

Analýza dat pro Neurovědy



RNDr. Eva Janoušová
doc. RNDr. Ladislav Dušek, Dr.

Jaro 2013

Blok 8

Jak analyzovat přežití pacientů.

Osnova

1. Analýza přežití
2. Coxův model proporcionálních rizik

1. Analýza přežití

Analýza přežití

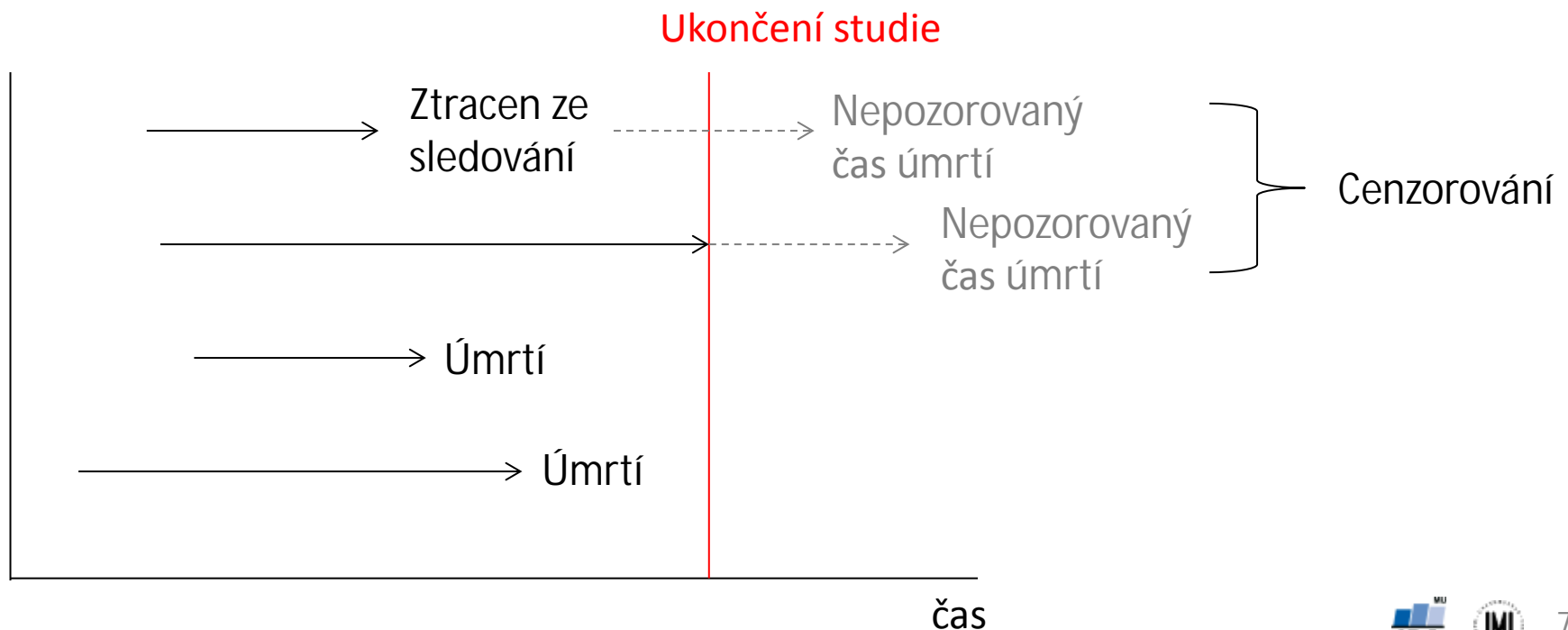
- Analýza doby od vstupní události do výskytu sledované (koncové) události.
- **Vstupní událost** – např.:
 - narození
 - začátek léčby
 - počátek nemoci
 - vstup do studie
 - dosažení remise (uzdravení pacienta)
 - operace
 - začátek používání přístroje
- **Koncová (sledovaná) událost** – např.:
 - úmrtí
 - výskyt progresse onemocnění
 - výskyt relapsu onemocnění (tzn. návrat onemocnění)
 - dosažení remise (uzdravení pacienta)
 - porucha přístroje
- Dobu mezi vstupní a koncovou událostí nazýváme jako **doba přežití**.

Analýza přežití

- Uplatnění analýzy přežití:
 - v lékařském výzkumu:
 - hodnocení celkového přežití („Overall Survival“ – OS)
 - hodnocení času do progresu („Progression-free Survival“ – PFS)
 - a další
 - v průmyslu
 - v zemědělství

Cenzorování

- Doba přežití dané osoby je cenzorována, pokud během pozorování nenastane sledovaná událost – důvody:
 - ztráta kontaktu s danou osobou (osoba se přestěhuje nebo přestane chodit na prohlídky)
 - v době uzavření studie se sledovaná událost ještě nevyskytla
 - pozorovaná osoba zemřela na jinou nemoc



