MUNI
MED

Populační hodnocení přežití onkologických pacientů

T. Pavlík, A. Šikyňová, Z. Bortlíček

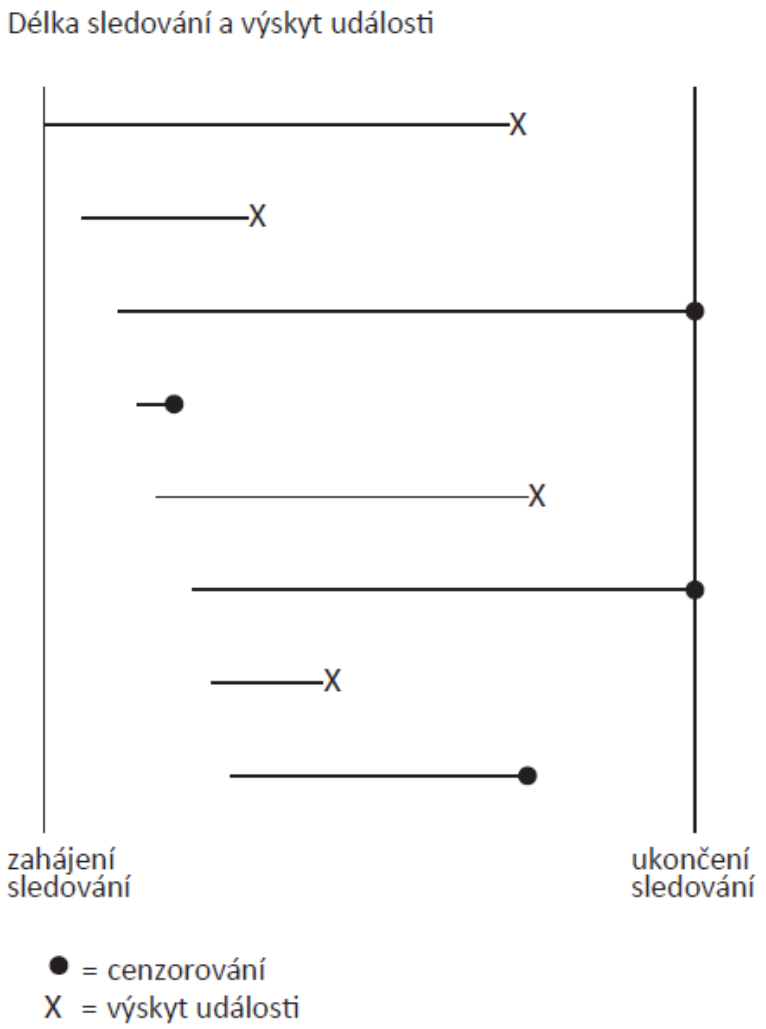
IBA LF MU, ÚZIS ČR

Analýza přežití

- **Hlavní charakteristikou odlišující data o přežití** (*survival data*) od ostatních typů dat, např. od klasického vnímání mortality jako podílu zemřelých pacientů v klinických aplikacích, je jejich **časová složka**.
- Data o přežití odráží nejen informaci o počtu, respektive podílu sledovaných událostí, ale zároveň nás také informují, kdy k dané události došlo.
- Analýza přežití zahrnuje matematicko-statistické metody pro hodnocení času do výskytu sledované události a vztah tohoto času s dalšími pozorovanými proměnnými.
- Klíčovým prvkem je **definice sledované události** (*event of interest*). Ta musí být stanovena jednoznačně a měla by být také snadno pozorovatelná či zjistitelná.

Cenzorování

- Definovaná událost se nemusí v průběhu sledování vyskytnout u všech subjektů = pozorování tedy není kompletní. Takový čas přežití označujeme jako cenzorovaný (*censored*).
- Subjekty bez sledované události nelze z hodnocení vyloučit, neboť nepřítomnost události je často pozitivní ukazatel.
- Příčiny cenzorování: ukončení sledování (např. uzavření databáze), ztráta kontaktu s pacientem, výskyt události, která výskyt námi definované události jednoznačně vylučuje: kompetitivní událost (*competing event*).
- **cenzorování zprava** (*right censoring*),
- **cenzorování zleva** (*left censoring*),
- **intervalové cenzorování** (*interval censoring*).



