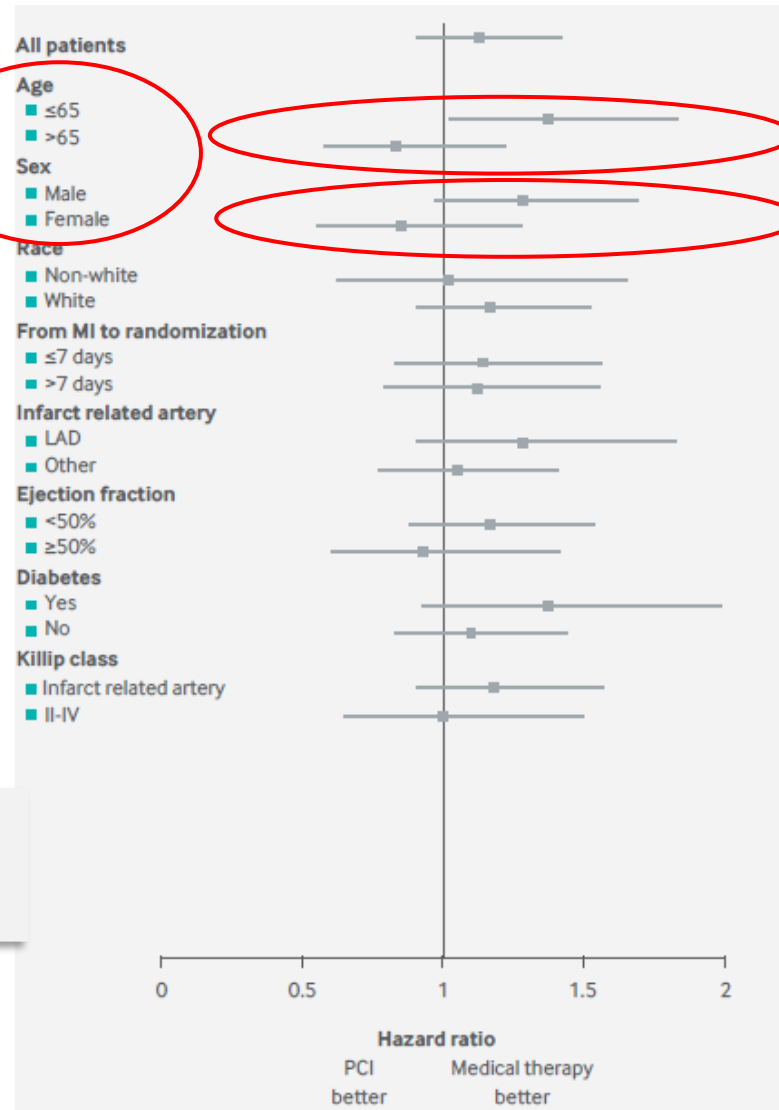
- Každý jednotlivý pacient má **nekonečné množství vlastností ovlivňujících účinnost/bezpečnost** konkrétní léčby
- Treatment guidelines jsou většinou tvořeny na základě **průměrného efektu** pro **průměrného pacienta**
- Využití klinických studií pro stanovení odpovědi konkrétního pacienta na léčbu vychází z konceptu „**reference class forecasting**“:
 - Efekt léčby konkrétního pacienta odhadujeme na základě průměrného efektu u obdobné skupiny pacientů
 - Jak moc je náš pacient **podobný** skupině pacientů zařazených do studie?



JAK (NE)ŘEŠÍ TRADIČNÍ KLINICKÉ STUDIE HETEROGENITU ODPOVĚDÍ NA LÉČBU?

I přesto, že v této konkrétní klinické studii se ukázal jasný klinicky signifikantní rozdíl účinnosti léčby dle věku a pohlaví, studie konstatovala **homogenitu léčebného efektu napříč všemi podskupinami (!)** (statistický test interakce podskupin byl nevýznamný z důvodu malého počtu pacientů)

„Forest plots“ mají za cíl zobrazit heterogenitu/homogenitu efektu léčby napříč podskupinami pacientů



Jaká je optimální léčba pro **mladé ženy** nebo **starší muže**?

Umíme odpovědět na základě výsledků této studie?



Posuzování efektu léčby v rámci podskupin na základě jednorozměrných analýz podskupin je chybné a zavádějící. Jsou dokumentovány konkrétní příklady kdy tento přístup vedl k dramaticky chybným závěrům (např. neúčinnost aspirinu v sekundární prevenci iktu u žen).

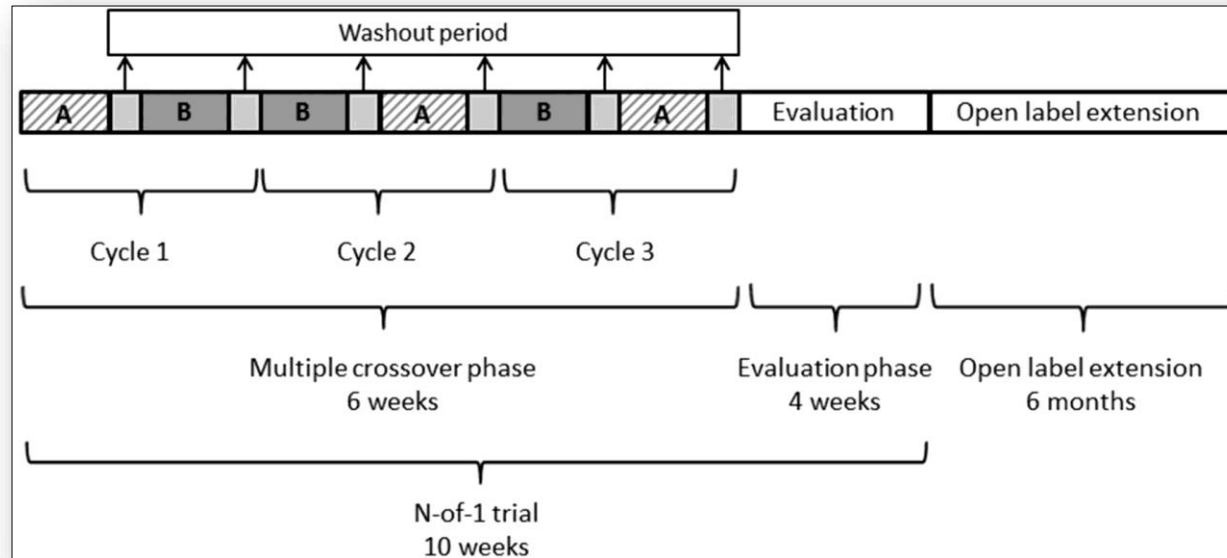
N-of-1 (single-patient) studie

- multi-crossover studie s cílem zjistit účinek léčby u individuálního pacienta a nastavit co nejefektivněji léčebný postup
 - každý léčivý přípravek podán opakovaně v několika cyklech pro potvrzení účinku
 - zobecnění závěrů pro širší populaci pacientů díky analýze série podobně designovaných N-of-1 studií pomocí pokročilých statistických metod, např:
 - metody meta-analýzy
 - lineární smíšené modely
 - Bayesovské modely
-

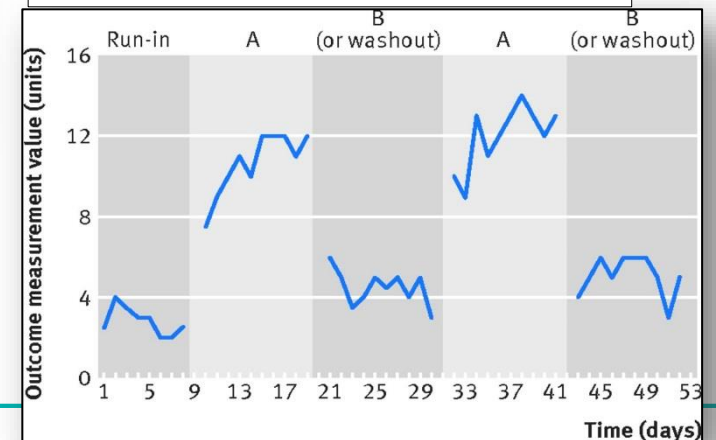
Jaký je princip a jaké jsou limitace N-of-1 studií?

Cílem studie je najít nejvhodnější léčbu pro konkrétního pacienta prostřednictvím testování různých typů léčby u jednoho pacienta odděleně v čase

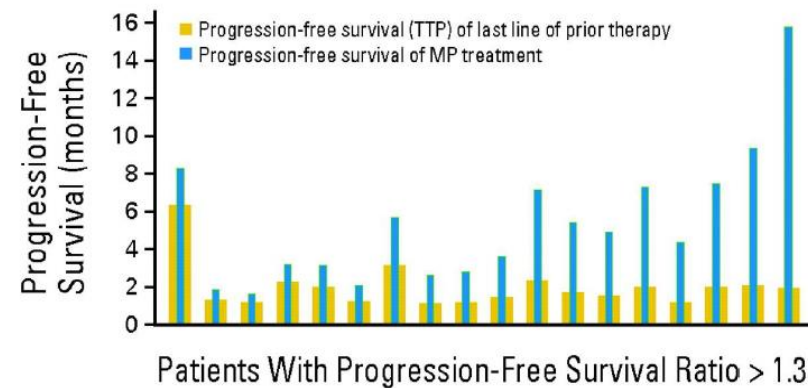
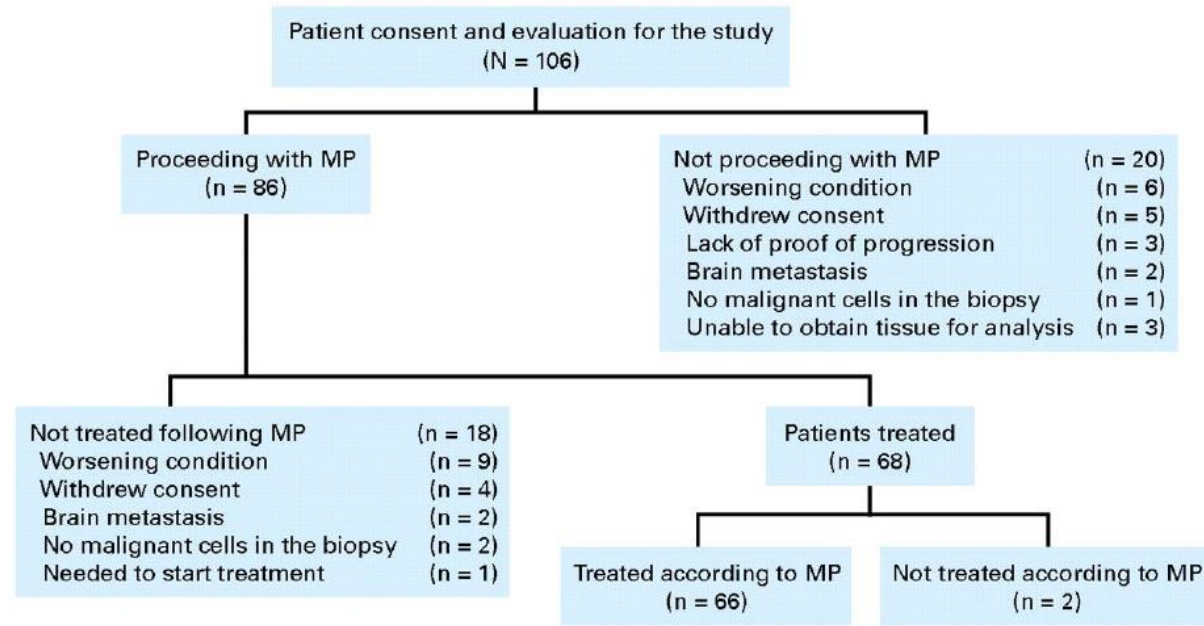
Typické uspořádání N-of-1 studie



Příklad výstupu N-of-1 studie:



N-of-1 studie: příklad



Výhody:

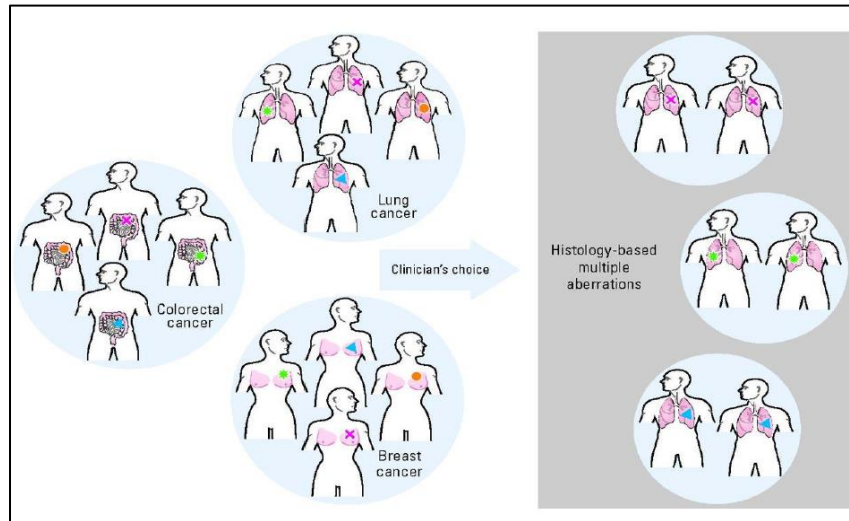
- eliminace vlivu inter-individuálních rozdílů mezi pacienty

Nevýhody:

- výsledek léčby může být ovlivněn více faktory (např. sekvencí předchozí léčby, změnou biologie onemocnění apod.)
 - Komplikovaná extrapolace a zobecnění výsledků
-

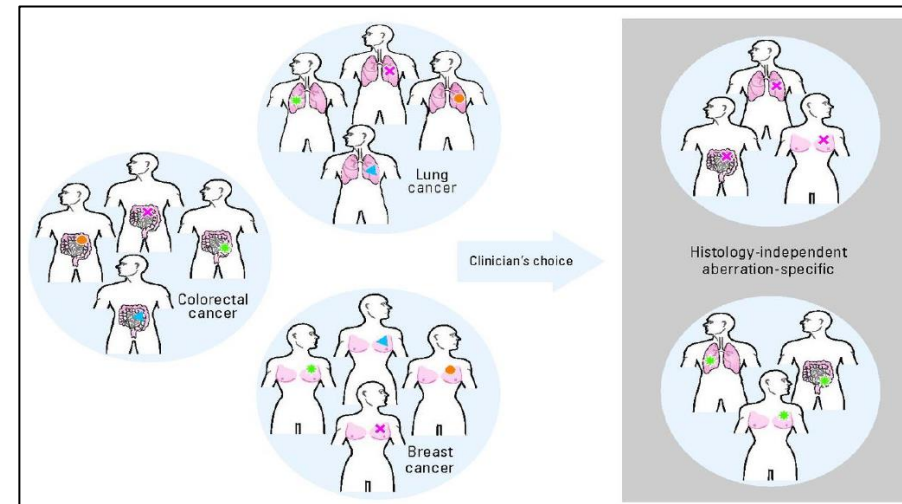
„Genotype-driven“ studie: klasifikace dle zohlednění histologie/aberrace