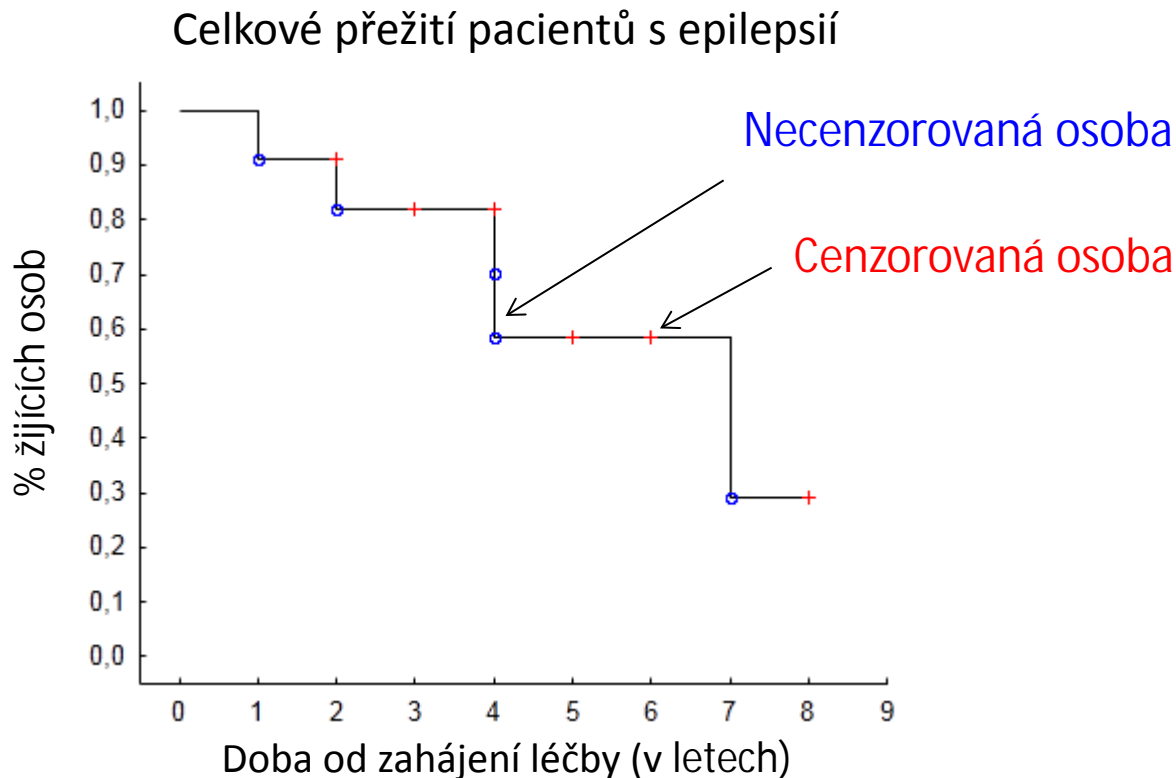
Analýza přežití – vstupní data

- Pro každého člověka dvojice hodnot:
 - T – čas přežití:
 - u osob, u nichž sledovaná událost nastala, je to čas od vstupní události do sledované události
 - u osob, u nichž sledovaná událost nenastala, je to čas od vstupní události do ukončení studie
 - C – identifikátor cenzorování – pouze 2 hodnoty:
 - 1 ... sledovaná událost nastala
 - 0 ... sledovaná událost nenastala (osoba je cenzorována)
- Cenzorovaná pozorování (tzn. osoby, u nichž sledovaná událost nenastala) nesmíme z analýzy vyhodit – obsahují informaci!!!

Neparametrické odhady křivky přežití

- V klinickém výzkumu i populačním modelování jsou pro popis přežití standardem neparametrické metody – Kaplan-Meierova metoda a metoda Life-tables.
1. **Kaplan-Meierova metoda** – založena na jednotlivých pozorovaných časech přežití, vhodná zejména pro hodnocení dat klinických studií – vyžaduje přesný záznam doby sledování.
 2. **Life-tables** – založena na agregaci pozorování do časových intervalů, vhodná zejména pro popis přežití na populační úrovni, kde není k dispozici tak kvalitní záznam doby sledování.