MUNI MED

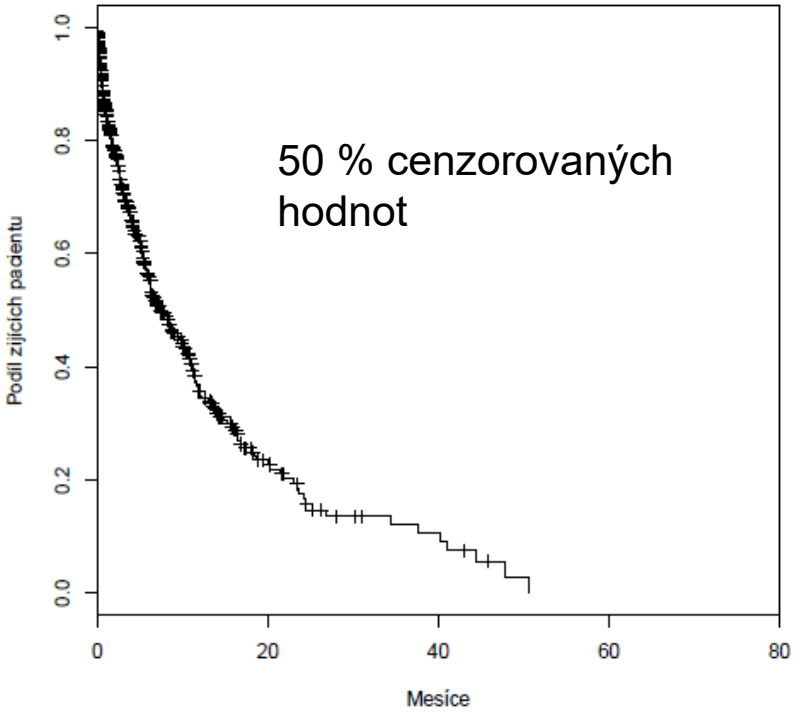
Funkce přežití

- **Čas přežití** (*survival time*) neboli dobu do výskytu sledované události reprezentujeme **nezápornou náhodnou veličinou *T***, která představuje buď skutečný nebo cenzorovaný čas přežití daného subjektu.
- **Funkce přežití** (*survival function*), označme ji $S(t)$, vyjadřuje pravděpodobnost, že se náhodná veličina T realizuje na reálné ose až za danou hodnotou t, což znamená pravděpodobnost, že čas přežití daného subjektu bude větší, než je zvolený čas t.

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$$

Vliv cenzorování na hodnocení přežití

– Cenzorování vnímáme jako ztrátu informace: procento cenzorovaných subjektů tak bývá měřítkem kvality sledování daného souboru, protože výrazně ovlivňuje kvalitu odhadů přežití.



MUNI MED

Význam délky sledování

- Cenzorování = ztráta informace, proto vysoké procento cenzorovaných časů přežití může vypovídat o nedostatečné délce sledování, kdy jsme nebyli schopni pozorovat dostatečné množství událostí, které hodnotíme.
- Cenzorování tak má vliv na požadovanou velikost souboru hodnocených subjektů, protože cenzorované časy nejsou rovnocenné s kompletními časy přežití a v přítomnosti cenzorování je nutné navýšit velikost souboru tak, abychom zajistili dostatečný počet kompletních pozorování.
- Předpoklad nezávislosti cenzorování a výskytu sledované události, tedy tzv. neinformativního cenzorování (non-informative censoring).

Odhad funkce přežití

- Při odhadu pravděpodobnosti přežití časů t_i je třeba zohlednit cenzorování. Cenzorované časy přežití totiž nelze hodnotit stejně jako kompletní pozorování, neboť nepřispívají k počtu událostí (d_i), ale zároveň je nelze z hodnocení vyřadit.