„histology-based“



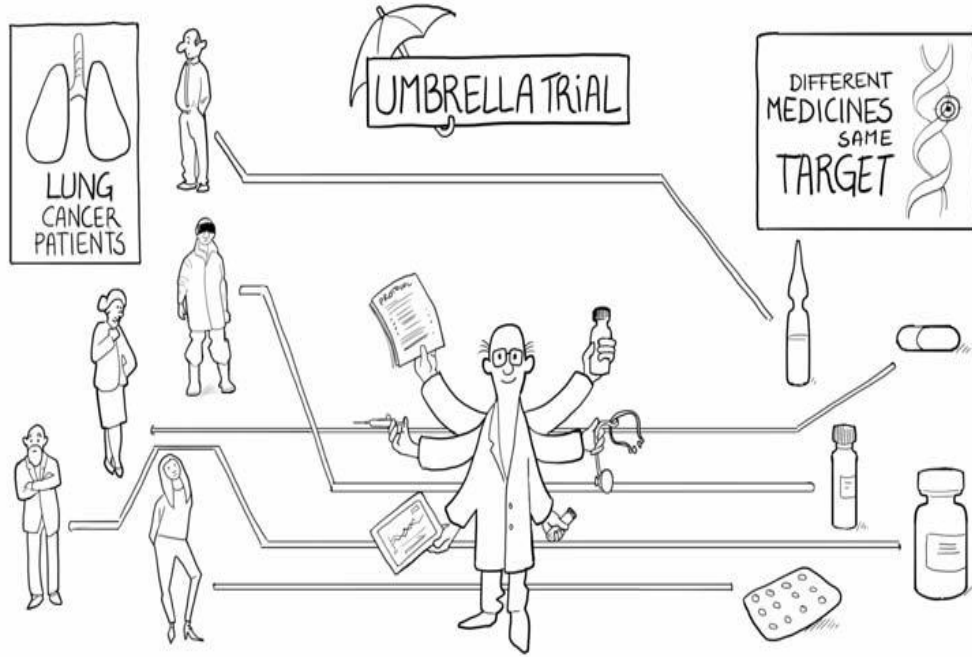
„UMBRELLA“ design

„histology-independent“,
„aberration –specific“



„BASKET“ design

„Umbrella“ design



- stejný histologický typ onemocnění
- paralelně testovány různé přípravky u podskupin pacientů s různými typy mutací
- hodnotí se specifická účinnost u jednotlivých molekulárních aberací

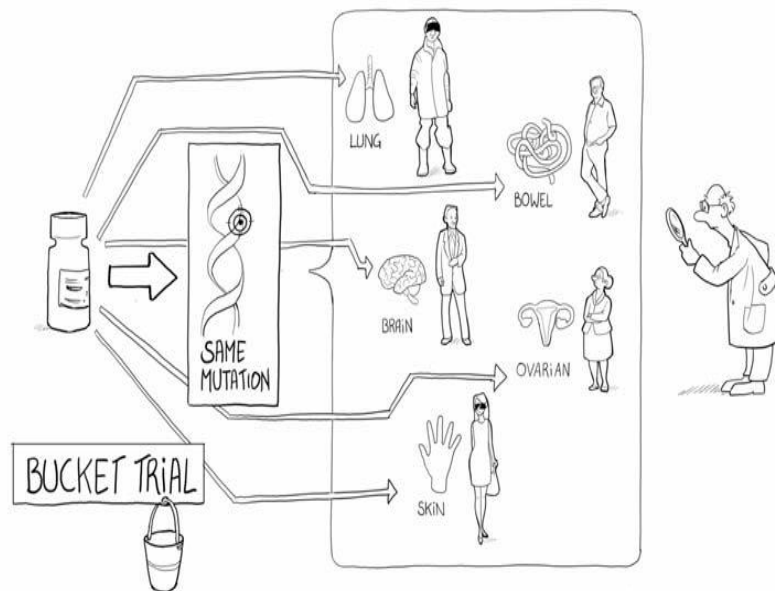
Výhody:

- snížený počet „screening failures“
- větší počet pacientů profitujících z léčby
- možnost posouzení více léčebných modalit v rámci jedné studie

Nevýhody:

- větší počet ramen studie (statistické konsekvence)
 - celkově vyšší počet zařazených pacientů
 - logisticky náročnější průběh studie
-

„Basket“ design



- odlišné histologické typy onemocnění
- stejný typ mutace / mechanismu účinku

„Basket“ design: Výhody a nevýhody

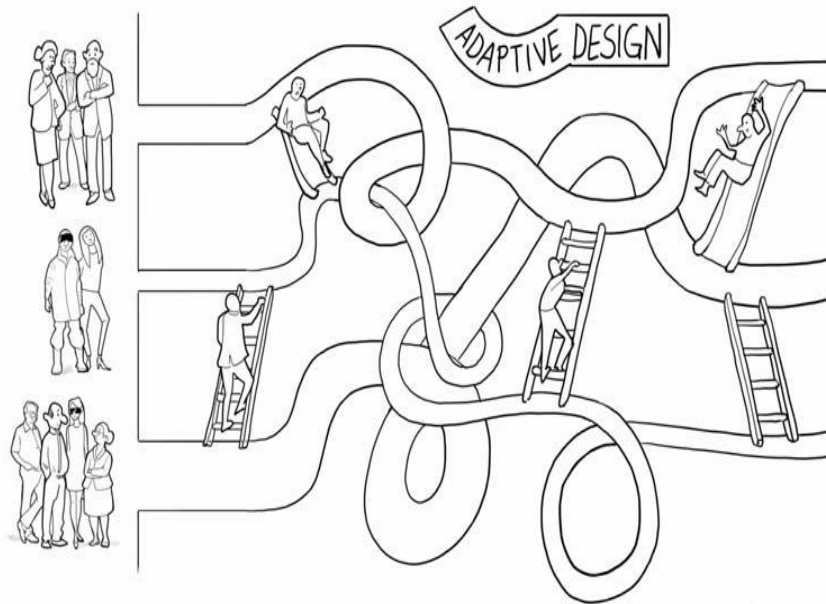
Výhody:

- možnost nalezení specifické léčby pro nádory odlišné histologie

Nevýhody:

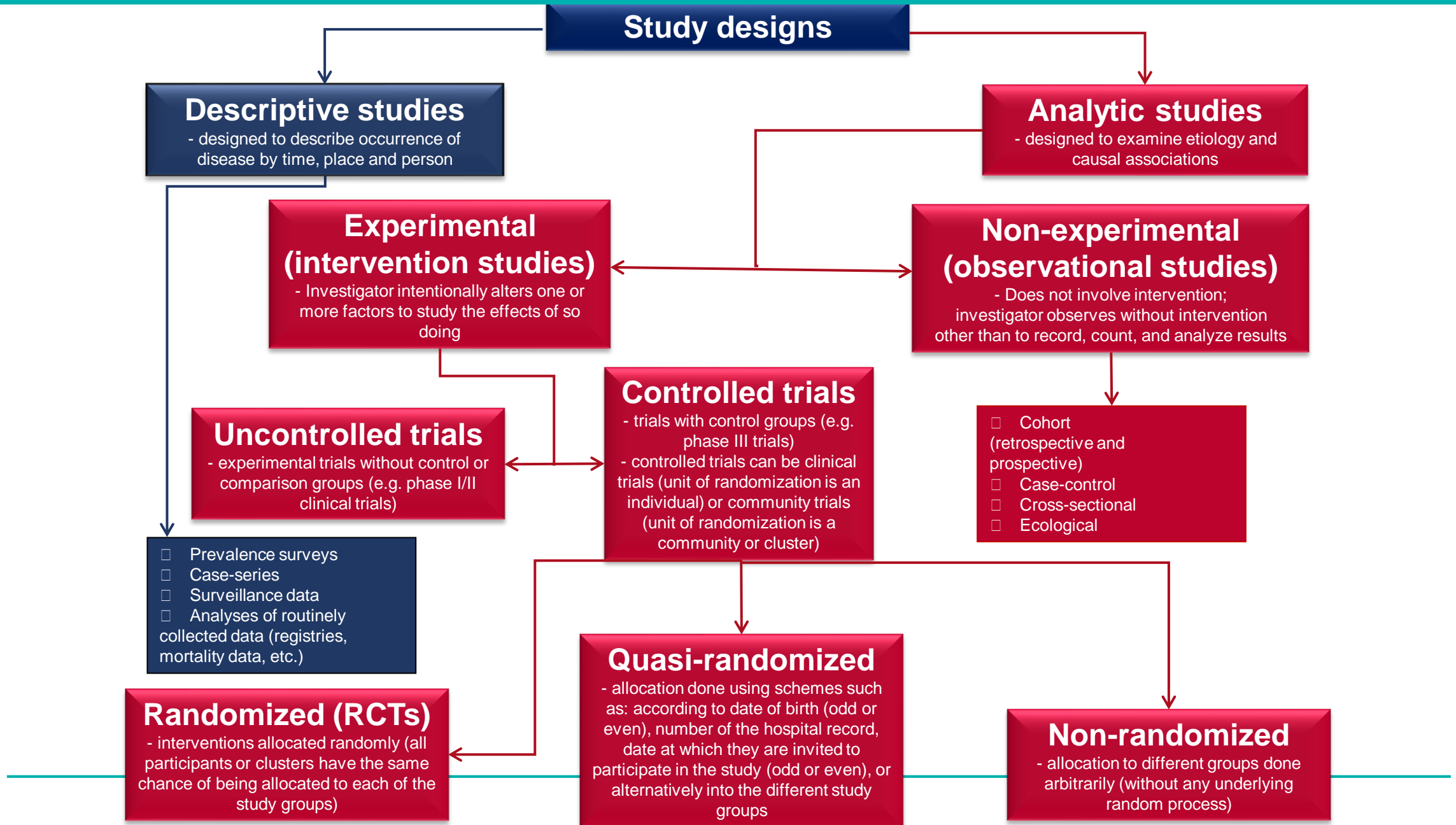
- možnost podcenění odlišné odpovědi u nádorů různých histologií (např. BRAF u melanomu a kolorekt)
-

Adaptivní design

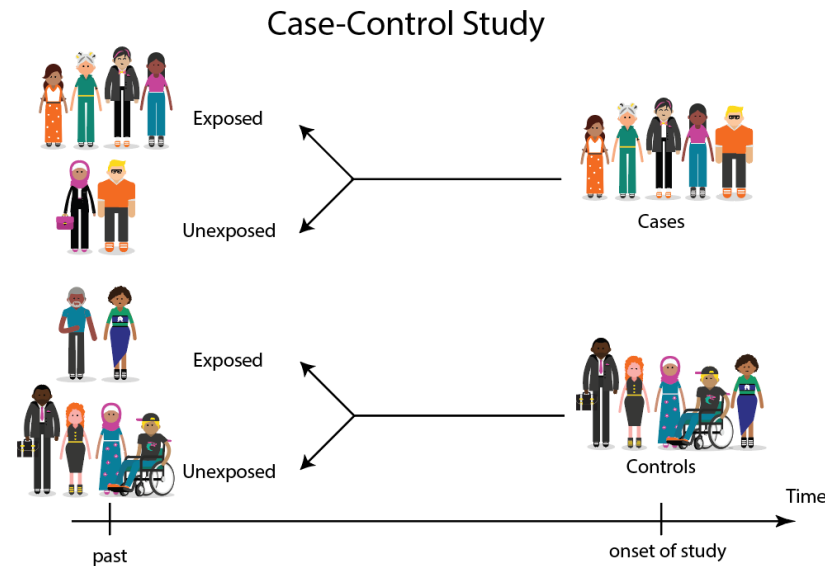


- Studie s adaptivním designem využívají nakumulované informace k modifikaci dalšího průběhu studie: **„We learn as we go“**
- FDA definice:
*.....an adaptive design clinical study is defined as a study that includes a **prospectively planned opportunity for modification of one or more specified aspects of the study design** and hypotheses based on analysis of data (usually interim data) from subjects in the study.*
- Hlavní prvky adaptivního designu:
 - + Interim analýza
 - + adaptivní randomizace

Klasifikace studií: design



Case-control studies (studie případů a kontrol)



Základní charakteristika:

V současnosti identifikujeme skupinu s výskytem sledovaného jevu (případy) a bez výskytu jevu (kontroly). Tyto skupiny „párujeme“ dle předem zvolených parametrů (např. věk, pohlaví apod.) a retrospektivně analyzujeme rozdíl ve výskytu faktorů potenciálně ovlivňujících výskyt sledovaného jevu.

Výhody:

- Relativně levné a rychlé
- Pomáhají identifikovat potenciální zdroje aktuálního problému



Nevýhody:

- Ne vždy dohledatelné údaje z historie
- Absence randomizace
- Nesnadné dohledání kontrol



Case-control studies (studie případů a kontrol): příklad

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

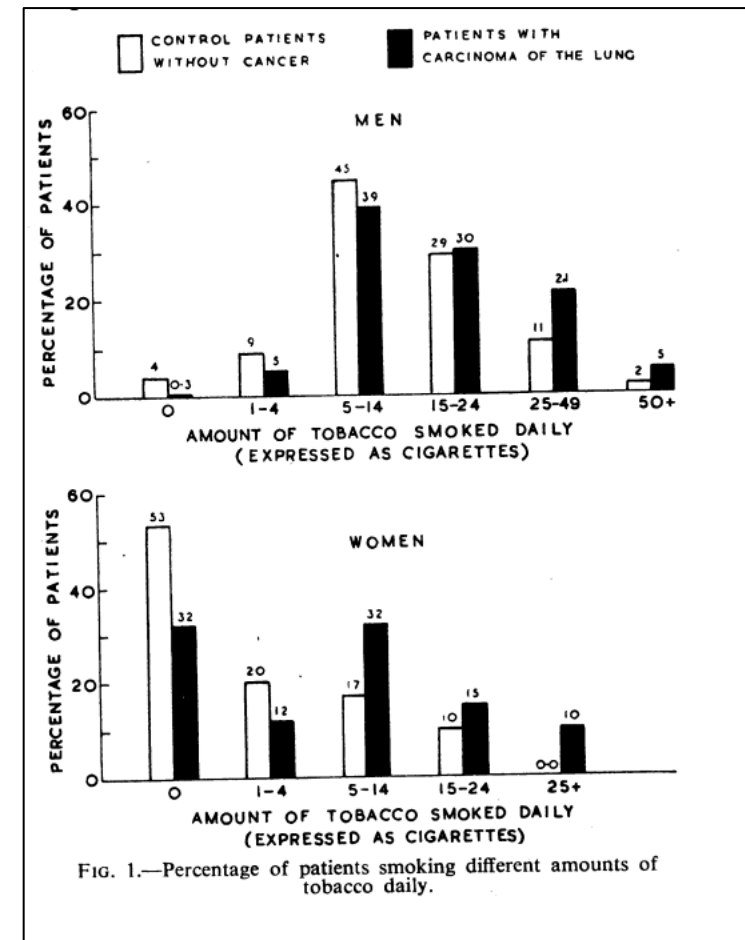
Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

Smokers and Non-smokers

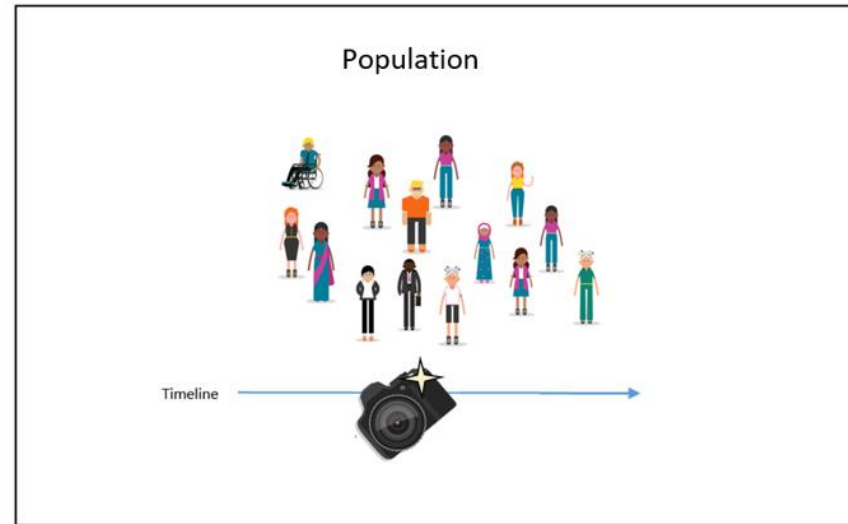
The simplest comparison that can be made to show whether there is any association at all between smoking and carcinoma of the lung is that between the proportion of lung-carcinoma patients who have been smokers and the proportion of smokers in the comparable group of subjects without carcinoma of the lung. Such a comparison is shown in Table IV.