Kaplan-Meierova křivka přežití



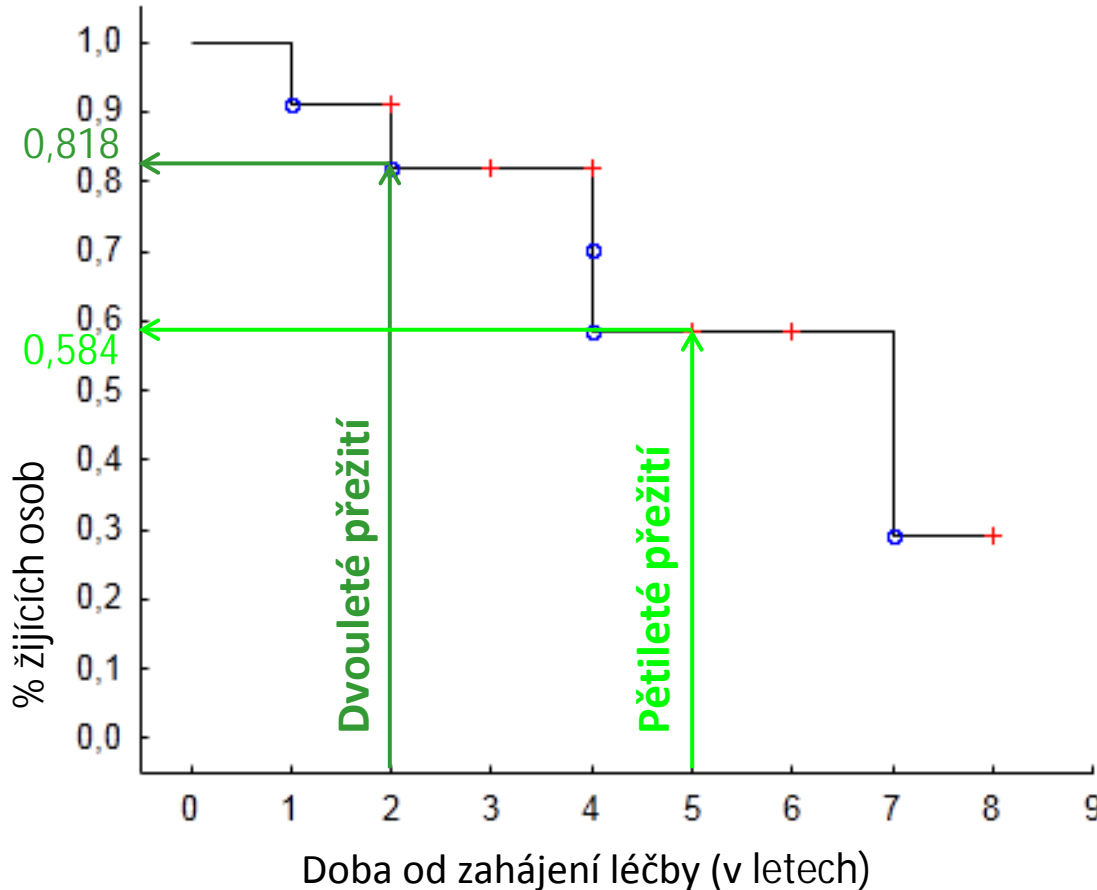
Vstupní data

Pacient	Čas	Úmrtí
1	1	1
2	2	1
3	2	0
4	3	0
5	4	1
6	4	1
7	4	0
8	5	0
9	6	0
10	7	1
11	8	0

- Křivka přežití odráží **procento žijících osob v daném čase**.
- Křivka přežití je **klesající**:
 - pokles křivky přežití nastane v čase koncové události u necenzorované osoby
 - velikost „schodu“ je dána počtem osob, kteří v daném čase zůstávají „v riziku“

Křivka přežití – hodnocení x-letého přežití

Celkové přežití pacientů s epilepsií



Kaplan-Meier (Product-limit) analysis (Data_preziti) Note: Censored cases are marked with +

	Time	Cumulative Survival	Standard Error
1	1	0,909	0,087
2	2	0,818	0,116
3+	2		
4+	3		
5	4	0,701	0,147
6	4	0,584	0,163
7+	4		
8+	5		
9+	6		
10	7	0,292	0,222
11+	8		

2-leté přežití: 81,8% (tzn. po dvou letech od zahájení léčby žije 81,8% pacientů)

5-leté přežití: 58,4% (tzn. po pěti letech od zahájení léčby žije 58,4% pacientů)

Křivka přežití – hodnocení x-letého přežití – výpočet intervalu spolehlivosti

Výpočet intervalu spolehlivosti pro x-leté přežití:

$$\text{Odhad pravděpodobnosti přežití v čase } t \pm 1,96 \times \text{SE (odhadu)}$$

Kaplan-Meier (Product-limit) analysis (Data_preziti) Note: Censored cases are marked with +

	Time	Cumulative Survival	Standard Error
1	1	0,909	0,087
2	2	0,818	0,116
3+	2		
4+	3		
5	4	0,701	0,147
6	4	0,584	0,163
7+	4		
8+	5		
9+	6		
10	7	0,292	0,222
11+	8		

Dolní mez IS:

Horní mez IS:

$$0,818 - 1,96 * 0,116 = 0,591 \quad 0,818 + 1,96 * 0,116 = 1,045$$

$$0,584 - 1,96 * 0,163 = 0,265 \quad 0,818 + 1,96 * 0,116 = 0,903$$

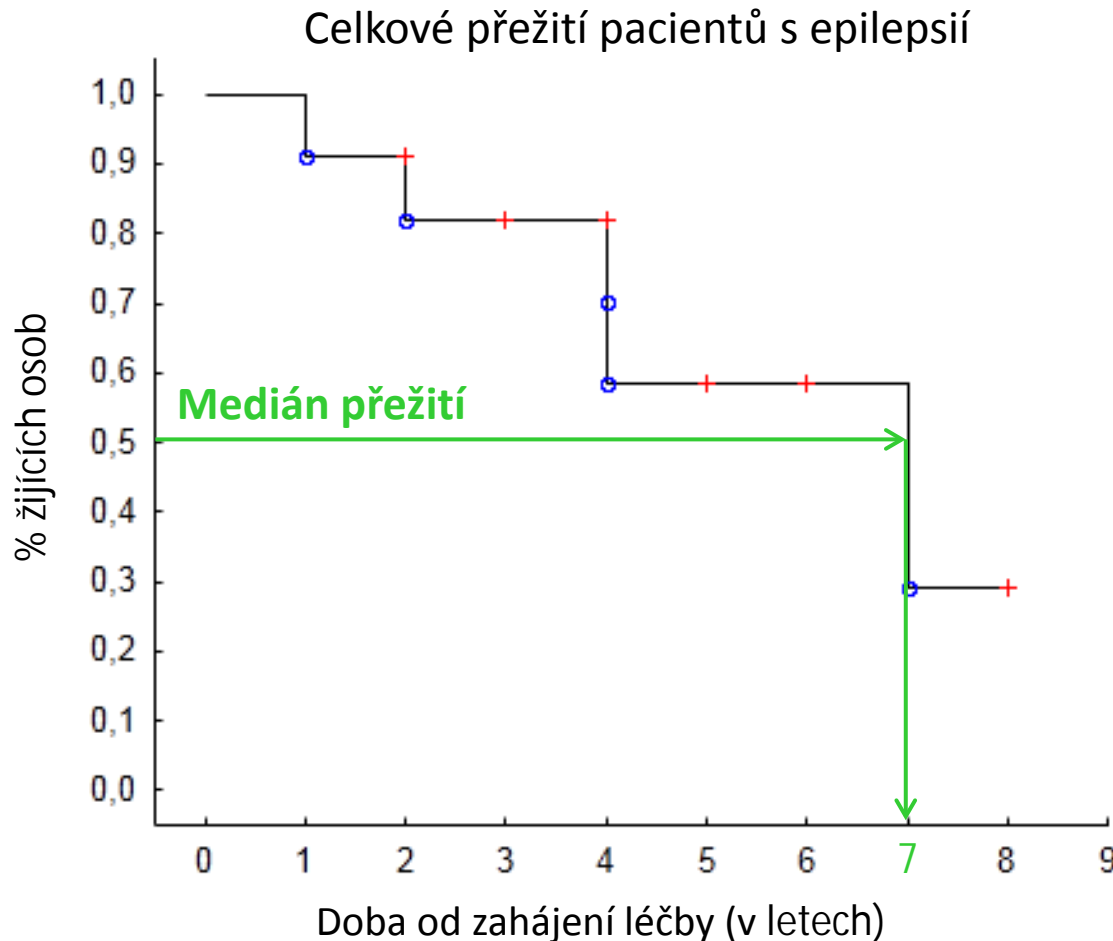
Závěr: 2-leté přežití: 81,8% (59,1%; 100,0%)
5-leté přežití: 58,4% (26,5%; 90,3%)