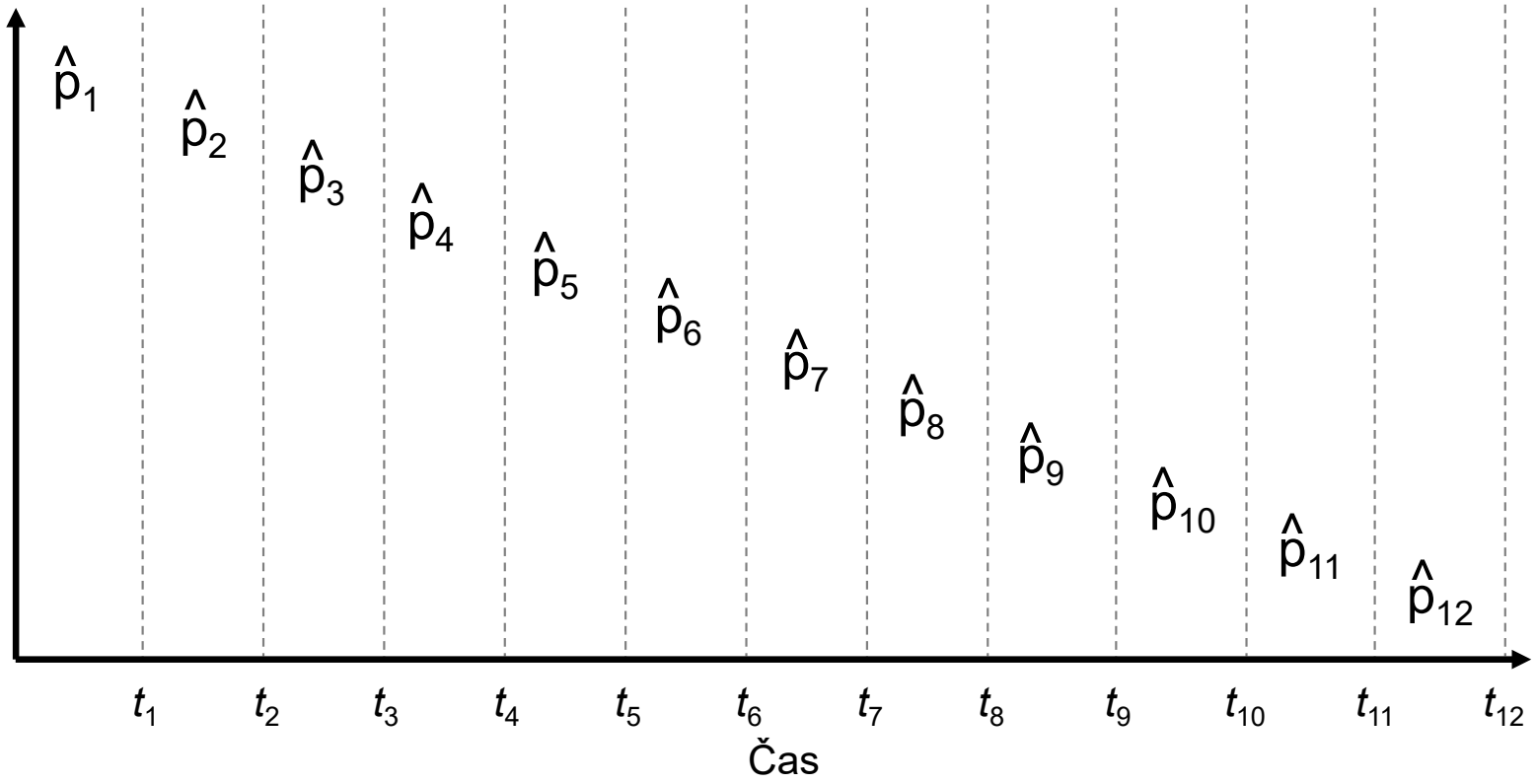
- **Parametrické metody** (*parametric survival analysis*) vyžadují specifikaci konkrétního rozdělení náhodné veličiny T.

- **Neparametrické metody** (*nonparametric survival analysis*) žádné zvláštní předpoklady ohledně rozdělení pravděpodobnosti náhodné veličiny T nevyžadují.

Kaplanův-Meierův odhad funkce přežití

- Nejpoužívanější odhad funkce přežití, standard pro hodnocení přežití v klinických studiích.
- Aby byl subjekt v čase *t* bez události (např. aby se pacient s nádorovým onemocněním dožil času *t*), nesmí se u něj událost vyskytnout v žádném čase t_i takovém, pro nějž platí, že $t_i < t$.

$p_1 - p_{12}$ označujeme jako intervalově specifické odhady



MUNI MED

Kaplanův-Meierův odhad funkce přežití

- Odhad pro detailní individuální data.
- Kaplanův-Meierův odhad pracuje s cenzorováním tak, že tato pozorování vypadávají ze skupiny subjektů v riziku ihned po zaznamenaném čase cenzorování.
- Kaplanův-Meierův odhad funkce přežití:

$$\hat{S}(t) = \prod_{t_i \le t} \hat{p}_i = \prod_{t_i \le t} \left(1 - \frac{d_i}{R_i}\right)$$

Počet sledovaných událostí v čase t_i

Počet subjektů v riziku sledované události v čase t_i

Intervalově specifický odhad přežití od času t_{i-1} do času t_i

MUNI MED

Metoda úmrtnostních tabulek (*life-table*)

- Odhad pro agregovaná populační data.
- Hlavní myšlenka odhadu zůstává stejná. Na rozdíl od Kaplanova-Meierova odhadu, kde byly časové intervaly určeny pozorovanými hodnotami časů přežití, zde pracujeme s předem definovanou sadou (většinou ekvidistantních a delších) *J* časových intervalů.
- Pracujeme s delšími časovými intervaly, pro odhad S(t) nám stačí pouze agregovaná data, tedy souhrnné údaje pro jednotlivé časové intervaly:

$$\hat{S}(J) = \prod_{j \le J} \hat{p}_j = \prod_{j \le J} \left(1 - \frac{d_j}{R_j - \frac{c_j}{2}} \right)$$

Počet sledovaných událostí v *j*-tém intervalu

Počet cenzorovaných časů v *j*-tém intervalu

Intervalově specifický odhad přežití pro *j*-tý interval

Počet subjektů v riziku na začátku *j*-tého intervalu

MUNI MED

Relativní přežití

- Chceme-li kvantifikovat mortalitu spojenou pouze se sledovanou diagnózou, musíme zohlednit pouze **vybranou příčinu** úmrtí.
- Jak?