TABLE IV.—*Proportion of Smokers and Non-smokers in Lung-carcinoma Patients and in Control Patients with Diseases Other Than Cancer*

Disease Group	No. of Non-smokers	No. of Smokers	Probability Test
Males:			
Lung-carcinoma patients (649)	2 (0.3%)	647	P (exact method) = 0.0000064
Control patients with diseases other than cancer (649)	27 (4.2%)	622	
Females:			
Lung-carcinoma patients (60)	19 (31.7%)	41	$\chi^2 = 5.76; n = 1$ 0.01 < P < 0.02
Control patients with diseases other than cancer (60)	32 (53.3%)	28	



Cross-sectional studies (průřezové studie)



Základní charakteristika:

Jednorázový „snímek“ vybraných charakteristik populace.

Výhody:

- Relativně levné a rychlé
- Možnost sběru dat velkého množství proměnných najednou



Neúhody:

- Identifikace možných asociací ne vždy kauzalit
- Nevhodné pro vzácné jevy (např. rare diseases)



Cross-sectional studies (průřezové studie): příklad

✓ Latest release

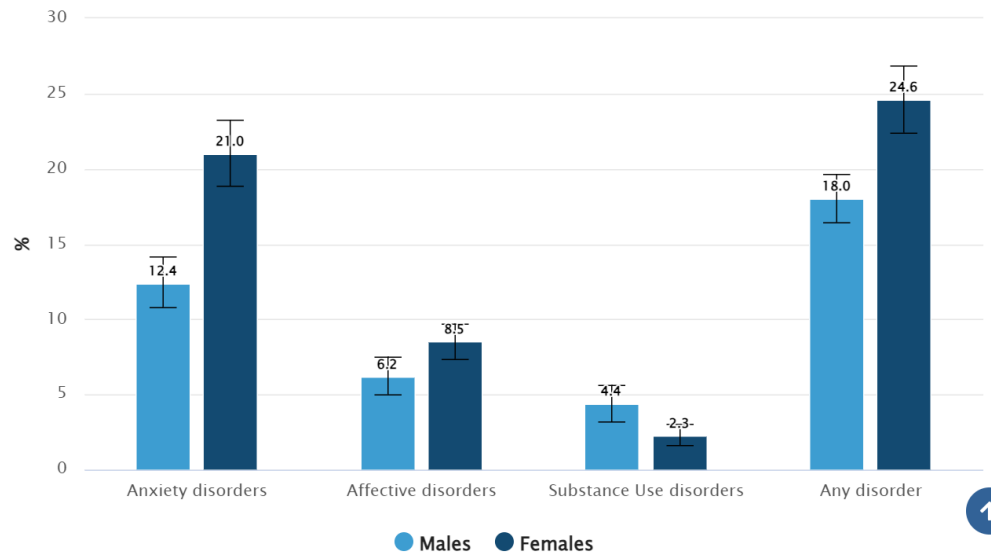
↓ Data download

National Study of Mental Health and Wellbeing

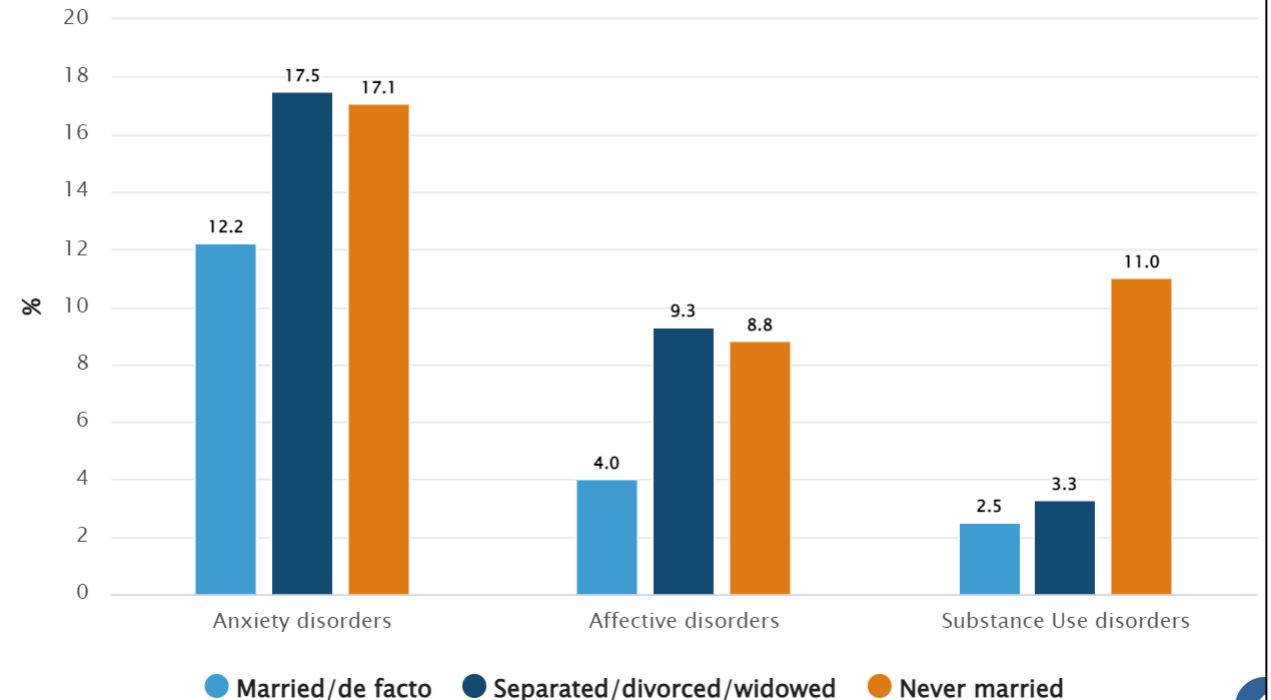
Summary statistics on key mental health issues including the prevalence of mental disorders and the use of services

Reference period 2020-21

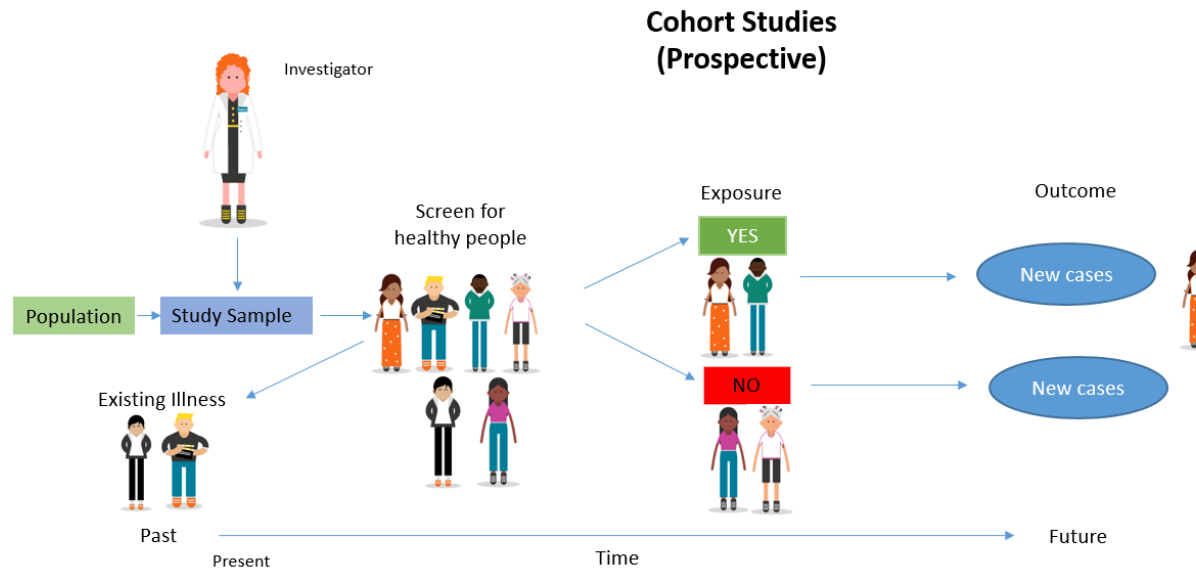
12-month mental disorders, by disorder group and sex, 2020-21



8. 12-month mental disorders(a), by marital status



Cohort studies (kohortové studie): prospektivní



Základní charakteristika:

Prospektivní sledování vybraných kohort, kde u části subjektů v čase nastane sledovaný jev.

Výhody:

- Jediný design sledující přímý vliv faktorů na sledovaný jev
- Vhodné i pro sledování expozice vzácným faktorům

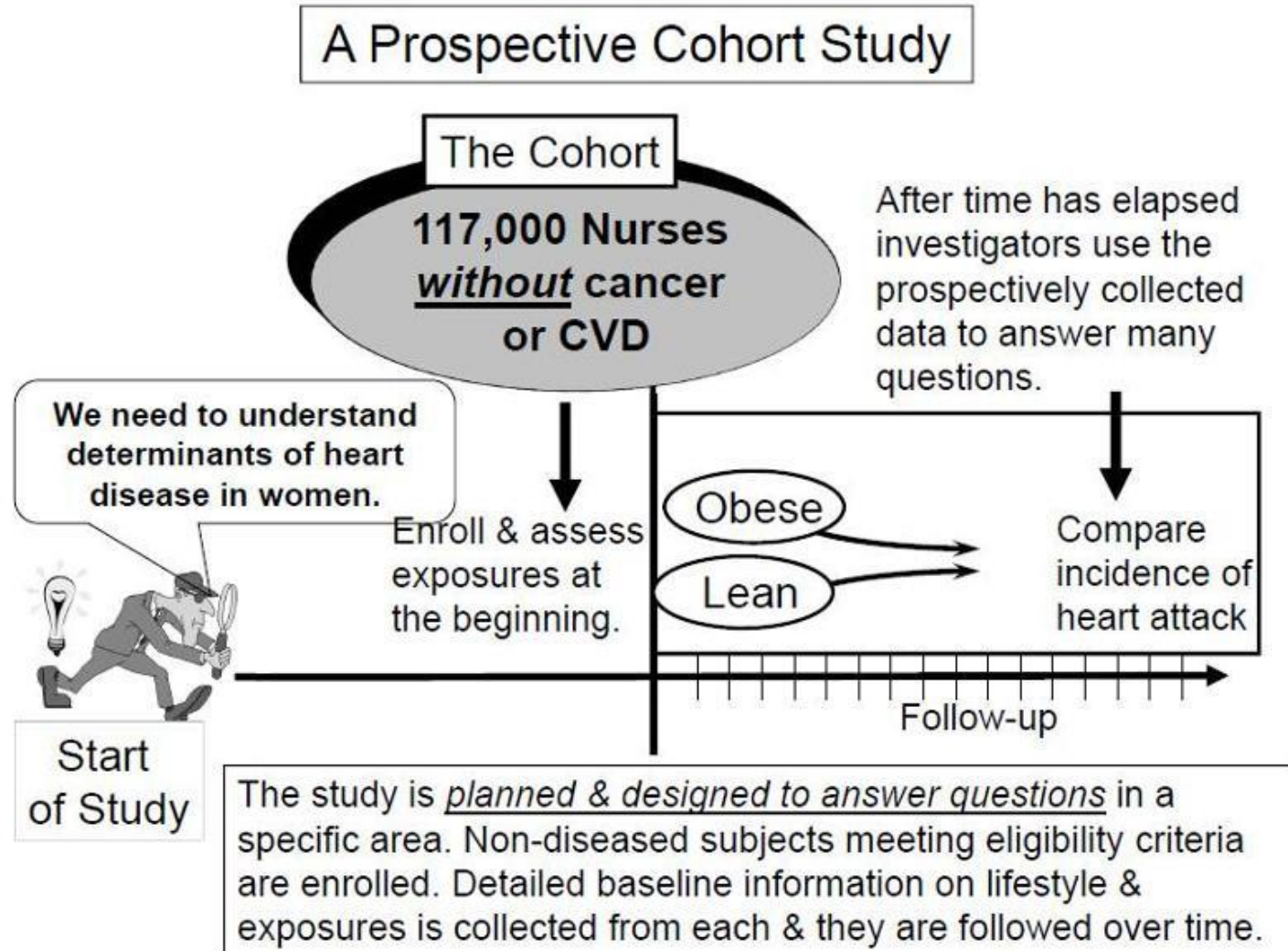


Nevýhody:

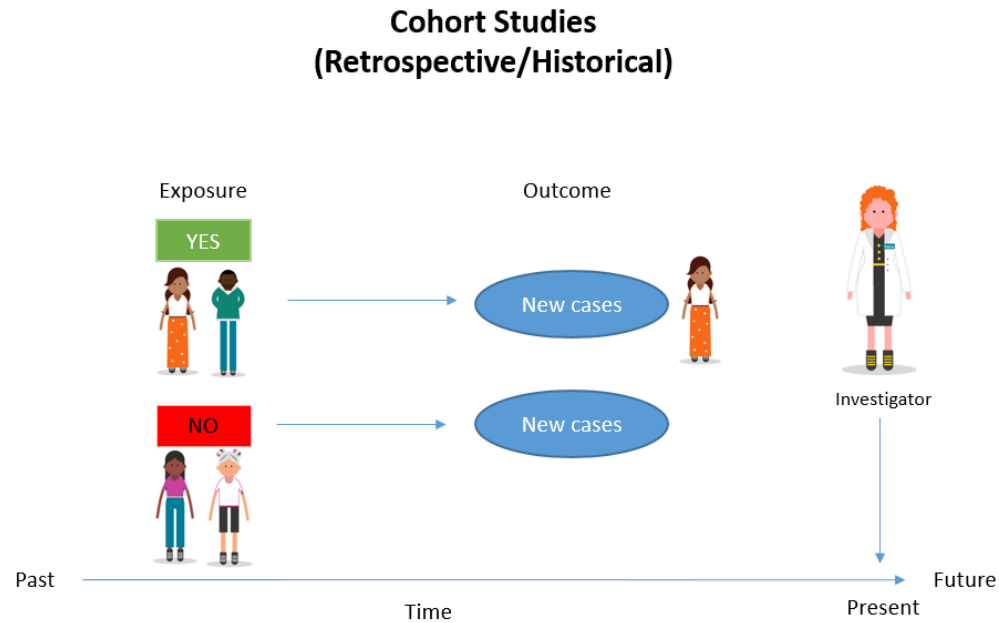
- Drahé a časově náročné
- Nevhodné pro sledování dlouho se vyvíjejících jevů



Cohort studies (kohortové studie): prospektivní: příklad



Cohort studies (kohortové studie): retrospektivní



Základní charakteristika:

Kohorty jsou definované retrospektivně a nejsou prospektivně sledovány.