Upozornění

- Pokud horní mez intervalu spolehlivosti vyjde větší než 1 (tzn. 100%), je třeba toto číslo nahradit hodnotou 1 (resp. 100%).
- Pokud dolní mez intervalu spolehlivosti vyjde menší než 0 (tzn. 0%), je třeba toto číslo nahradit hodnotou 0 (resp. 0%).
- **Procento žijících lidí totiž nemůže být větší než 100% a menší než 0% !**

Křivka přežití – medián přežití

Medián přežití je čas od vstupní události, v němž je pravděpodobnost přežití 50%, tedy čas, kterého se podle očekávání dožije polovina sledovaných pacientů.



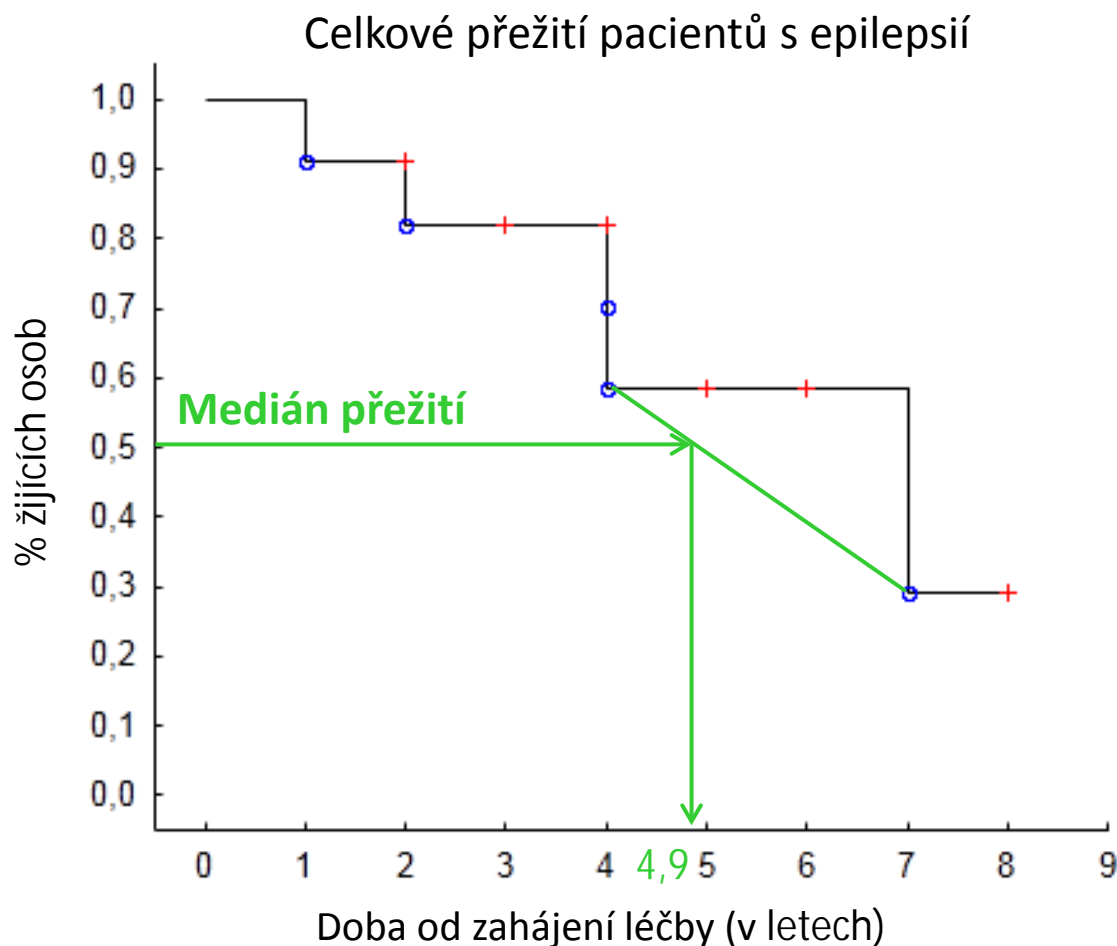
Závěr:

Medián přežití je 7 let.

(Tzn. čas, kterého se dožila polovina sledovaných pacientů, je 7 let.)

Křivka přežití – medián přežití v softwaru STATISTICA

Software STATISTICA provádí interpolaci – počítá aproximativní medián. Nevadí to, pokud je velký počet sledovaných událostí.