1. Výpočet **přežití specifického dle diagnózy** (*cause-specific survival*)
2. Výpočet **relativního přežití** (*relative survival*)

Relativní přežití

- **Poměr celkového přežití a tzv. očekávaného přežití**, které vyjadřuje mortalitu v obecné populaci odpovídající sledované skupině pacientů věkem a pohlavím.
- Cílem výpočtu je **odfiltrovat mortalitu spojenou s dalšími diagnózami**. Relativní přežití odráží pouze mortalitu spojenou se sledovaným onemocněním.

$$R(t) = \frac{S(t)}{S^*(t)}$$

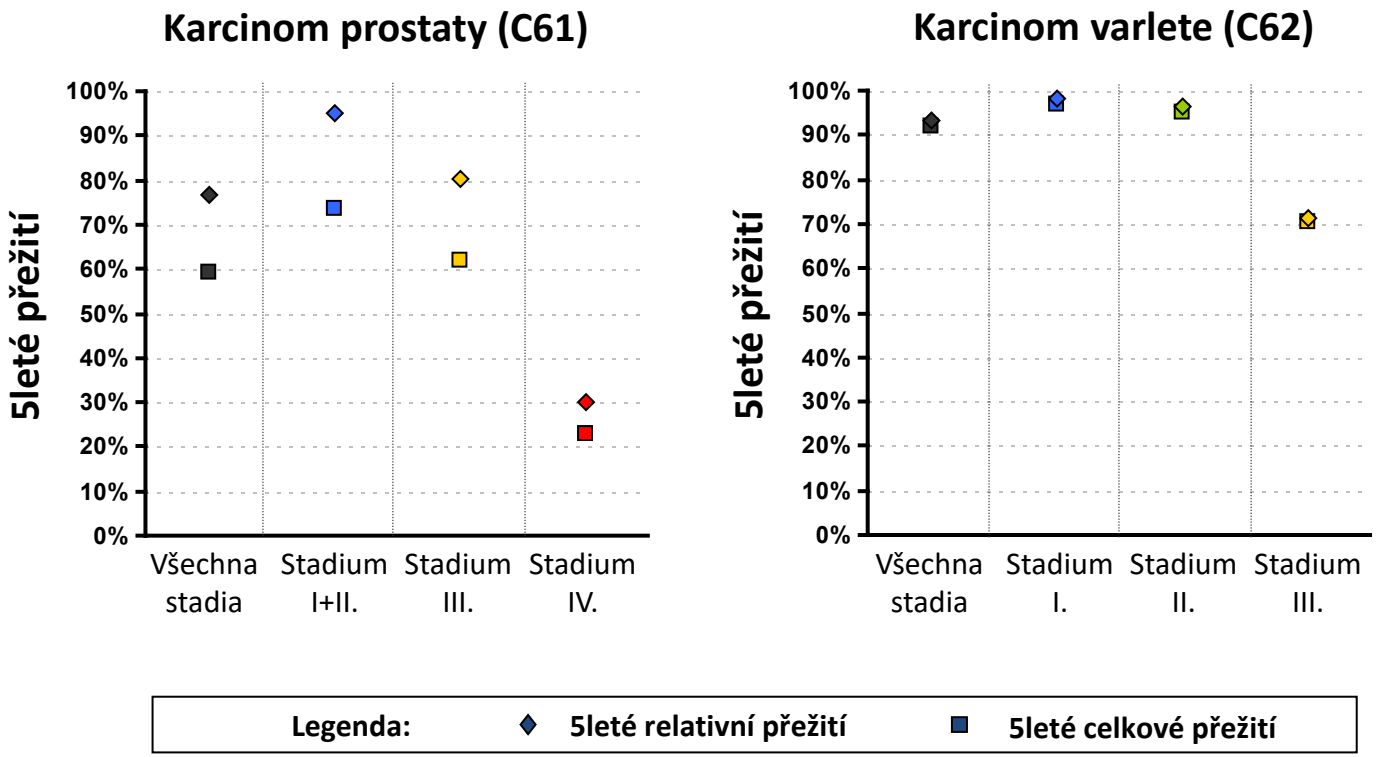
celkové přežití

očekávané přežití

MUNI MED

Relativní přežití u různých diagnóz

– Období 2005-2009



MUNI MED

Metody odhadu očekávaného přežití

- Principem výpočtu očekávaného přežití je výpočet přežití z populačních mortalitních tabulek odpovídajícího srovnatelné skupině (vzhledem k věku, pohlaví a roku diagnózy) z obecné populace, u které předpokládáme, že prakticky není zasažena sledovaným onemocněním.
- Metody výpočtu očekávaného přežití:

1. **Edererova metoda I**
2. **Edererova metoda II**
3. **Hakulinenova metoda**
4. **Metoda dle Pohar-Perme**

- Všechny metody vycházejí ze stejných datových podkladů, tedy úmrtnostních tabulek pro danou populaci (stát). Mezinárodní zdroj: www.mortality.org.