Výhody:

- Jediný design sledující přímý vliv faktorů na sledovaný jev
- Vhodné i pro sledování expozice vzácným faktorům

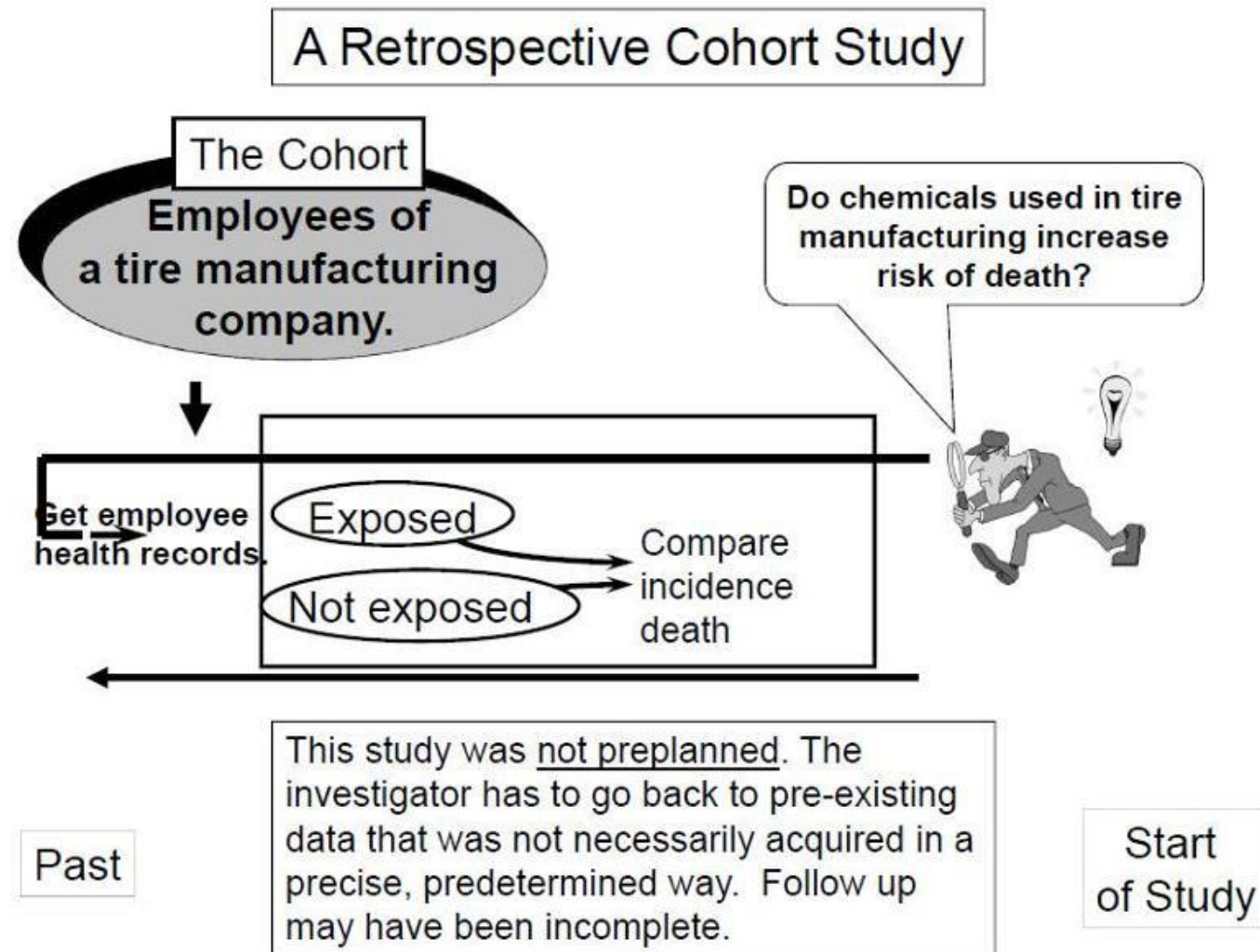


Nevýhody:

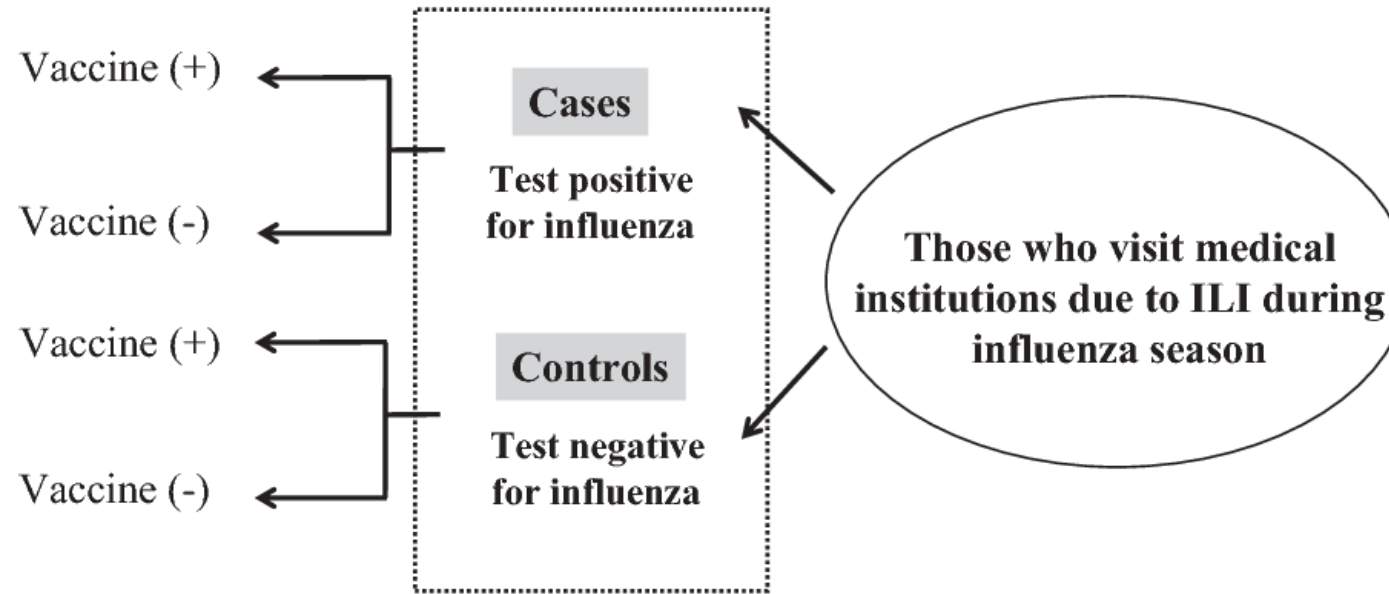
- Drahé a časově náročné
- Nevhodné pro sledování dlouho se vyvíjejících jevů



Cohort studies (kohortové studie): retrospektivní



Test-negative design



Základní charakteristika:

Případy a kontroly jsou do studie zařazeny na základě stejného klinického stavu a rozděleny podle provedení laboratorního testu.

Výhody:

- Případy a kontroly srovnatelnější než u klasické case-control studie
- Minimalizace falešně pozitivních případů



Nevýhody:

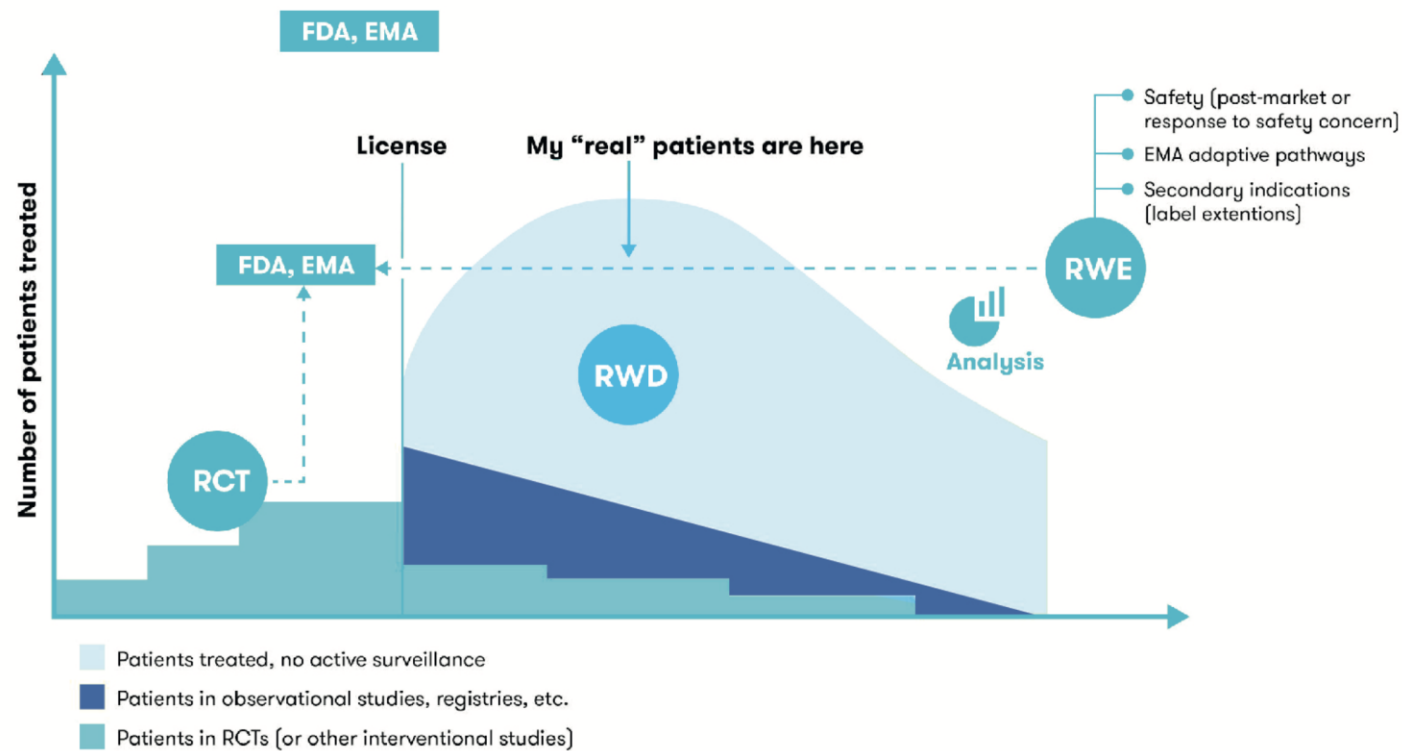
- Vyžadují velké soubory (poměr cases/test negative)
- Drahé a časově náročné



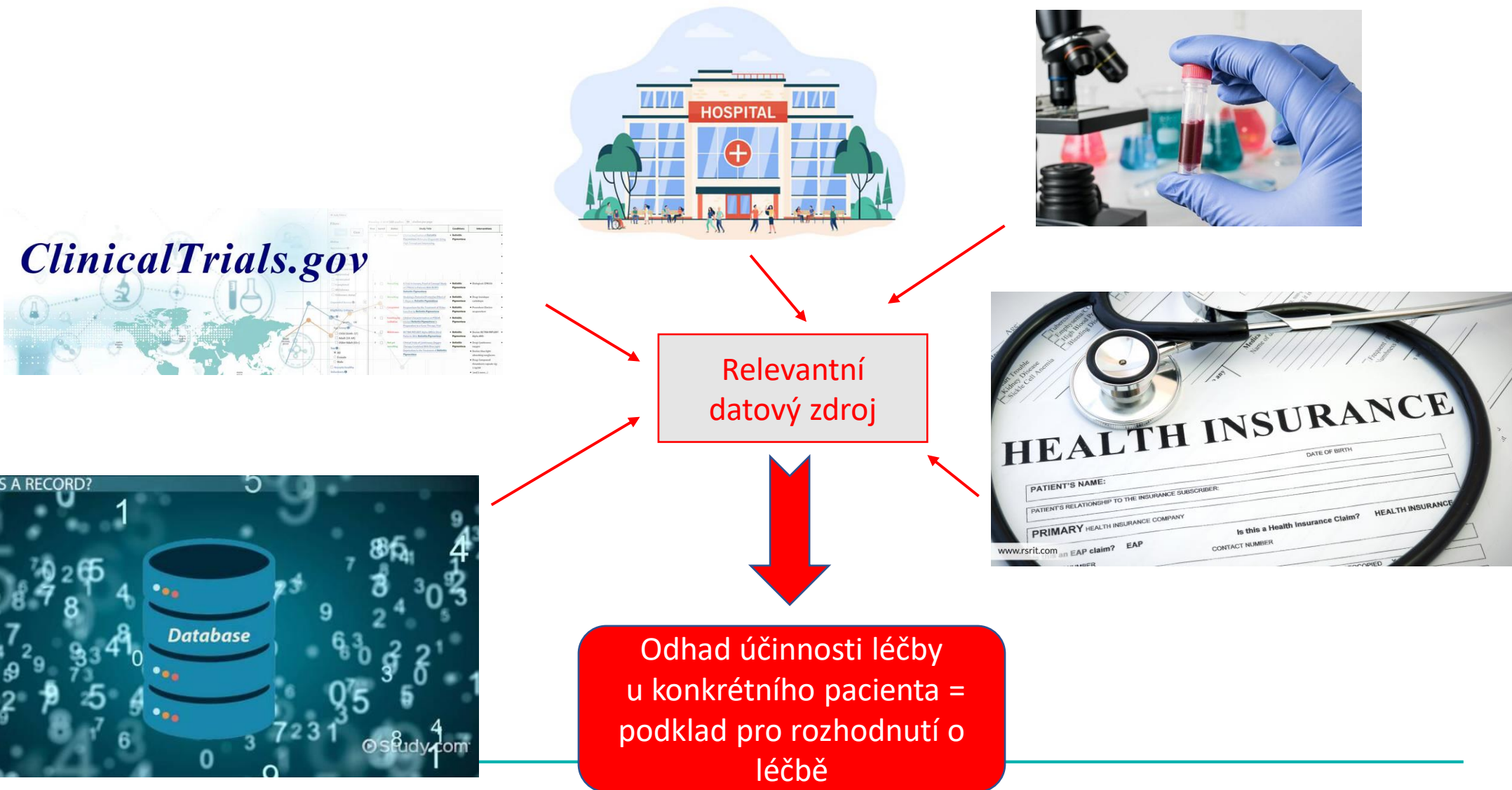
Pro zajištění co nejúčinnější
a nejbezpečnější léčby pacientů
nestačí jen **vyvíjet a registrovat nové přípravky.**

Je nutné současně **inteligentně využívat
existující data** z klinické praxe.

RWD: EXISTUJÍCÍ PRAVDIVÁ DATA O ÚČINNOSTI A BEZPEČNOSTI LÉČBY V KLINICKÉ PRAXI



RWD: OBROVSKÝ „SPÍCÍ“ POTENCIÁL INFORMAČNÍHO ZDROJE



2018



RWD a RA: Změna paradigmatu

FRAMEWORK FOR FDA'S REAL-WORLD EVIDENCE PROGRAM

Contains Nonbinding Recommendations

2019



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

FDA 2021

IN THIS SECTION

← News & Events for Human Drugs

FDA approves new drug based on real-world evidence



Today, the U.S. Food and Drug Administration (FDA) has approved a new use for Prograf (tacrolimus) based on real-world evidence (RWE) of effectiveness in combination with other immunosuppressants to prevent organ rejection in adult and pediatric patients receiving lung transplantation.



The European Medicines Agency (EMA) is establishing a coordination centre to provide timely and reliable evidence on the use, safety and effectiveness of medicines for human use, including vaccines, from real world healthcare databases across the European Union (EU).

This capability is called the **Data Analysis and Real World Interrogation Network (DARWIN EU®)**.