Survival Time	
25'th percentile (lower quartile)	3,2
50'th percentile (median)	4,9
75'th percentile (upper quartile)	6,0

Závěr:
Medián přežití je 4,9 let.

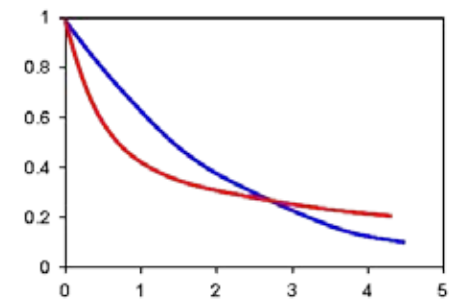
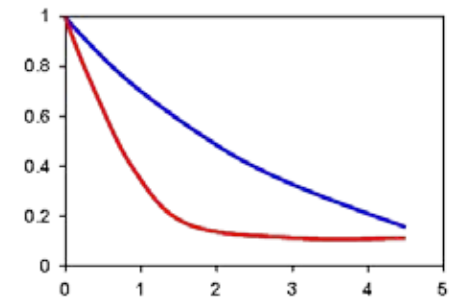
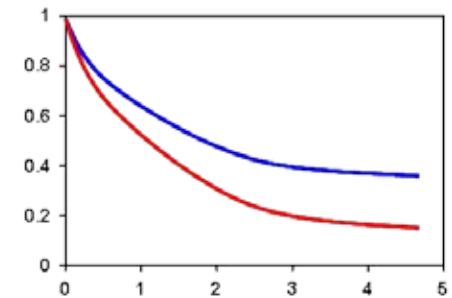
Poznámky

- **Sestrojovat křivky přežití za každou cenu je mnohdy zavádějící** – zvláště v případě použití stratifikačních kritérií vedoucích k nízkým N ve skupinách. è riziko zkreslení a dezinterpretace výsledků!!
- Podíl cenzorovaných pozorování je důležitou charakteristikou – je vhodné uvádět:
 - Podíl pacientů ztracených ze sledování (lost to follow-up).
 - Podíl pacientů „bez události“ na konci studie.
- S křivkami přežití by měla být vždy reportována **maximální a minimální doba sledování** dosažitelná ve studii (dáno začátkem náboru a datem ukončení studie) – samozřejmě ve vztahu k události, která nás zajímá.
- Je nutné mít na paměti **nízkou věrohodnost „konce“ křivky přežití** – zůstává-li ve studii 10 pacientů nebo méně, „skoky“ v přežití s každou další událostí jsou dramatické.