Statistické vyléčení

- **Když se odhady intervalově specifického relativního přežití „přiblíží“ hodnotě 1**, lze říci, že se mortalita sledovaných pacientů v daném intervalu dostala na úroveň mortality populační. V tomto případě pak mluvíme o tzv. **statistickém vyléčení**.
- Nelze zaměňovat pojmy klinické a statistické vyléčení:
  - Pojem **klinické vyléčení** chápeme na úrovní jedince jako vymizení všech klinických projevů nemoci.
  - Pojem **statistické vyléčení** chápeme na úrovni skupiny pacientů jako srovnání mortality s populační úrovní.

MUNI MED

Specifika populačního hodnocení přežití

- Tendence vztahovat výsledky k léčbě
- Výrazný vliv demografických (věk) a klinických (stadium onemocnění) charakteristik
- Tendence analyzovat nejaktuálnější data

Specifika populačního hodnocení přežití

- Tendence vztahovat výsledky k léčbě

  → Nutná správná definice analyzovaného souboru

- Výrazný vliv demografických (věk) a klinických (stadium onemocnění) charakteristik

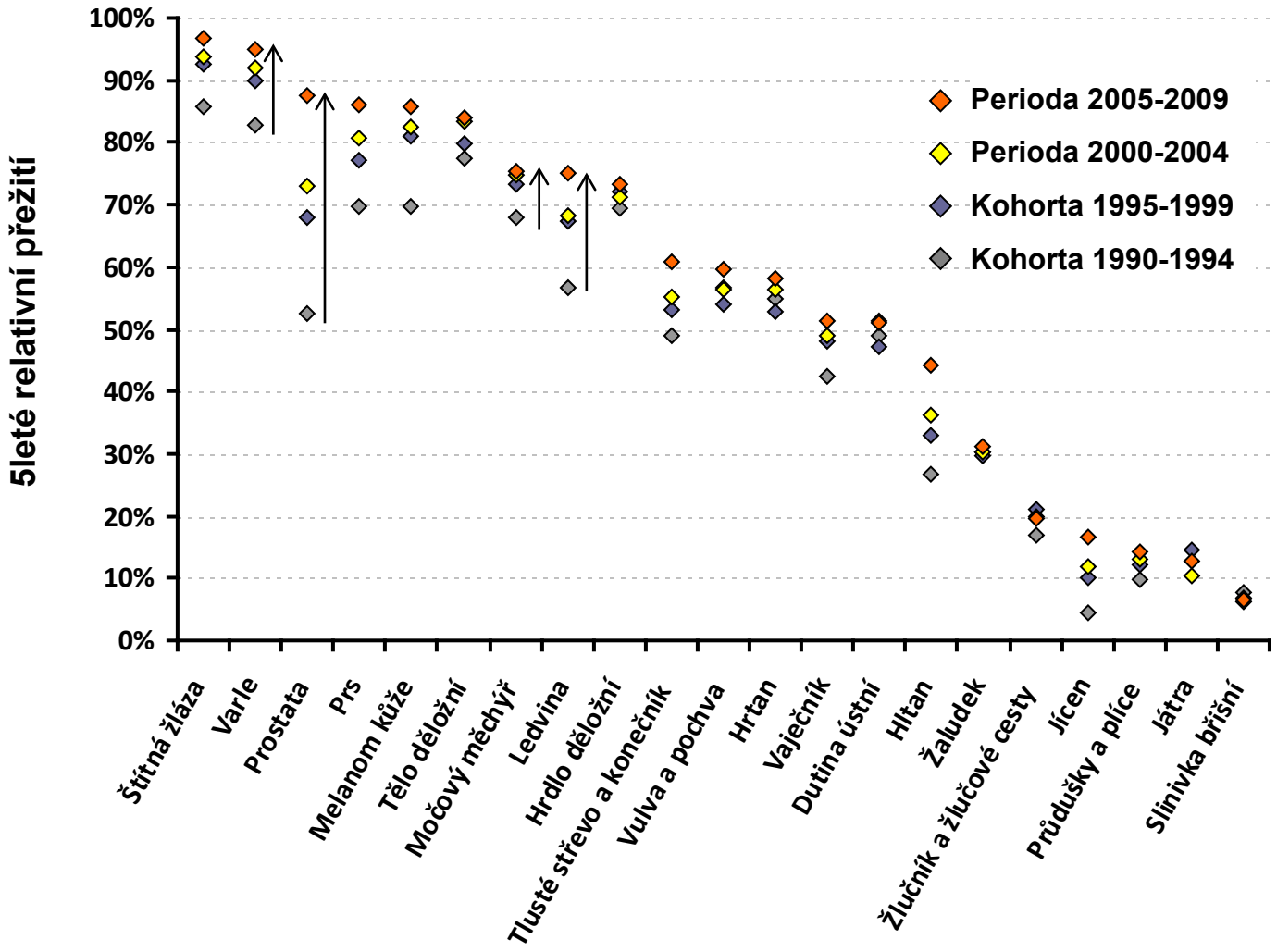
  → Nutná standardizace výsledků

- Tendence analyzovat nejaktuálnější data

  → Nutná správná volba analyzovaného období

Příklad

– **Vidíme velká zlepšení, která ale mohou souviset se změnou struktury pacientů**



MUNI MED

Věková standardizace

- Nutná kvůli **srovnatelnosti výpočtu v různých populacích pacientů**
- Výpočet relativního přežití v několika věkových kategoriích
- Vážení dílčích odhadů váhami příslušných věkových kategorií

- Klíčovým prvkem je definice vah

1. Dle věkové struktury onkologických pacientů v ČR
2. Dle věkové struktury onkologických pacientů v Evropě/ve světě

Věková standardizace

– Nutná kvůli **srovnatelnosti výpočtu v různých populacích pacientů**