2021

comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document or the Real-World Evidence Program, please email CDERMedicalPolicy-RealWorldEvidence@fda.hhs.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

October 2021
Real-World Data/Real-World Evidence (RWD/RWE)



delivering rapid access to and analysis of representative, longitudinal RWD throughout a product's lifecycle



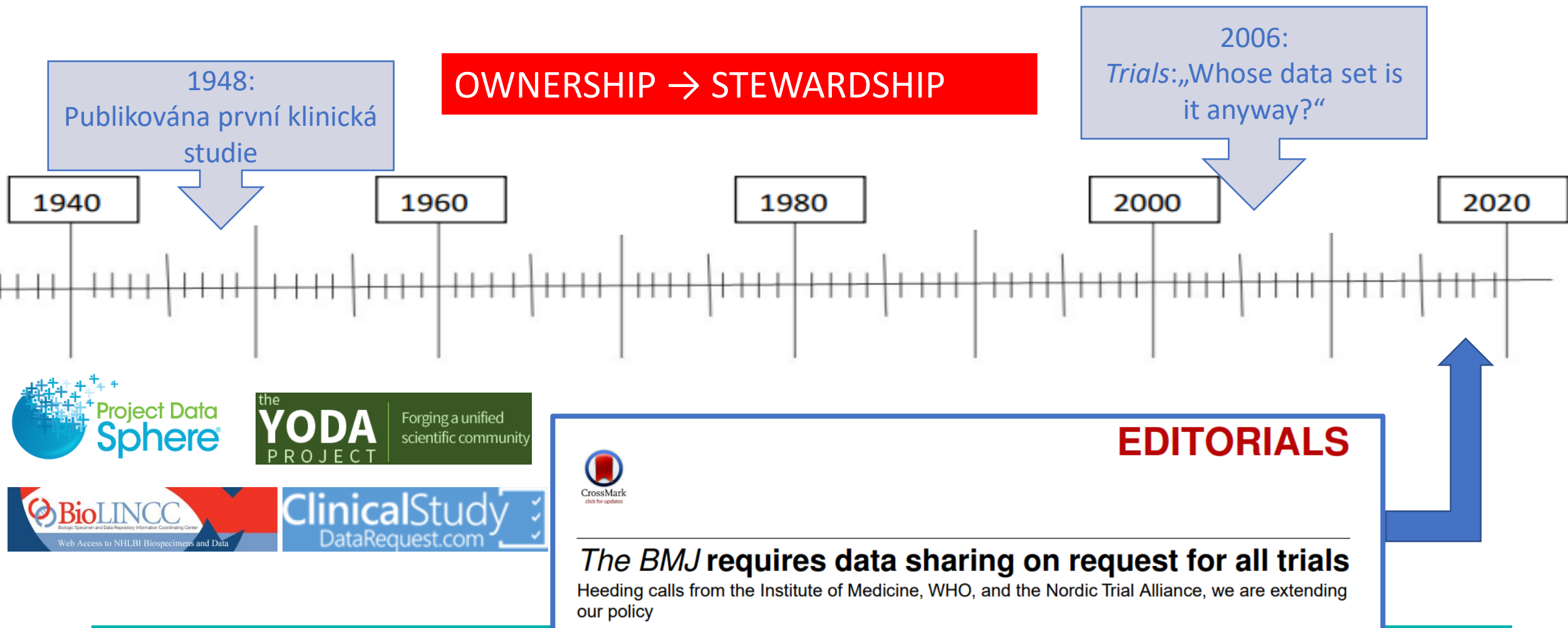
Develop a capacity that will enable the Agency to rapidly and securely access and analyse large amounts of healthcare data



Accelerate the implementation of a learning regulatory system based on electronic health records and other routinely collected RWD



Sdílení dat z klinických studií: *Data transparency revolution*



Datový standard pro sdílení dat z klinických studií

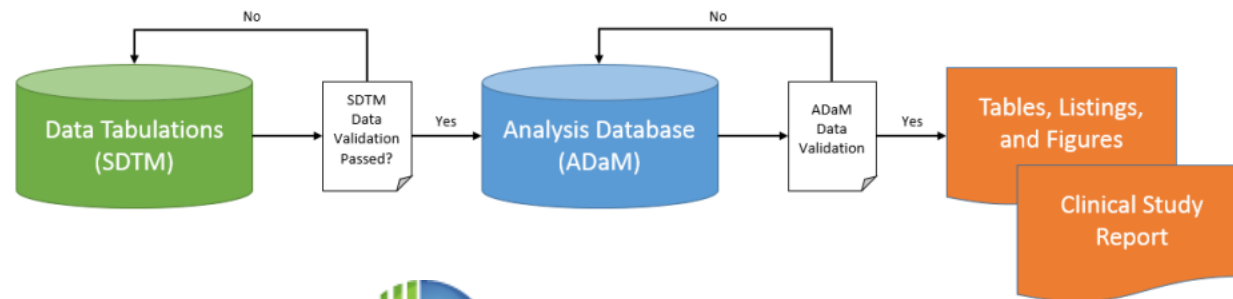


Uppsala
Monitoring
Centre

WHODrug



MedDRA



- Do roku 2000 bez standardizace
- 1997 – založení CDISC

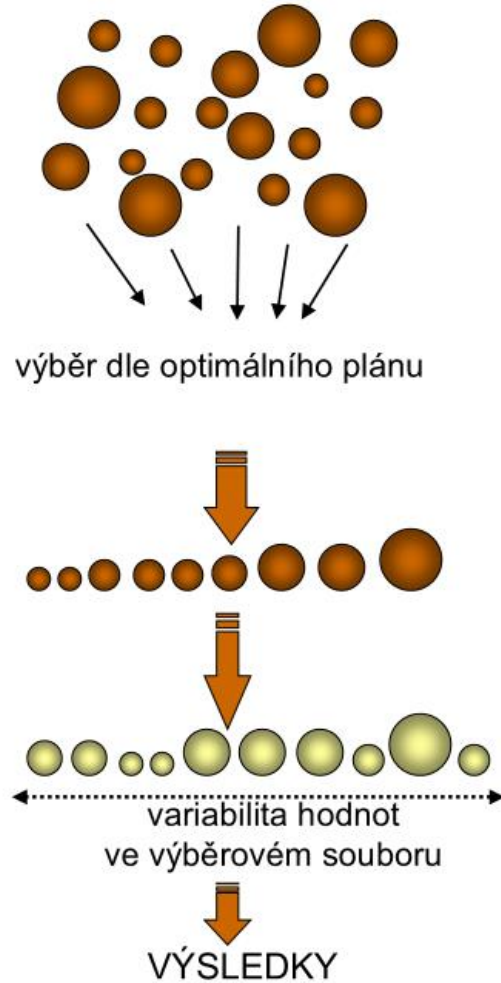
- 2016 – FDA vyžaduje SDTM
- 2019 – FDA vyžaduje ADaM
- 2020 – PMDA vyžaduje SDTM & ADaM

6.

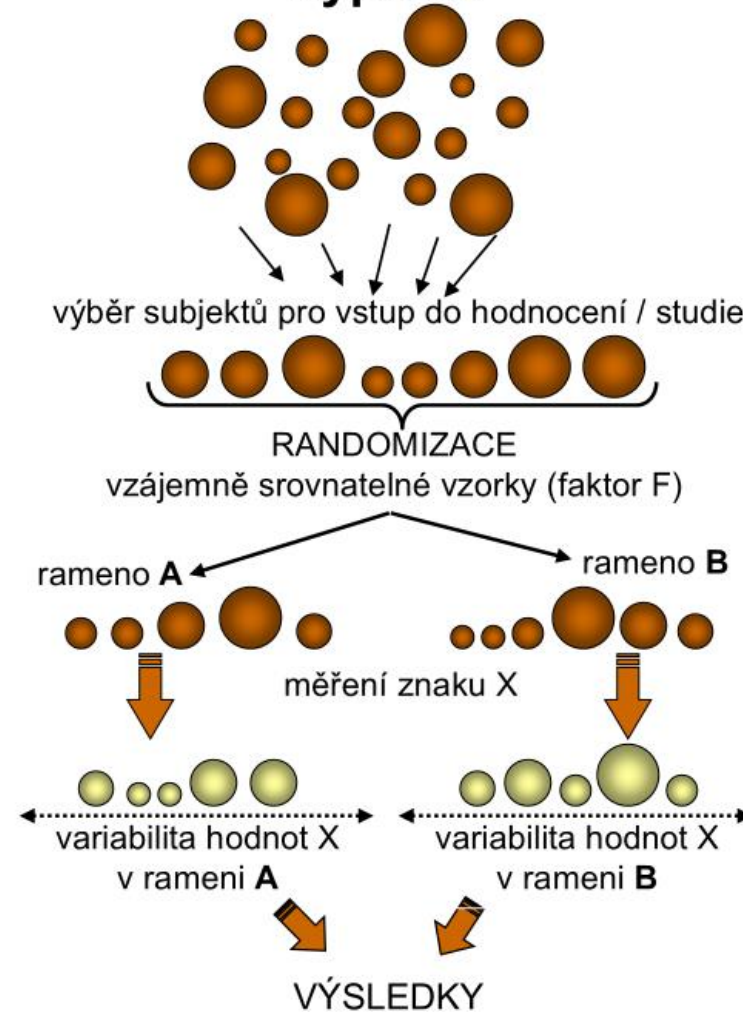
Základy deskriptivní statistiky

Typy statistických analýz v klinických studiích

Deskriptivní statistika a intervalové odhady znaků



Srovnání více skupin a testování hypotéz



Základní typy dat a jejich využití jako endpointů

Data poměrová



Kolikrát ?

Data intervalová



O kolik ?

Data ordinální



Větší, menší ?

Data nominální

Rovná se ?

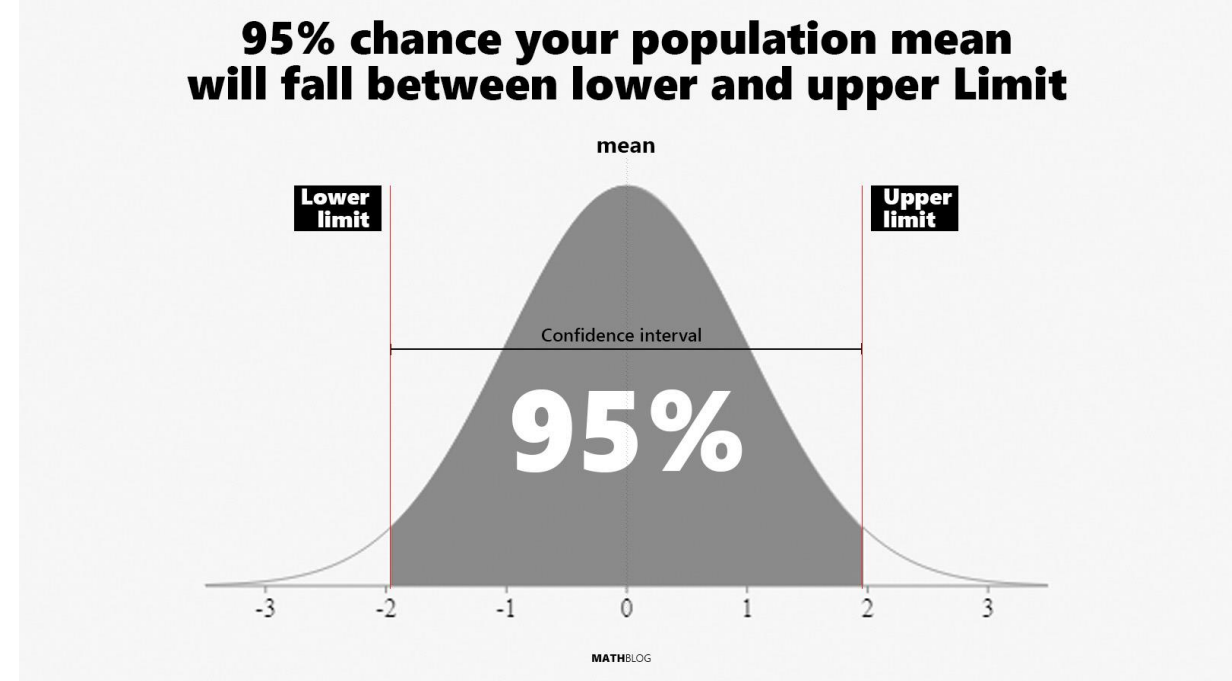
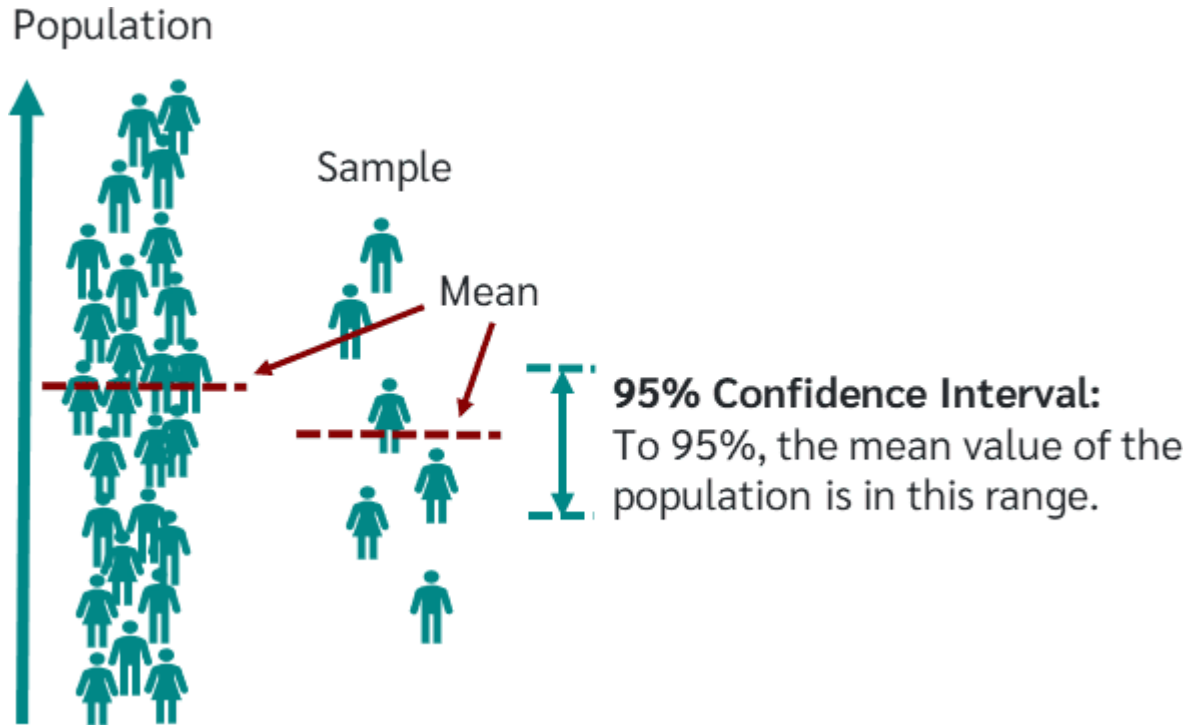
**Spojitá
data**

Kategoriální otázky

**Diskrétní
data**

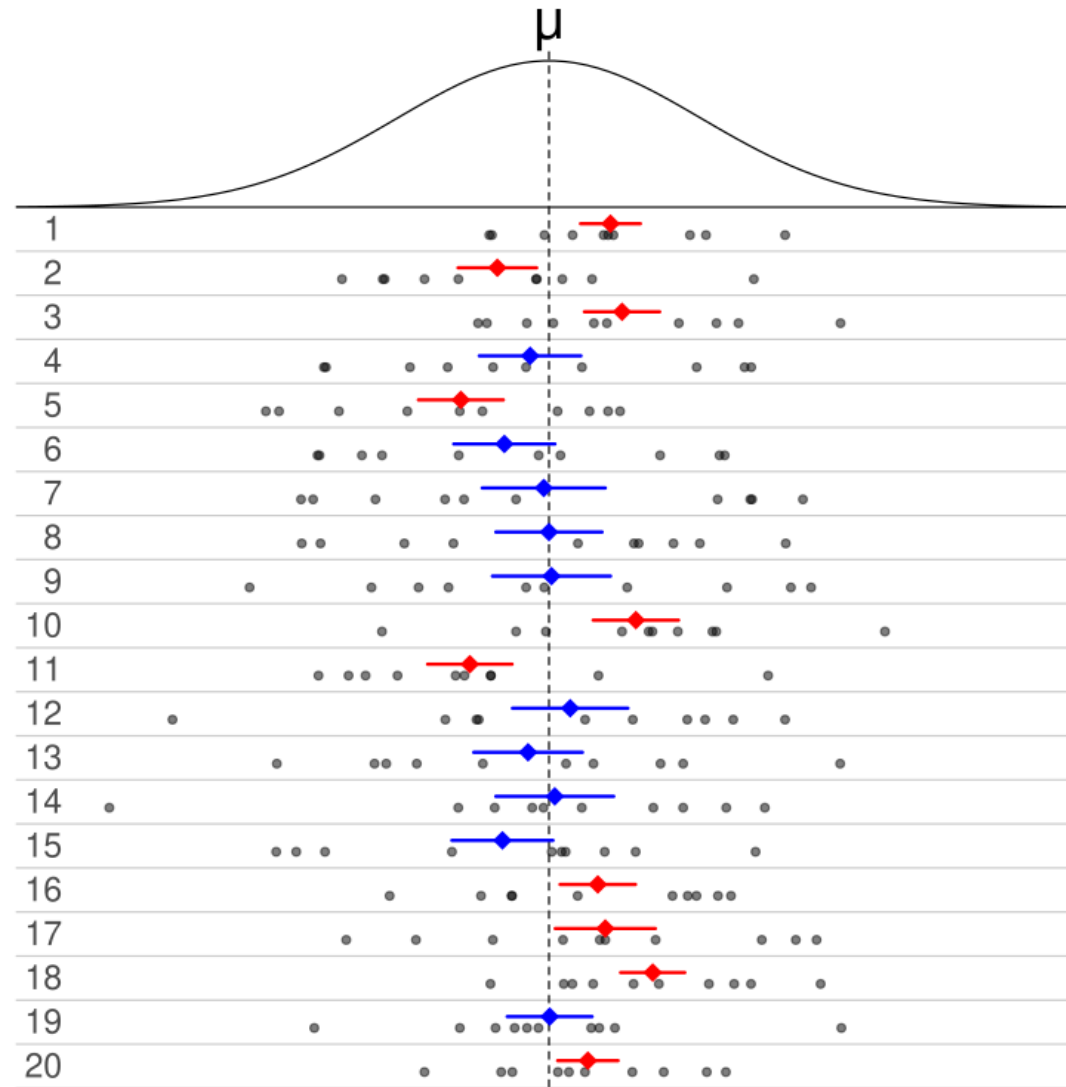
Otázky „Ano/Ne“

Confidence interval



Confidence intervals provide a range of plausible values for an unknown population parameter. They are constructed based on sample data and provide an estimate of the parameter along with a level of confidence. For example, a 95% confidence interval indicates that if the sampling process is repeated many times, 95% of the intervals constructed would contain the true population parameter. Confidence intervals are useful for quantifying the uncertainty associated with an estimate and providing a range of values within which the true parameter is likely to fall.