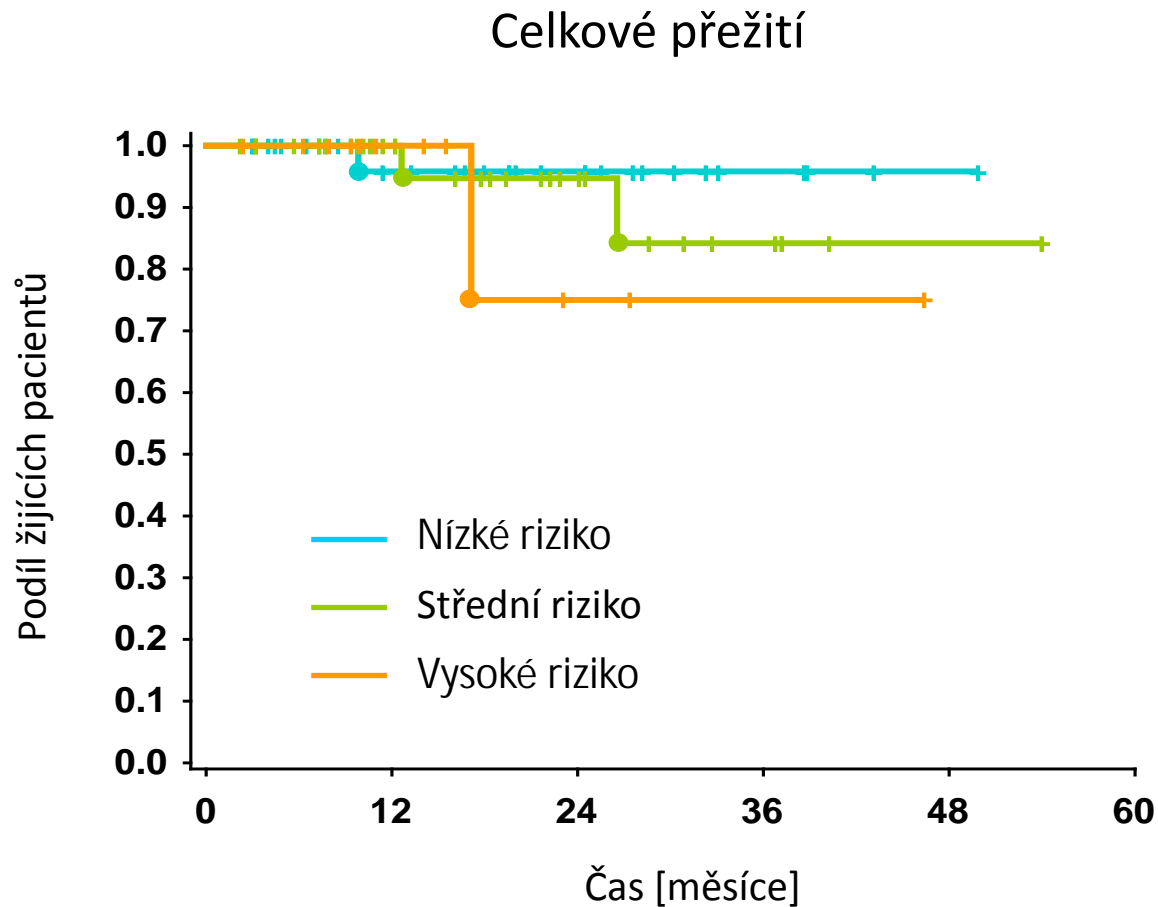
Srovnání křivek přežití

- Častým cílem klinického výzkumu je srovnání přežití dvou a více skupin pacientů
- Standardem v analýze klinických dat jsou opět neparametrické testy:
 - Log-rank test
 - Gehanův-Wilcoxonův test
- **Log-rank test** je zaměřen na srovnání očekávaných a pozorovaných počtů událostí v jednotlivých skupinách
- **Gehanův-Wilcoxonův test** umožňuje klást větší důraz na rozdíly v raných fázích sledování pacientů.

$S(t)$ v čase

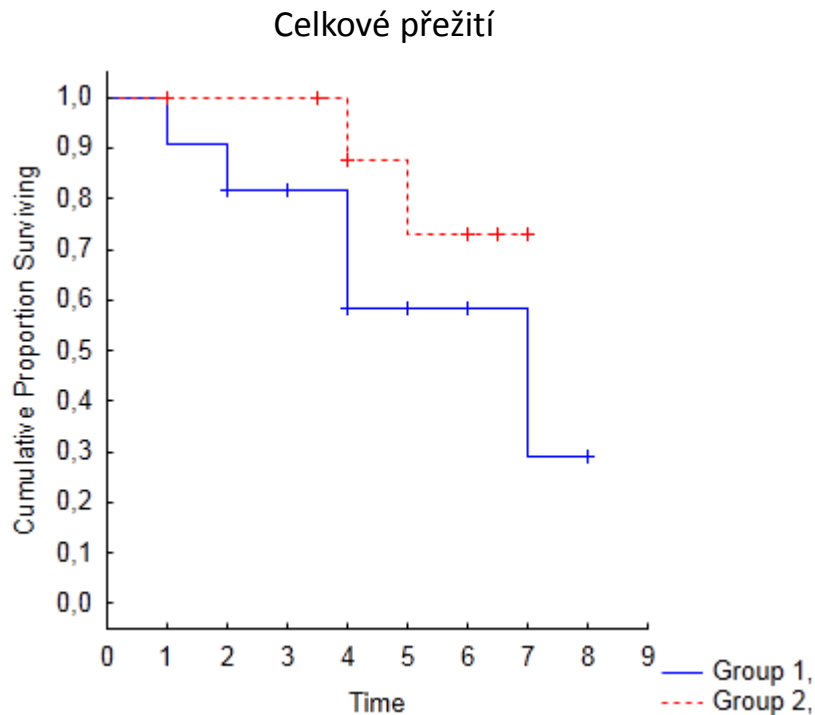


Srovnání křivek přežití – ukázka



Srovnání dvou křivek přežití – příklad

- Příklad:** Srovnajte přežití pacientů, kteří byli léčeni dvěma různými preparáty.



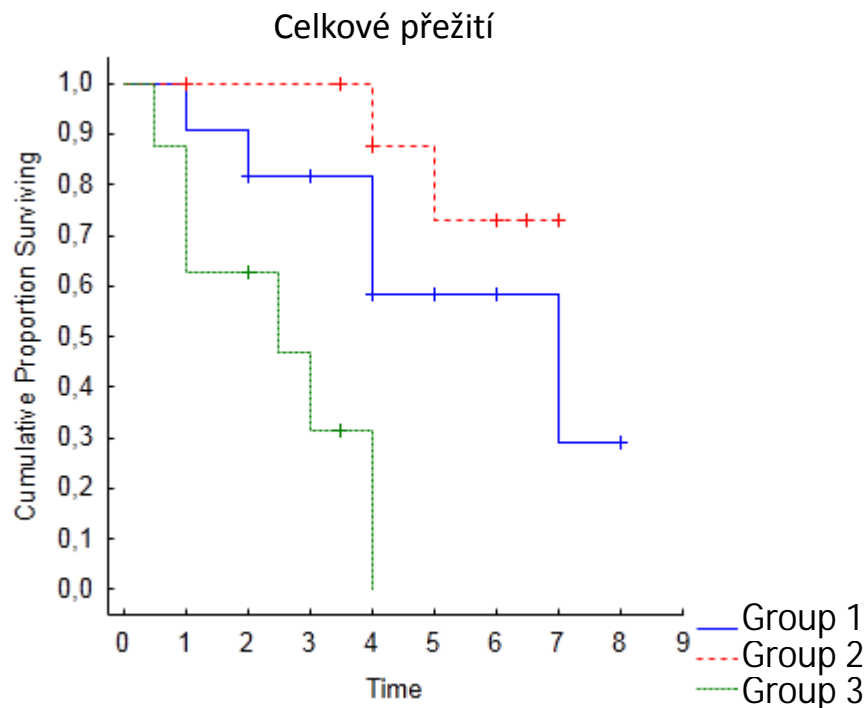
Life Table for Group 1 and Group 2
(Data_preziti_2kat) Group 1: Code 1,
Group 2: Code 2

	Group 1: Cum.%	Group 2: Cum.%
1,00	100,0	100,0
1,78	90,9	100,0
2,56	81,3	100,0
3,33	81,3	100,0
4,11	56,3	87,5
4,89	56,3	87,5
5,67	56,3	72,9
6,45	56,3	72,9
7,22	28,2	72,9
8,00	28,2	

Log-rank test: $p=0,231$

Srovnání třech křivek přežití – příklad

- **Příklad:** Srovnejte přežití pacientů se třemi diagnózami (CML, CLL a AML).