Populace		**Celkem**
2000 - 2004	Subjekty v riziku	1 000
	5leté přežití	**37,5 %**
2005 - 2009	Subjekty v riziku	1 000
	5leté přežití	**37,5 %**
2010 - 2014	Subjekty v riziku	1 000
	5leté přežití	**37,5 %**

Stejné 5leté přežití v rámci všech tří období

MUNI MED

Věková standardizace

– Nutná kvůli **srovnatelnosti výpočtu v různých populacích pacientů**

Populace		**Celkem**	**do 50 let**	**50-69 let**	**70 a více let**
2000	Subjekty v riziku	1 000	200	300	500
- 2004			20,0%	30,0%	50,0%
	5leté přežití	**37,5 %**	**50,0%**	**40,0%**	**31,0%**
2005	Subjekty v riziku	1 000	100	250	650
- 2009			10,0%	25,0%	65,0%
	5leté přežití	**37,5 %**	**53,0%**	**43,0%**	**33,0%**
2010	Subjekty v riziku	1 000	50	150	800
- 2014			5,0%	15,0%	80,0%
	5leté přežití	**37,5 %**	**55,0%**	**45,0%**	**35,0%**

Věková standardizace

– Nutná kvůli **srovnatelnosti výpočtu v různých populacích pacientů**

Populace	5leté přežití	do 50 let	50-69 let	70 a více let
2000 – 2004	**33,3%**	200 20,0% **50,0%**	300 30,0% **40,0%**	500 50,0% **31,0%**
2005 – 2009	**35,5%**	100 10,0% **53,0%**	250 25,0% **43,0%**	650 65,0% **33,0%**
2010 – 2014	**37,5%**	50 **5,0%** **55,0%**	150 **15,0%** **45,0%**	800 **80,0%** **35,0%**

MUNI MED

Věková standardizace

– Nutná kvůli **srovnatelnosti výpočtu v různých populacích pacientů**

Populace	5leté přežití	do 50 let	50-69 let	70 a více let
		200	300	500
		20,0%	[illegible]	50,0%
2000 – 2004	**33,3%**	**50,0**	[illegible]	**31,0%**
		10[illegible]	[illegible]	650
		10,0[illegible]	[illegible]	65,0%
2005 – 2009	**35,5%**	**53,0%**	[illegible]	**33,0%**
		50	150	800
		5,0%	15,0%	80,0%
2010 – 2014	**37,5%**	**55,0%**	**45,0%**	**35,0%**