Confidence interval



Hypothesis testing

	<i>Experiment Does Not Reject H_0</i>	<i>Experiment Rejects H_0</i>
<i>H_0 is Actually True</i>	Correct Outcome	Incorrect Outcome Type I Error (α)
<i>H_0 is Actually False</i>	Incorrect Outcome Type II Error (β)	Correct Outcome Power ($1 - \beta$)

Hypothesis testing is a statistical procedure used to make decisions or draw conclusions about a population based on sample data. The process involves formulating a null hypothesis (H_0) and an alternative hypothesis (H_a). The null hypothesis represents the status quo or no effect, while the alternative hypothesis suggests a specific effect or difference. The steps of hypothesis testing typically include choosing a significance level (α), calculating a test statistic, and comparing it to a critical value or p-value to determine the statistical significance of the results.

Aplikovaná analýza přežití

Analýza přežití je v současnosti nejvíce používána ve dvou oborech:

„Lifetime data“ v lékařském výzkumu

- „Overall Survival (OS)“
- „Time to Progression (TTP)“
- „Time to Treatment Failure (TTF)“
- „Duration of Response“
- „Relapse Free Survival“
- jiné

„Reliability studies“ v průmyslu

- zátěžové zkoušky součástí
-

- Deskriptivní popis a modelování přežití pacientů v jedné skupině, odhad charakteristik křivky přežití (medián přežití apod.)**
 - Srovnání přežití ve dvou nebo více skupinách pacientů např. s odlišnou léčbou**
 - Vícerozměrná analýza vlivu prognostických faktorů na přežití**
-

Základní vlastností dat typu přežití je skutečnost, že určitá pozorování mohou být „censorována“, což znamená, že sledovaný jev (např. úmrtí, relaps apod.) u pacienta ve sledovaném období nenastal.

Tato „censorovaná“ pozorování však není možné ignorovat, protože v sobě nesou zásadní informaci o účinnosti hodnocené léčby.

Praktický příklad dat přežití

*Data pacientů s angina pectoris ve studii s 15letým follow-up
(Mayo Clinic, Gehan 1969)*

Survival time [years]	Number of patients known to survive at beginning of interval	Number of patients lost to follow up
0-1	2418	0
1-2	1962	39
2-3	1697	22
3-4	1523	23
4-5	1329	24
5-6	1170	107
6-7	938	133
7-8	722	102
8-9	546	68
9-10	427	64
10-11	321	45
11-12	233	53
12-13	146	33
13-14	95	27
14-15	59	23
15-16	30	

Modelová klinická studie:

Analyzovaný

počet pacientů:.....4 pacienti

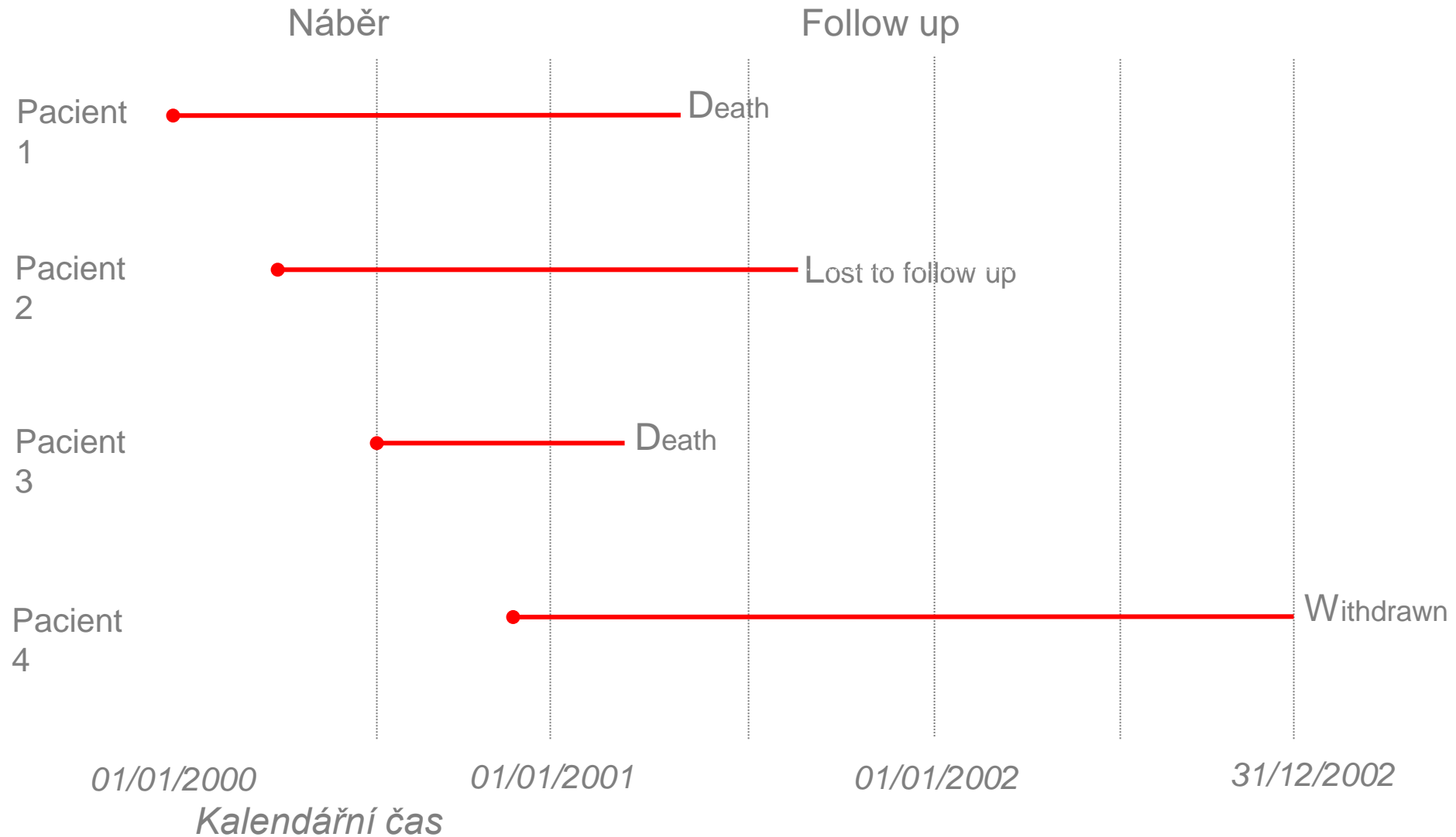
Primární endpoint:..... Overall Survival
(OS)

Období náběru

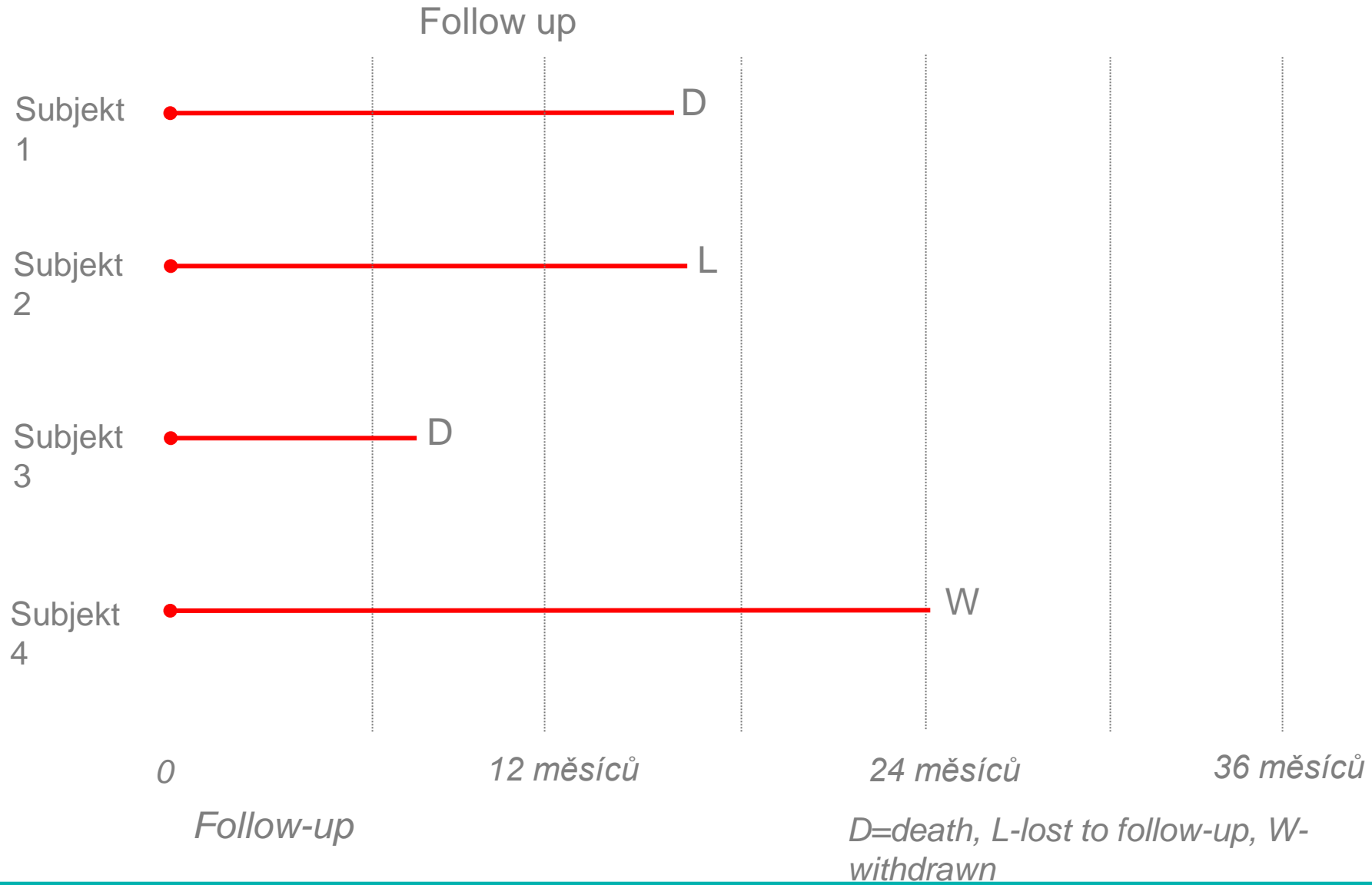
(=Accrual Period):.....12 měsíců

Minimální follow up:.....24 měsíců

Příklad cenzorovaných dat v klinických studiích



Příklad cenzorovaných dat v klinických studiích



Cíl studie:

Zhodnocení přínosu udržovací léčby na prodloužení doby do relapsu u dětských pacientů s akutní lymfoblastickou leukémií

Design studie:

Po dosažení kompletní remise (CR) po primární léčbě, byli pacienti randomizováni do dvou ramen:

- Placebo
- 6-mercaptopurine (6-MP)

Primární endpoint:

Time to Progression (TTP)

Výsledek modelové studie

Celkem randomizováno 42 pacientů (1:1):

Placebo: 21 z 21 pacientů relabovalo

6-MP: 12 z 21 pacientů při ukončení studie stále v CR

Doba sledování pacientů v týdnech:

Placebo: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

6-MP: 6, 6, 6, 6*, 7, 9*, 10, 10*, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35*
(*cenzorovaná pozorování)

Možnosti analýzy dat modelové studie

1/ Srovnání průměru nebo mediánu doby do relapsu

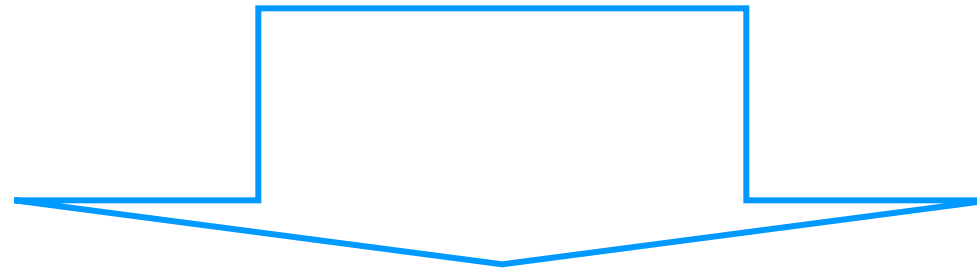
- není možné, neznáme u cenzorovaných pacientů

2/ Srovnání podílu relapsů v obou skupinách

(100% pro placebo, 43% pro 6-MP)

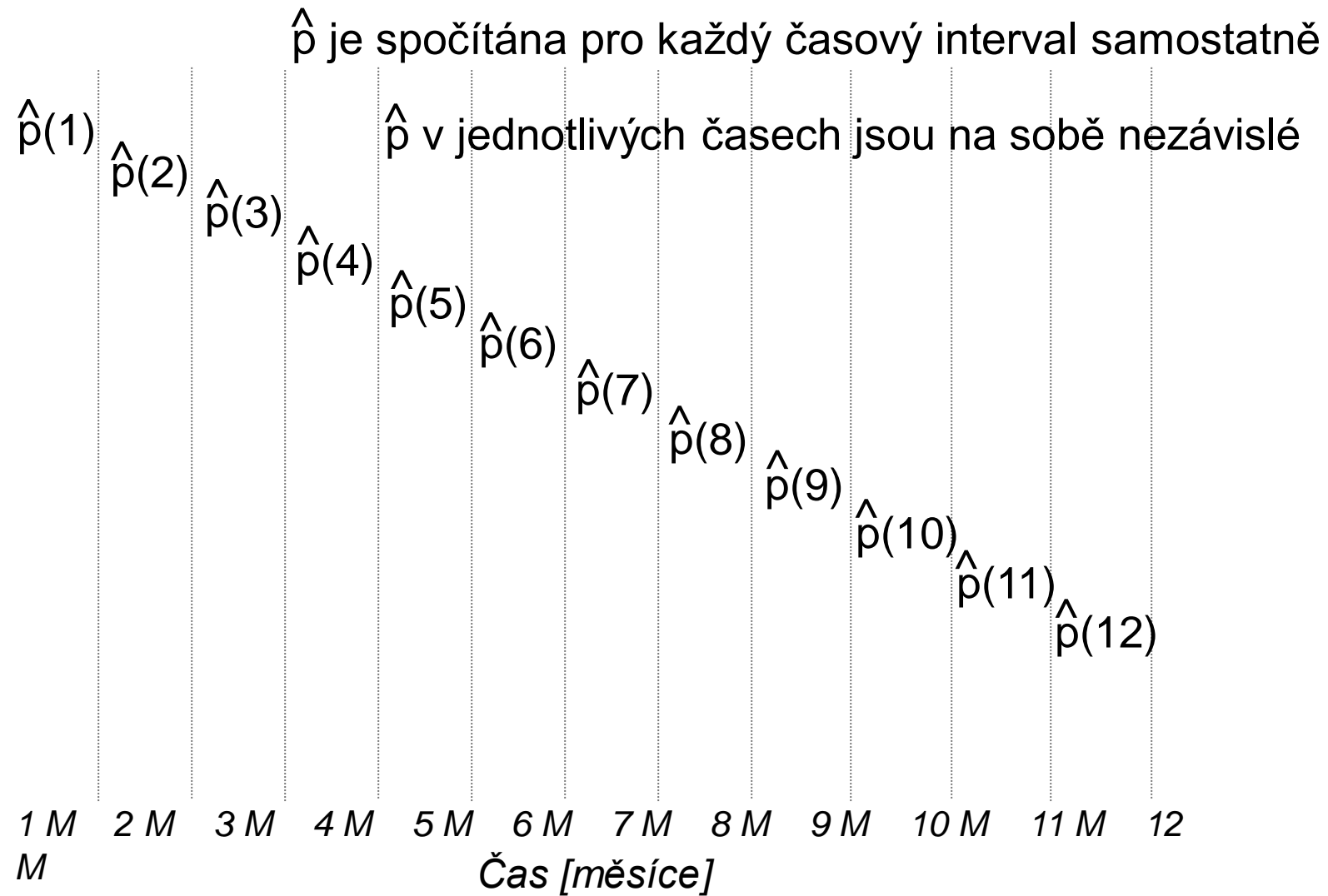
- nepřináší nám informaci o prodloužení doby do relapsu