	Group 1	Group 2	Group 3
,50	100,0	100,0	100,0
1,33	90,9	100,0	62,5
2,17	81,3	100,0	62,5
3,00	81,3	100,0	46,9
3,83	81,3	100,0	28,1
4,67	56,3	86,7	0,0
5,50	56,3	72,2	0,0
6,33	56,3	72,2	0,0
7,17	28,2	72,2	0,0
8,00	28,2		0,0

$p=0,007$

2. Coxův model proporcionálních rizik

Coxův model proporcionálních rizik

- Analýza vlivu prognostických faktorů onemocnění na přežití pacientů, na dosažení remise apod.
- **Příklad:** Testování vlivu binární proměnné (např. užívání léčby B) na celkové přežití.

Výsledek analýzy:

Variable	Hazard ratio (HR) (poměr rizik)	95% conf. Int.	P-value
DRUG B	2.18	1.4 – 3.5	0.001

Interpretace:

U pacientů užívajících v období před vstupem do studie přípravek B, je více jak dvojnásobně vyšší riziko úmrtí ve sledovaném období než u pacientů neužívajících přípravek B.

Coxův model proporcionálních rizik

- **Příklad:** Testování vlivu kategoriální proměnné (např. věk při diagnóze) na celkové přežití.

Výsledek analýzy:

Age group	HR	95% conf. Int.	P-value
1: [20-29]	1.0		
2: [30-34]	3.31	1.37-8.01	<0.001
3: [35-39]	3.72	1.51-9.14	<0.001
4: [40-54]	6.43	2.56-16.12	<0.001

Interpretace:

S rostoucím věkem při diagnóze roste riziko úmrtí ve sledovaném období. Nárůst rizika je vztažen k věkové skupině 20 - 29 let.

Coxův model proporcionálních rizik

- **Příklad:** Testování vlivu spojité **proměnné** (např. věk při diagnóze) na celkové **přežití**.

Výsledek analýzy:

Variable	HR	95% conf. Int.	P-value
AGE _[5 years interval]	1.50	1.3 – 1.8	<0.001

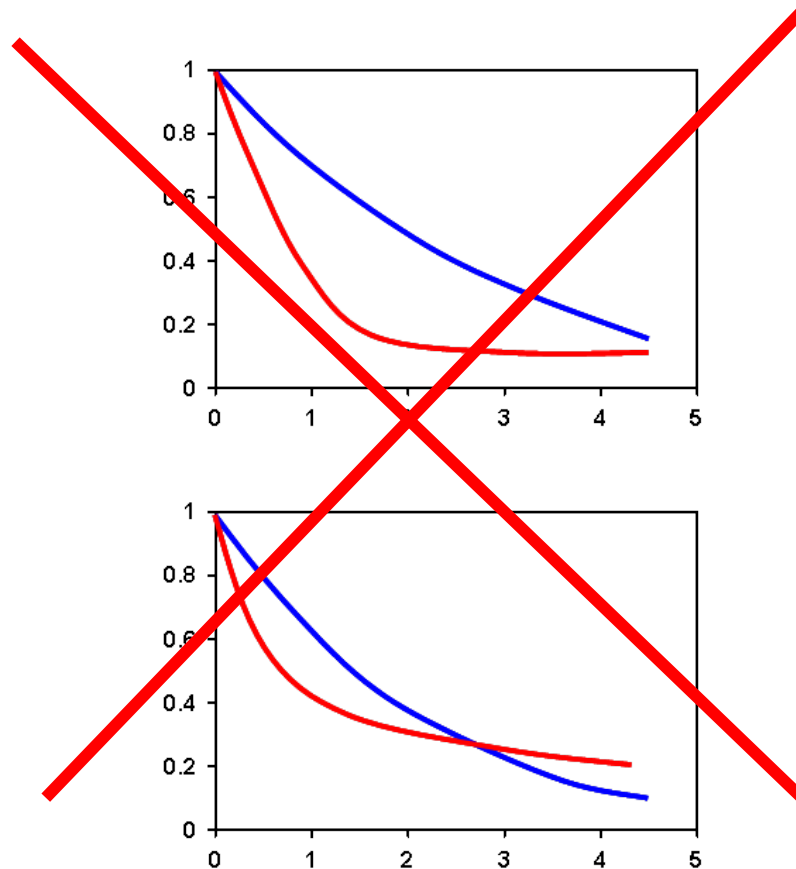
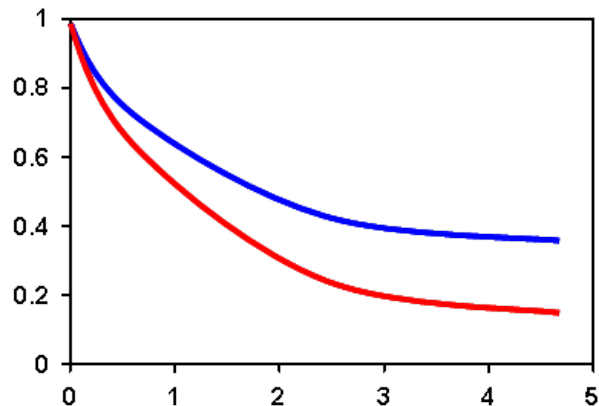
Interpretace:

Nárůst věku při diagnóze o 5 let zvyšuje riziko úmrtí 1,5-krát.

Coxův model proporcionálních rizik – důležité!