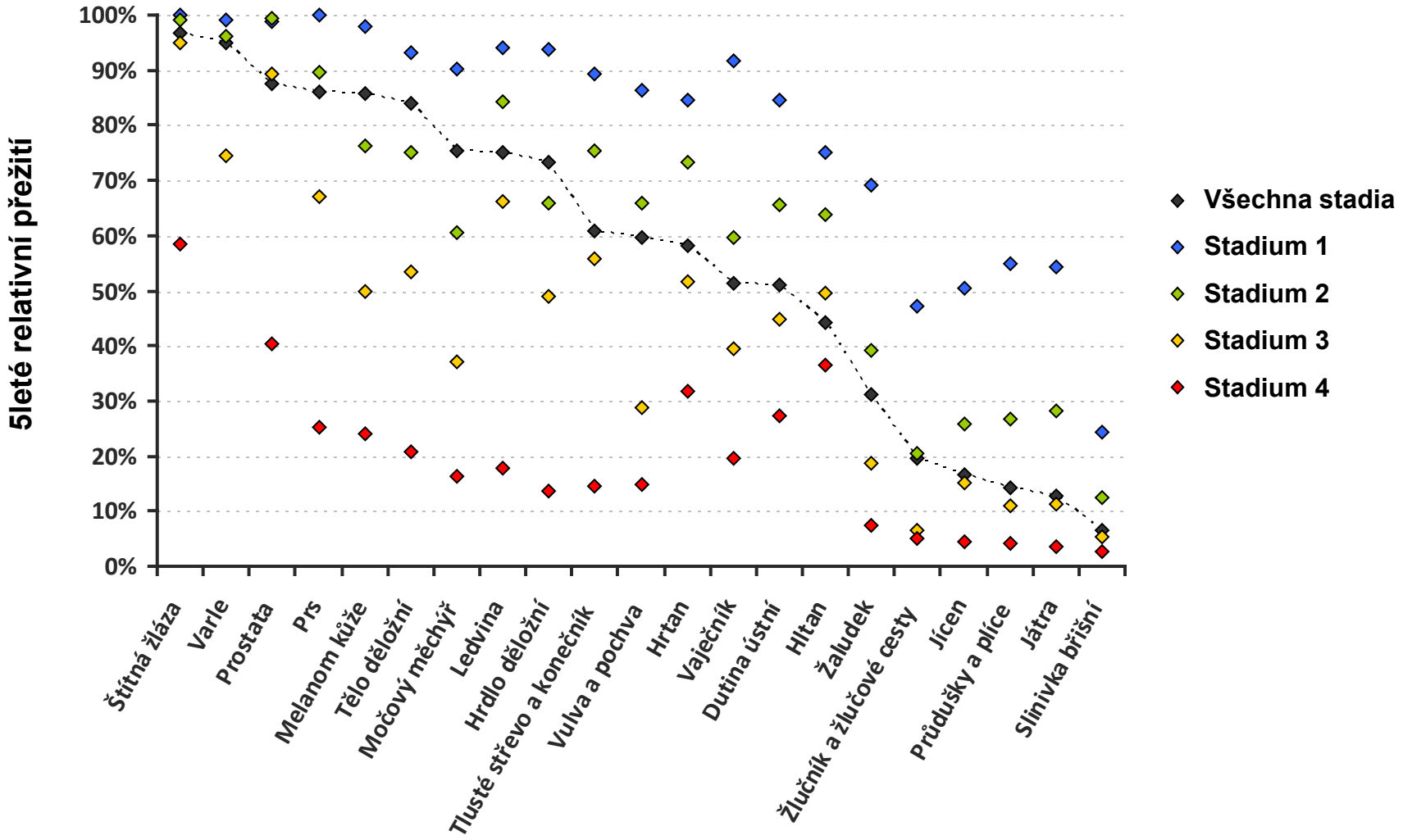
Zlepšení 5letého přežití v rámci tří období

MUNI MED

Vliv stadia

Analýza periody 2005-2009

– **Stadium onemocnění má v onkologii často větší vliv na výsledek léčby než věk pacienta**



Zdroj: Národní onkologický registr, ÚZIS ČR

MUNI MED

Časové období pro hodnocení přežití

- **Analýza kohorty** – zahrnuje pouze pacienty, kteří od data diagnózy mohli dosáhnout *x*-letého přežití. Všichni pacienti tedy mají ukončenou dobu sledování *x* let od diagnózy. Nemusí odrážet aktuální trendy.
- **Analýza periody** – zahrnuje pacienty, kteří se dožili definovaného období. Odráží aktuální epidemiologické trendy. Hodnotíme pouze pacienty, kteří se dožili recentního období.

- - - - - Kohortová analýza ——— Analýza časové periody

	Kalendářní rok sledování pacienta										
Kalendářní rok diagnózy	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
2000	1	1 / 2	2 / 3	3 / 4	4 / 5	5					
2001		1	1 / 2	2 / 3	3 / 4	4 / 5	5				
2002			1	1 / 2	2 / 3	3 / 4	4 / 5	5			
2003				1	1 / 2	2 / 3	3 / 4	4 / 5	5		
2004					1	1 / 2	2 / 3	3 / 4	4 / 5	5	
2005						1	1 / 2	2 / 3	3 / 4	4 / 5	5
2006							1	1 / 2	2 / 3	3 / 4	4 / 5
2007								1	1 / 2	2 / 3	3 / 4
2008									1	1 / 2	2 / 3
2009										1	1 / 2
2010											1

. pacienti diagnostikovaní a sledovaní v daných letech poskytují informaci o x-letém a y-letém přežití

MUNI MED

Definice souboru pacientů v NOR

- **Dospělí pacienti** – nádory v dětském věku nelze z hlediska hodnot přežití označit za srovnatelné s nádory dospělých.
- **Nenulová hodnota času přežití** – záznamy o pacientech s nulovou hodnotou přežití logicky nevedou k věcně relevantní informaci o přežití (jedná se zejména o pacienty diagnostikované jako DCO nebo na základě pitvy).
- **Pacienti s prvním zhoubným novotvarem** – pacienti s druhým a dalším novotvarem mohou zkreslovat pohled na přežití v souvislosti s primární protinádorovou léčbou.
- **Všichni pacienti** jako epidemiologicky reprezentativní soubor vs. **pacienti s protinádorovou léčbou** jsou reprezentativní vzhledem k hodnocení kvality léčebné péče v ČR.