- u cenzorovaných pacientů je možné, že při delším follow-up by k relapsu došlo



**Nutnost použití jiné techniky odhadu přežití:
kumulativní pravděpodobnost přežití**

Základní princip odhadu kumulativní pravděpodobnosti přežití



Výpočet kumulativní pravděpodobnosti přežití

Podmíněná pravděpodobnost přežití 1. měsíc od data diagnózy:

$$\hat{p}(1) = \frac{\text{počet pacientů vstupujících do studie} - \text{počet pacientů zemřelých 1. měsíc}}{\text{počet pacientů vstupujících do studie}}$$

Podmíněná pravděpodobnost přežití 6. měsíc od data diagnózy:

$$\hat{p}(6) = \frac{\text{počet pacientů „at risk“ v 6. měsíci} - \text{počet pacientů zemřelých 6. měsíc}}{\text{počet pacientů „at risk“ v 6. měsíci}}$$

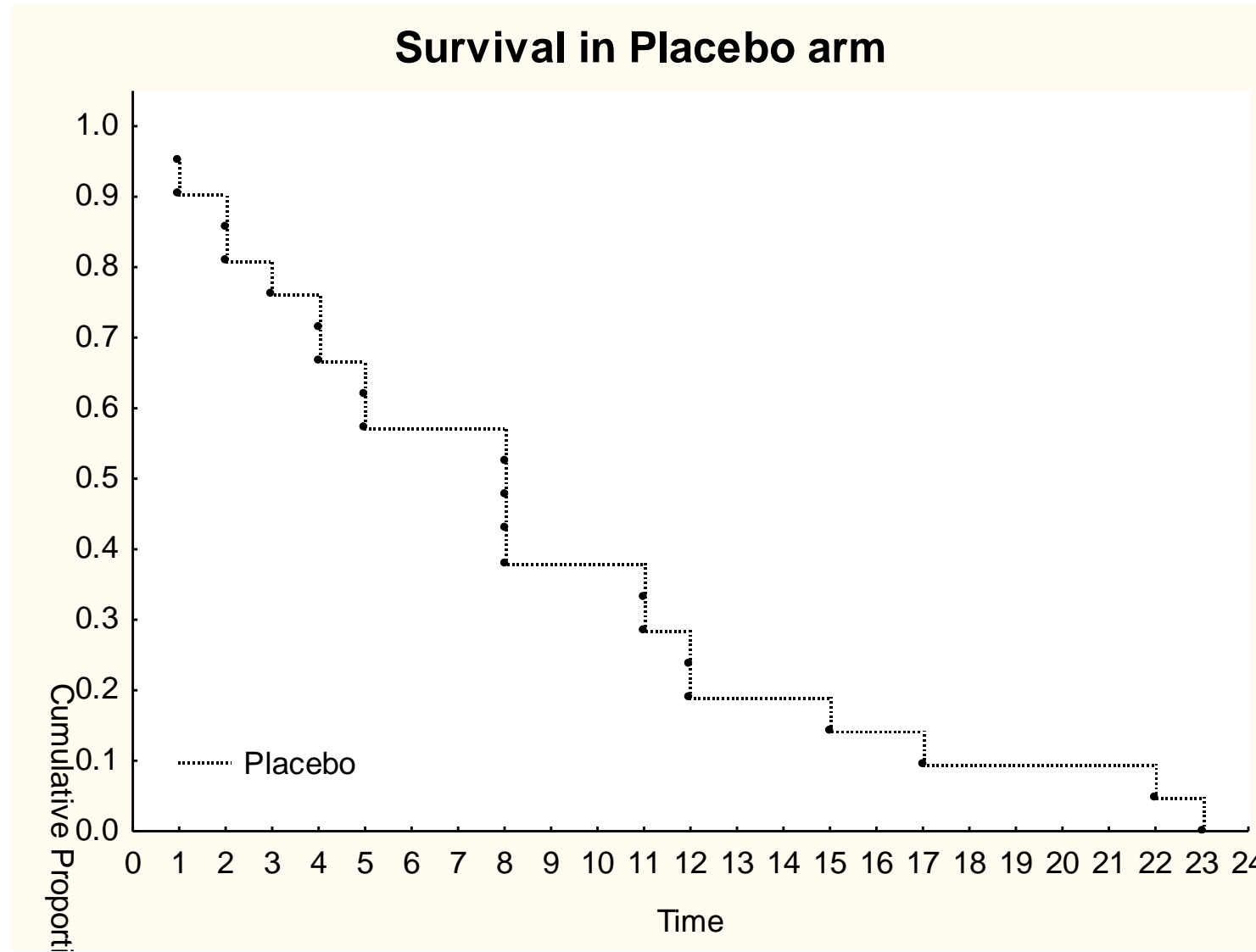
Kumulativní pravděpodobnost přežití 12 měsíců od data diagnózy:

$$P(12) = \hat{p}(1) \times \hat{p}(2) \times \hat{p}(3) \dots \times \hat{p}(12)$$

Výpočet kumulativní pravděpodobnosti přežití: Placebo

Doba do progrese $t_{(j)}$	Počet cenzorovaných	Počet relapsů d_j	Počet pacientů „at risk“ n_j	Podmíněná pravděpodobnost přežití $p_j=(n_j-d_j)/n_j$	Kumulativní pravděpodobnost přežití $P(t)$
1	0	2	21	$19/21=0,905$	0,905
2	0	2	19	$17/19=0,895$	$0,905 \times 0,895 = \mathbf{0,810}$
3	0	1	17	$16/17=0,941$	$0,810 \times 0,941 = \mathbf{0,762}$
4	0	2	16	$14/16=0,875$	$0,762 \times 0,875 = \mathbf{0,667}$
5	0	2	14	$12/14=0,857$	$0,667 \times 0,857 = \mathbf{0,571}$
8	0	4	12	$8/12=0,667$	$0,571 \times 0,667 = \mathbf{0,381}$
11	0	2	8	$6/8=0,750$	$0,381 \times 0,750 = \mathbf{0,286}$
12	0	2	6	$4/6=0,667$	$0,286 \times 0,667 = \mathbf{0,191}$
15	0	1	4	$3/4=0,750$	$0,191 \times 0,750 = \mathbf{0,143}$
17	0	1	3	$2/3=0,667$	$0,143 \times 0,667 = \mathbf{0,095}$
22	0	1	2	$1/2=0,500$	$0,095 \times 0,500 = \mathbf{0,048}$
23	0	1	1	$0/1=0,000$	$0,048 \times 0,000 = \mathbf{0,000}$

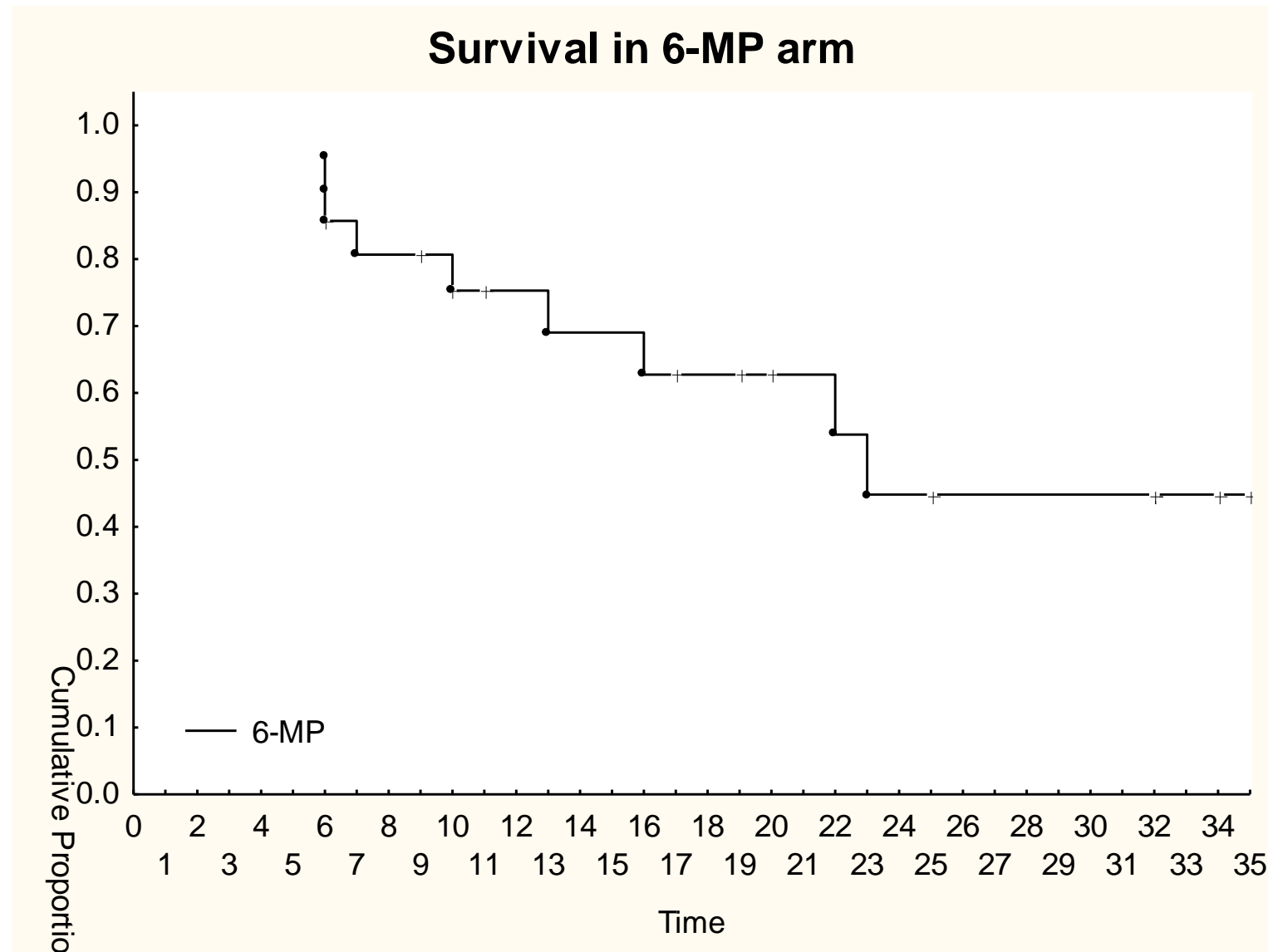
Kaplan-Meier křivka přežití



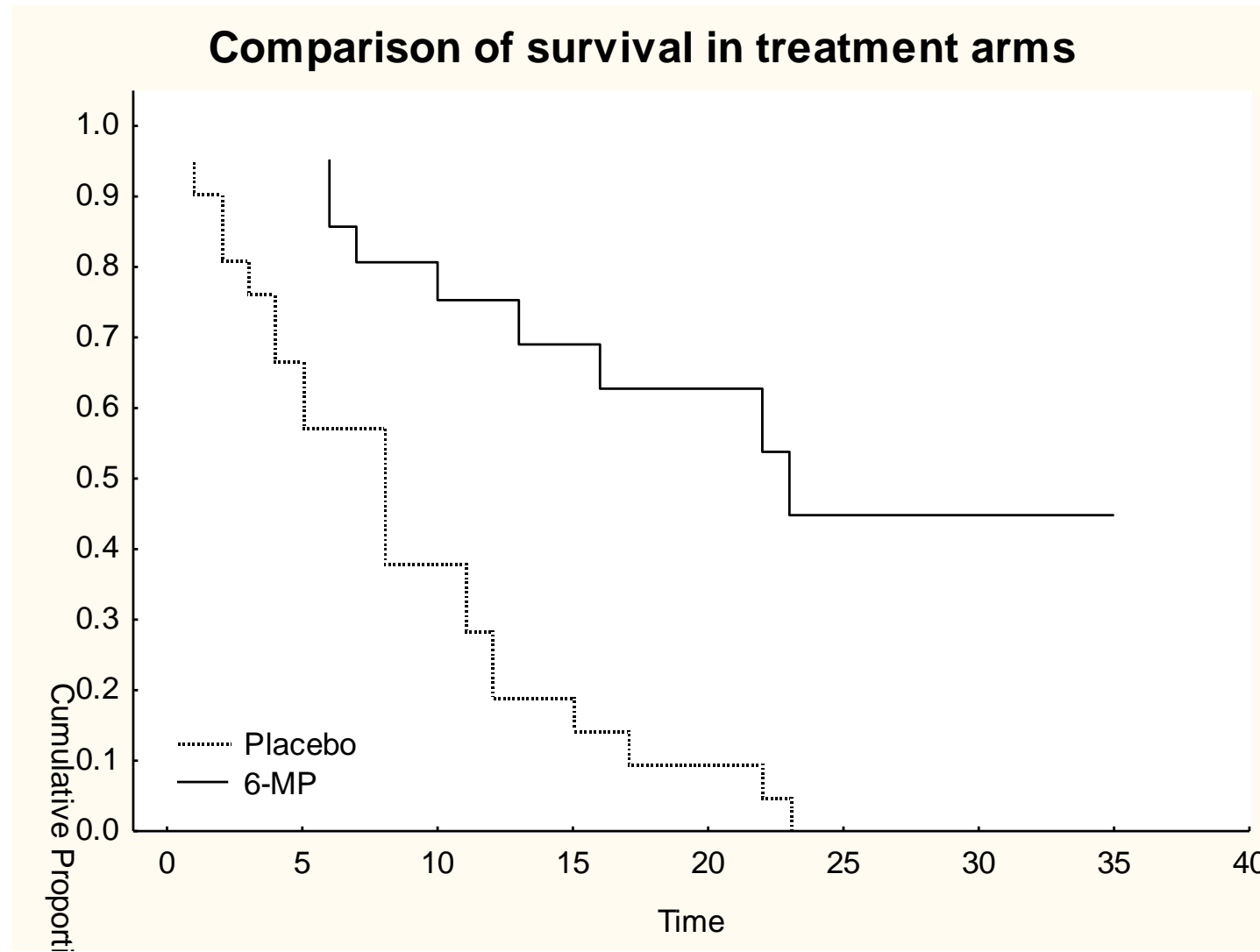
Výpočet kumulativní pravděpodobnosti přežití: Aktivní rameno