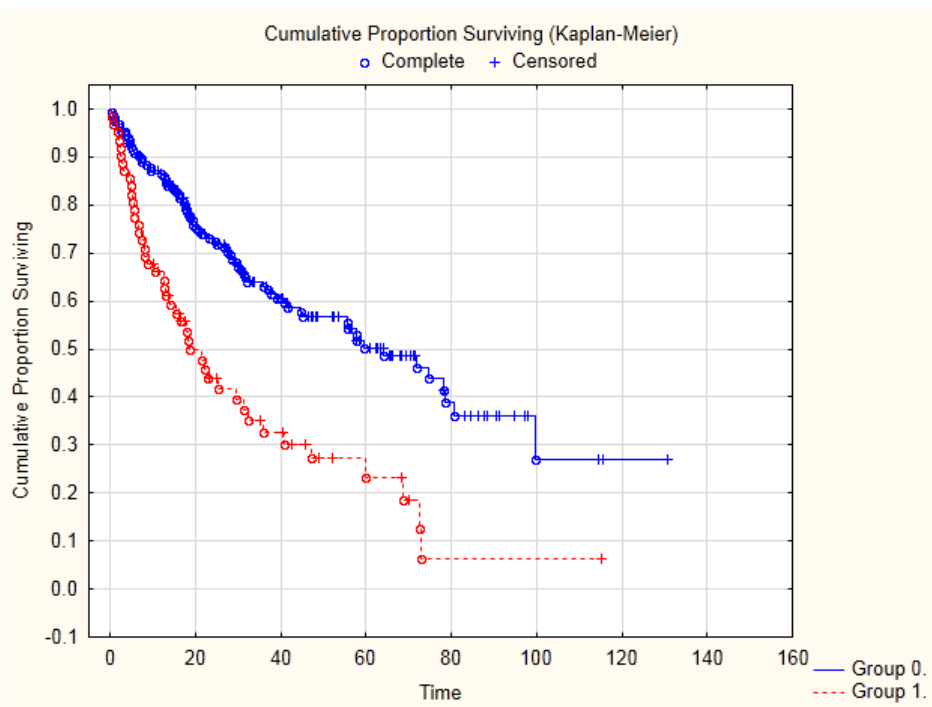
- Coxův model má smysl počítat, pokud se křivky přežití od sebe postupně oddalují. Pokud se křivky oddálí a pak se zase přiblíží nebo pokud se dokonce křivky překříží, tak nelze Coxův model pro výpočet poměru rizik (hazard ratio) použít!

OK



Coxův model proporcionálních rizik – příklad

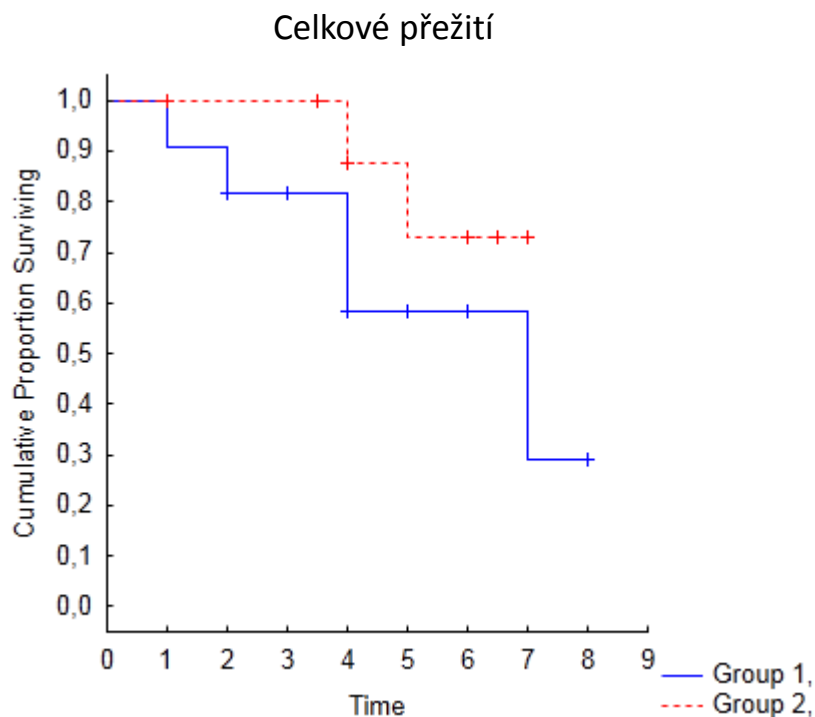
- **Příklad:** Vypočtete poměr rizika úmrtí podle ECOG skóre.



Parameter Estimates (Data_Cox)									
	Standard Error	Chi-square	P value	95% Lower CL	95% Upper CL	Hazard Ratio	95% Hazard Ratio Lower CL	95% Hazard Ratio Upper CL	
ECOG1	0,192764	20,60900	0,000006	0,497283	1,252903	2,399098	1,644247	3,500491	

Coxův model proporcionálních rizik – úkol

- Úkol: Vypočtete poměr rizika úmrtí podle typu léčby.



	Parameter Estimates (Data_preziti_2kat)						
	Chi-square	P value	95% Lower CL	95% Upper CL	Hazard Ratio	95% Hazard Ratio Lower CL	95% Hazard Ratio Upper CL
Skupina	1.204534	0.272417	-2.58078	0.727986	0.395977	0.075715	2.070905

Poděkování...

Příprava výukových materiálů předmětu „DSAN01 Analýza dat pro Neurovědy“ je finančně podporována prostředky projektu FRVŠ č. 942/2013 „Inovace materiálů pro interaktivní výuku a samostudium předmětu Analýza dat pro Neurovědy“