MUNI
MED

Workshop: Populační hodnocení přežití onkologických pacientů

T. Pavlík, A. Šikyňová, Z. Bortlíček

IBA LF MU, ÚZIS ČR

Pokyny před workshopem

- Jste rozdělení do skupin a každá skupina má přiřazenu jednu z onkologických diagnóz, na kterou se v průběhu workshopu zaměří
  - C18–C20: Zhoubný novotvar kolorekta
  - C25: Zhoubný novotvar slinivky břišní
  - C34: Zhoubný novotvar průdušky a plíce
  - C50: Zhoubný novotvar prsu
  - C53: Zhoubný novotvar hrdla děložního
  - C61: Zhoubný novotvar prostaty
- Nainstalovat si R, verze 4.2.2 (či vyšší)
- Převzít a připravit si potřebné datové vstupy (flash disk)

Cíle workshopu – odpovědět na otázky

1. Lze říci, že se přežití vybrané diagnózy mezi obdobími 2005-2009 a 2010-2014 zlepšuje? Čemu lze změny v přežití přisoudit?
2. Lze říci, že zlepšení v přežití souvisí/koreluje se zkvalitněním léčby?
3. Můžeme u vybrané diagnózy mluvit v horizontu 5 až 10 let o statistickém vyléčení? Je rozdíl mezi stadii onemocnění?
4. Existují statisticky významné rozdíly v přežití na onkologickou diagnózu dle toho, zda se u pacienta jedná o první nebo druhý (a další) ZN?
5. Čemu přisuzujete výraznou změnu v 5letém celkovém přežití u pacientů mezi lety 2010-2014 a 2015-2021.

Odhad funkce přežití v R

Funkce `survfit()`, která je součásti knihovny *survival*

```
survfit(Surv(time, status) ~ x, data =…, weights = …, …)
```

- `time` - časy událostí nebo cenzorovaní
- `status` - indikátor cenzorování
- `x` - proměnná charakterizující skupinu (v případě odhadu pro celá data x=1)
- `type` - specifikuje typ odhadu funkce přežití `("kaplan-meier","fleming-harrington", "fh2")`
- `conf.type` - specifikuje nastavení pro interval spolehlivosti

Testování rovnosti křivek přežití

```
survdiff(Surv(time, status) ~ x, rho = 0)
```

- rho - specifikuje typ testu: 0 - Coxův–Mantelův test (log-rank test) , 1 - Gehan–Wilcoxonův test

MUNI MED

Odhad funkce relativního přežití v R

Funkce `rs.surv()`, která je součásti knihovny *relsurv*

```
rs.surv(Surv(time, status) ~ x, data =…, ratetable = …, rmap = …, …)
```

Argumenty funkcie analogické funkcii `survfit`

- `method` - specifikuje metodu odhadu funkce relativního přežití (`"ederer1", "erderer2", "pohar-perme", "hakulinen"`)
- `ratetable` – tabulka miery výskytu události (ve formátu retetable)
- `rmap = list(age = VekDg*365.25, year = RokDg, sex = Pohlavi)`
  namapováni názvů proměnných v datech a názvu tabulky `ratetable`

MUNI MED

Výstupy objektů fit a summary(fit)

- `$time` - časy událostí nebo cenzorovaní (v případě fit), nebo časy událostí (v případě `summary`(fit))
- `$n` - počet pozorovaní
- `$n.risk` - počet jedinců v riziku v odpovídajících časech `$time`
- `$n.event` - počet událostí v odpovídajících časech `$time`
- `$n.censor` - počet cenzorování v odpovídajících časech `$time`
- `$surv` - odhad funkce přežití v odpovídajících časech `$time`
- `$std.err` - odhad směrodatné odchylky kumulativní rizikové funkce (v případě fit), nebo funkce přežití (v případě `summary`(fit)) v odpovídajících časech `$time`
- `$lower` - dolní hranice IS pro funkci přežití v odpovídajících časech `$time`
- `$upper` - horní hranice IS pro funkci přežití v odpovídajících časech `$time`

MUNI
MED

Prezentace výsledků workshopu

- Odpovědět na otázky v kontextu dané diagnózy…
- Zapojte fantazii