Doba do progrese $t_{(j)}$	Počet cenzorovaných	Počet relapsů d_j	Počet pacientů „at risk“ n_j	Podmíněná pravděpodobnost přežití $p_j=(n_j-d_j)/n_j$	Kumulativní pravděpodobnost přežití $P(t)$
6	1	3	21	$18/21=0,857$	0,857
7	0	1	17	$16/17=0,941$	$0,857 \times 0,941 = \mathbf{0,807}$
9	1	0	16		
10	1	1	15	$14/15=0,933$	$0,807 \times 0,933 = \mathbf{0,753}$
11	1	0	13		
13	0	1	12	$11/12=0,917$	$0,753 \times 0,917 = \mathbf{0,690}$
16	0	1	11	$10/11=0,909$	$0,690 \times 0,909 = \mathbf{0,628}$
17	1	0	10		
19	1	0	9		
20	1	0	8		
22	0	1	7	$6/7=0,857$	$0,628 \times 0,857 = \mathbf{0,538}$
23	0	1	6	$5/6=0,833$	$0,538 \times 0,833 = \mathbf{0,448}$

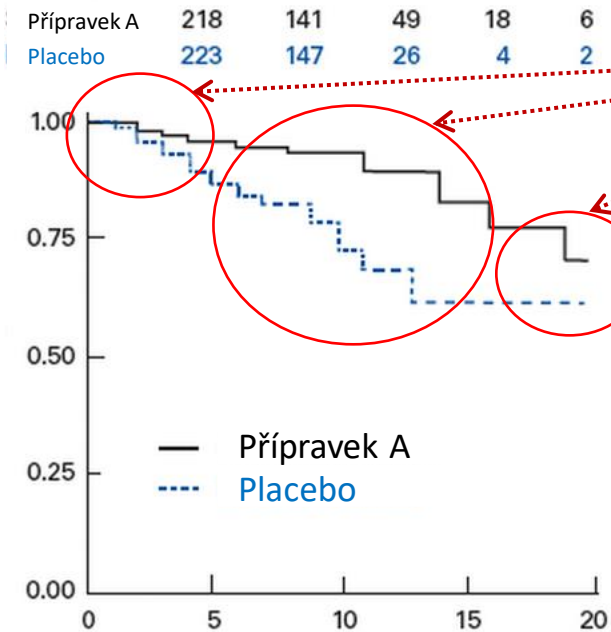
Kaplan-Meier křivka přežití



Srovnání přežití v obou ramenech



Srovnání výpočtu RR, OR a HR v jedné studii



Křivky přežití pro přípravek A a Placebo jsou v průběhu follow-up vzájemně odlišné

Hazard Ratio (HR) vyjadřuje poměr rizika nastání události v průběhu celého follow-up a je tedy nejspolehlivějším vyjádřením srovnání rizika v obou ramenech

Při interpretaci křivek přežití bývá často přeceňován jejich „konec“, který je však z důvodu velmi nízkého počtu pacientů „at risk“ velmi nespolehlivý pro jakékoliv smysluplné interpretace.

Hazard ratio

Vyjadřuje odhad poměru rizika úmrtí, progresu, nebo relapsu (podle definice endpointu) mezi srovnávanými skupinami

$$\text{HR} = \frac{\text{Riziko relapsu na experimentální léčbě}}{\text{Riziko relapsu na placebu}}$$

< 1

Léčba je lepší

= 1

Bez efektu

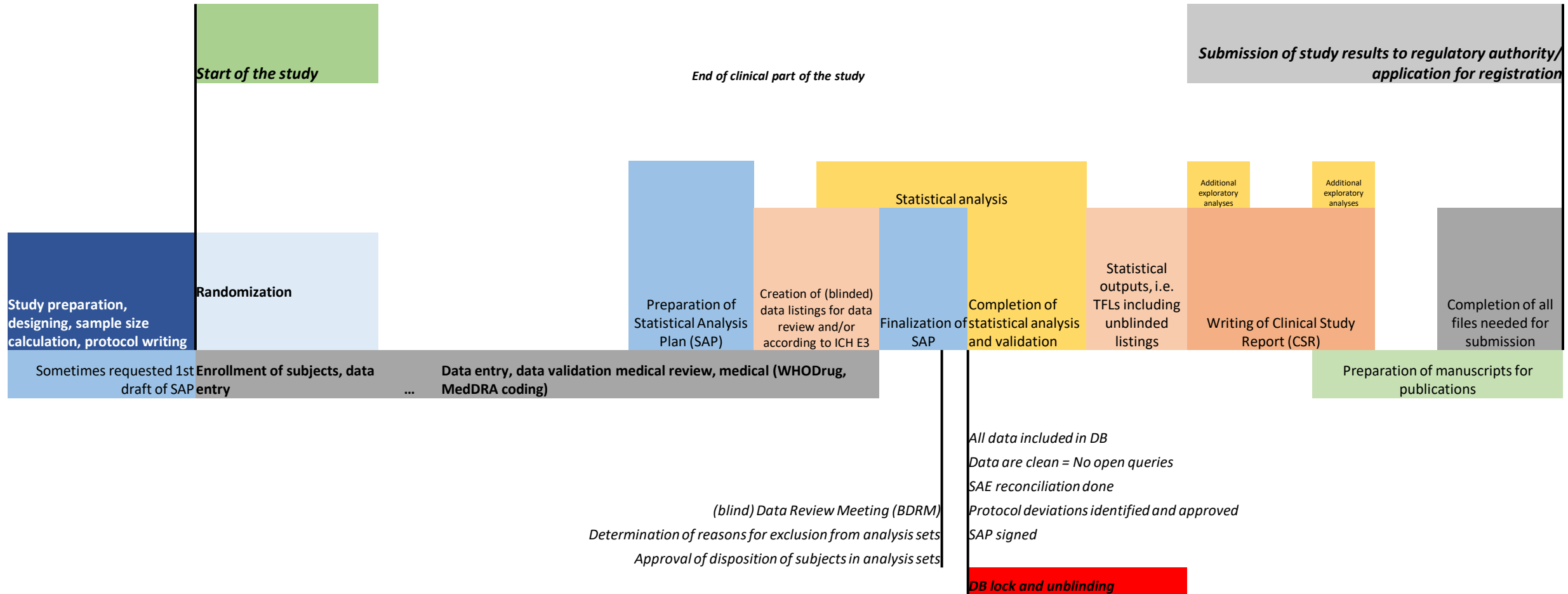
> 1

Léčba je horší

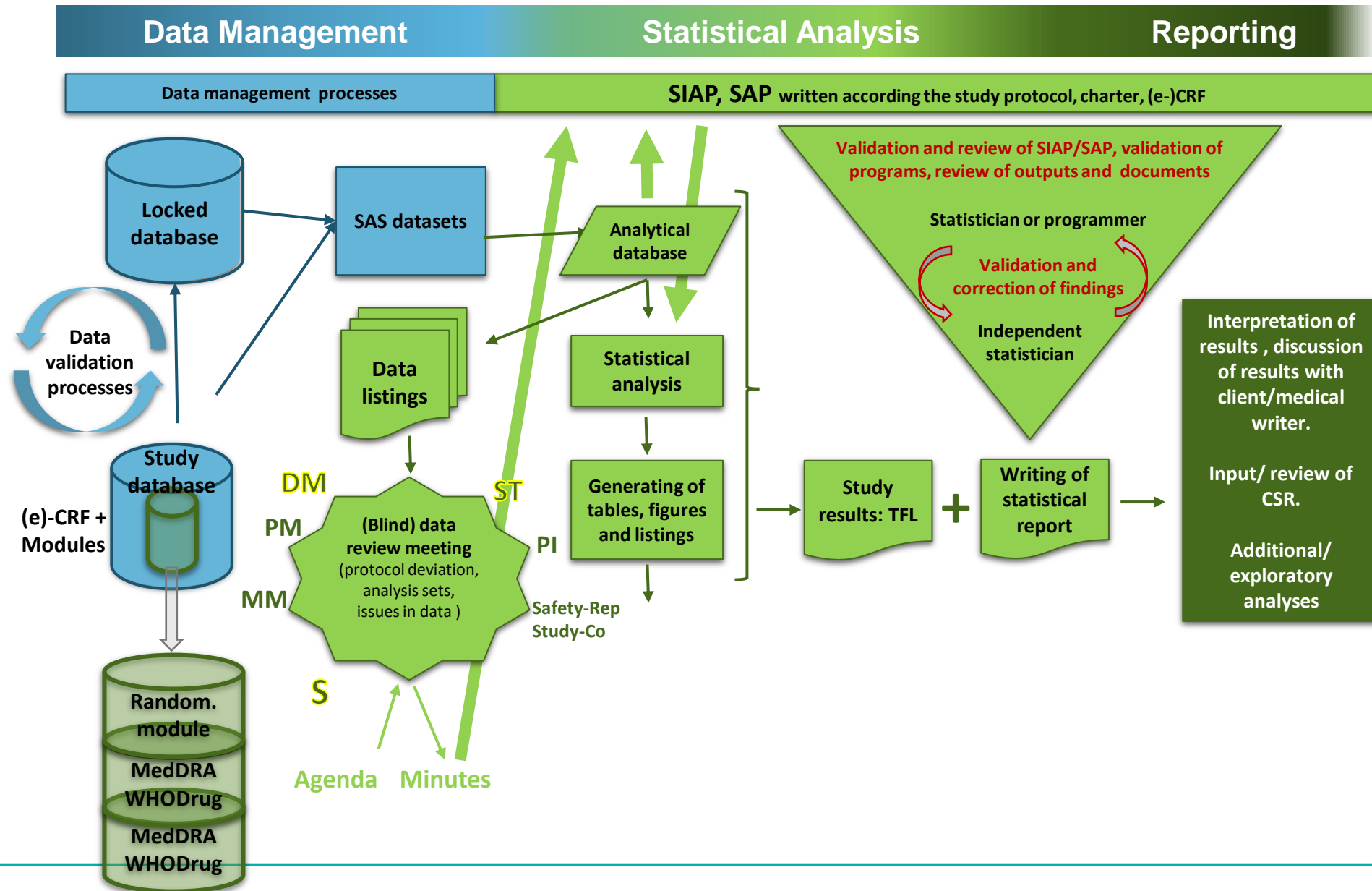
Interval spolehlivosti (např. pro HR=0,68: 0,42-0,88) nesmí obsahovat hodnotu 1 aby byl výsledek statisticky signifikantní.

Role statistika v klinických studiích

Zapojení statistika do klinické studie



Proces statistického zpracování dat klinického hodnocení



Zapojení statistika do klinické studie

Protokol studie



ICH Guidelines

Q	<ul style="list-style-type: none">• Stability• Impurities testing• GMP
S	<ul style="list-style-type: none">• Carcinogenicity• Genotoxicity• Reprotoxicity
E	<ul style="list-style-type: none">• Clinical trials• Pharmacogenomics
M	<ul style="list-style-type: none">• MedDRA• CTD• Electronic Standards



TABLE OF CONTENTS	
Abbreviations And Definitions Of Terms	
PROTOCOL SYNOPSIS	
Table 1: Schedule of Study Visits, Procedures & Measurements	
Figure 1: Study Diagram	
1	Background Information And Rationale
1.1	Introduction
1.2	Name And Description Of Investigational Product Or Description Of Intervention
1.3	Findings From Non-Clinical And Clinical Studies
1.3.1	Non-Clinical Studies
1.3.2	Clinical Studies
1.4	Selection Of Drugs And Dosages
1.5	Compliance Statement
1.6	Discussion Of Relevant Literature And Data
2	Study Objectives
2.1	Primary Objective
2.2	Secondary Objectives
3	Investigational Plan
3.1	General Schema Of Study Design (overview)
3.1.1	Screening Phase
3.1.2	Treatment Phase
3.1.3	Follow-Up Phase (if applicable)
3.2	Randomization And Blinding
3.3	Study Duration, Enrollment And Number Of Sites
3.3.1	Duration Of Study
3.3.2	Total Number Of Study Sites/Total Number Of Subjects Projected
3.4	Study Population
3.4.1	Inclusion Criteria
3.4.2	Exclusion Criteria
4	Study Procedures
4.1	Screening Visit
4.2	Treatment Phase
4.2.1	Visit 1
4.2.2	Visit 2 (etc.)
4.3	Follow-Up Phase
4.4.1	Visit
4.5	Unscheduled Visits
4.6	Concomitant Medication
4.7	Rescue Medication Administration
4.8	Subject Completion/Withdrawal
4.8.1	Early Termination Study Visit
5	Study Evaluations and Measures
5.1	Screening And Baseline Evaluations (Procedures & Measurements)
5.1.1	Physical Exams
5.1.2	Laboratory Tests
5.1.3	Other Procedures
5.2	Efficacy Evaluations
5.2.1	Diagnostic Tests, Scales, Measures, Etc.
5.3	Pharmacokinetic Evaluation (if applicable)
5.4	Safety Evaluations/Measurements
6	Statistical Considerations
6.1	Primary Endpoint
6.2	Secondary Endpoints
6.3	Statistical Methods
6.3.1	Baseline Data
6.3.2	Efficacy Analysis
6.3.3	Safety Analysis
6.4	Sample Size And Power
6.5	Interim Analysis (if applicable)
7	Study Medication (Adapt for Other Interventions)
7.1	Description
7.1.1	Packaging
7.1.2	Labeling
7.1.3	Dosing
7.1.4	Treatment Compliance And Adherence
7.1.5	Drug Accountability
8	Safety Management
8.1	Clinical Adverse Events
8.2	Adverse Event Reporting
8.3	Definition Of An Adverse Event
8.4	Definition Of A Serious Adverse Event (SAEs)
8.5	IRB/IEC Notification Of SAEs
8.6	Investigator Reporting Of SAEs to Sponsor
8.7	Medical Emergencies
9	Study Administration
9.1	Treatment Assignment Methods
9.1.1	Randomization
9.1.2	Blinding
9.1.3	Unblinding
9.2	Data Collection And Management
9.3	Regulatory And Ethical Considerations
9.3.1	Data And Safety Monitoring Plan
9.3.2	Risk Assessment
9.3.3	Potential Benefits Of Trial Participation
9.3.4	Risk-Benefit Assessment
9.4	Informed Consent/Assent Process
9.5	Payment To Subjects/Families
9.6	Confidentiality
10	Publication
11	References

Publikace výsledků studie v renomovaném časopise není zárukou jejich správnosti!

The screenshot shows the CONSORT website header with the logo and navigation menu. A red box highlights a text snippet from a BMJ article. A red arrow points from the text 'Welcome to the CONSORT' to the highlighted box. Below the box, there are links for 'Read more' and 'Read more news stories'.

CONSORT
TRANSPARENT REPORTING of TRIALS

Home CONSORT Statement Extensions

Welcome to the CONSORT
CONSORT is for
Trials, encompass various
CONSORT Group to alleviate
reporting of randomized cont

The main product of CONSORT
which is an evidence-based, minimum set of recommendations for
reporting RCTs. It offers a standard way for authors to prepare
reports of trial findings, facilitating their complete and transparent
reporting, and aiding their critical appraisal and interpretation.

Unfortunately significance tests of baseline differences are still common⁽²³⁾ ⁽³²⁾ ⁽²¹⁰⁾; they were reported in half of 50 RCTs trials published in leading general journals in 1997.⁽¹⁸³⁾ Such significance tests assess the probability that observed baseline differences could have occurred by chance; however, we already know that any differences are caused by chance. Tests of baseline differences are not necessarily wrong, just illogical.⁽²¹¹⁾ Such hypothesis testing is superfluous and can mislead investigators and their readers. Rather, comparisons at baseline should be based on consideration of the prognostic strength of the variables measured and the size of any chance imbalances that have occurred.⁽²¹¹⁾

<http://www.bmj.com/content/348/bmj>
[Read more](#)
[Read more news stories](#)

**P-values in baseline table
control group are
common in high quality**

MJ Knol, RHH Groen

Table 1. Total number of RCTs, number of RCTs reporting *p*-values in the baseline table, and number of RCTs without baseline table published in 2008, 2009 or 2010 in major general medical and cardiovascular journals

Journal	Total N	<i>p</i> -value in baseline table N (%)	No baseline table N
JAMA	149	86 (58.5%)	2
NEJM	321	169 (54.3%)	10
Circulation	141	59 (47.2%)	16
European Heart Journal	86	43 (43.9%)	12
Annals of Internal Medicine	74	15 (21.1%)	3
BMJ	150	11 (7.7%)	7
Lancet	254	5 (2.0%)	10
Total	1175	388 (34.8%)	60