

Zápočtová úloha

Vypracoval: Radovan Zeman
Datum: 7. listopadu 2019

1 Popis dat

Data obsahují 135 diagnostikovaných pacientů a veličiny survival time (čas přežití, ve dnech), treat (léčba, standard nebo placebo), cell (typ buněk, squamous, small, large nebo adeno), cens (ukončení léčby, died nebo censored), Karnofsky score (KAR, stupeň hospitalizaci), disease duration (doba od diagnózy, v měsících), age (věk v letech), prith (předchozí léčba, none nebo some).

V tab. 1 jsou základní statistiky numerických veličin, v tab. 2 pak četnosti jednotlivých kategorií. Časy přežití pacientů jsou zobrazeny na obr. 1, jsou zde rozlišeny typy buněk, léčba a cenzorování. Na obr. 2 jsou krabicové diagramy veličin cell a treat. Obr. 3 a 4 znázorňují závislost numerických proměnných.

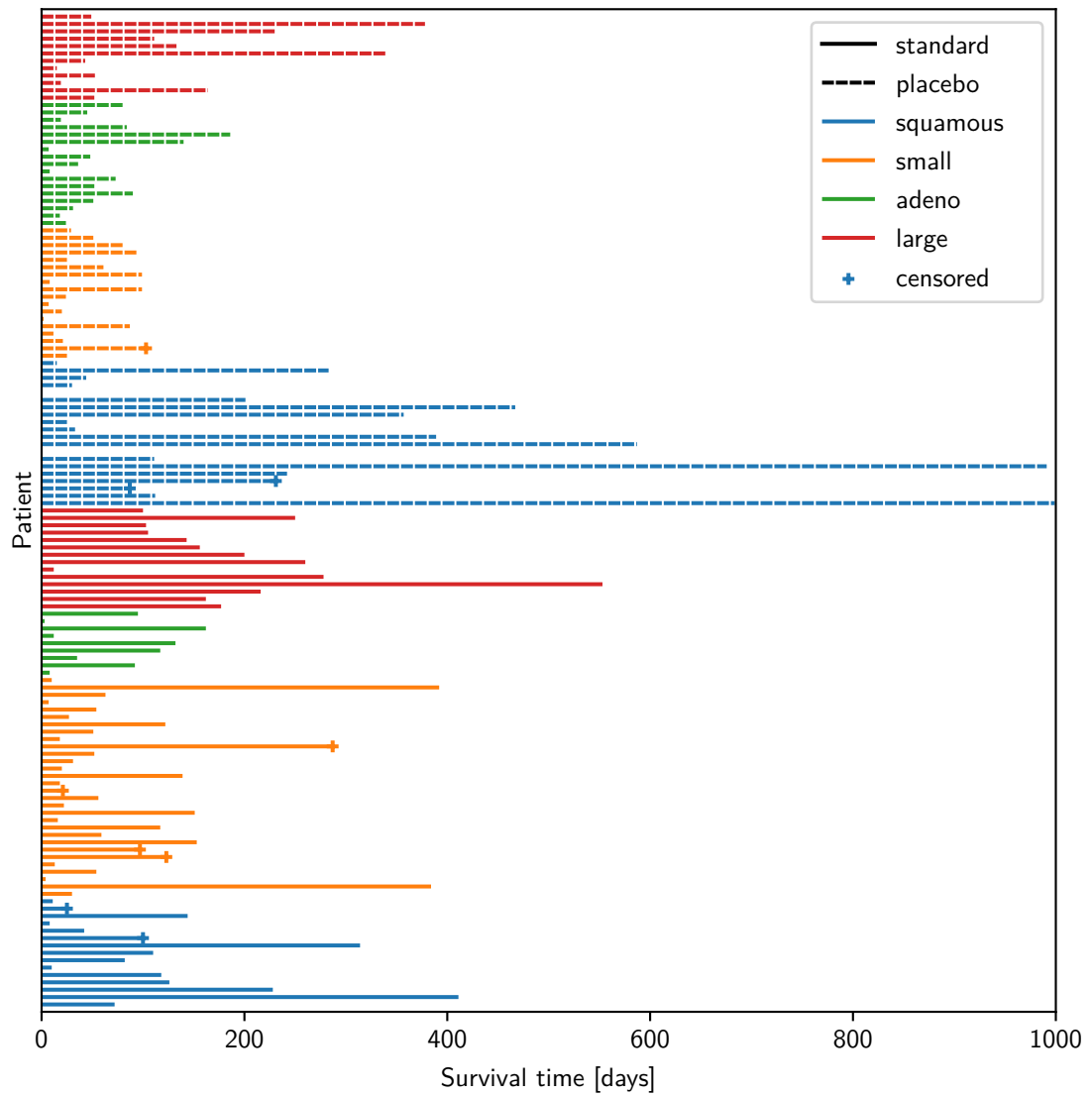
Z grafů vyplývá, že zatímco u typů large, adeno a small léčba zvýšila střední dobu dožití, v případě typu squamous se v průměru déle dožívali uživatelé placeba. Přesto jsou u typu squamous vidět nejdelší doby přežití. Naopak typy small a adeno je mají nejkratší bez ohledu na léčbu.

Statistika	Survival time [dny]	KAR [-]	Disease duration [měsíce]	Age [roky]
Průměr	121,5	58,0	8,7	58,3
Směrodatná odchylka	158,7	19,7	10,7	10,6
Minimum	1,0	10,0	1,0	34,0
Maximum	999,0	90,0	87,0	81,0
1. kvartil	24,5	40,0	3,0	50,5
3. kvartil	143,5	70,0	11,0	66,0

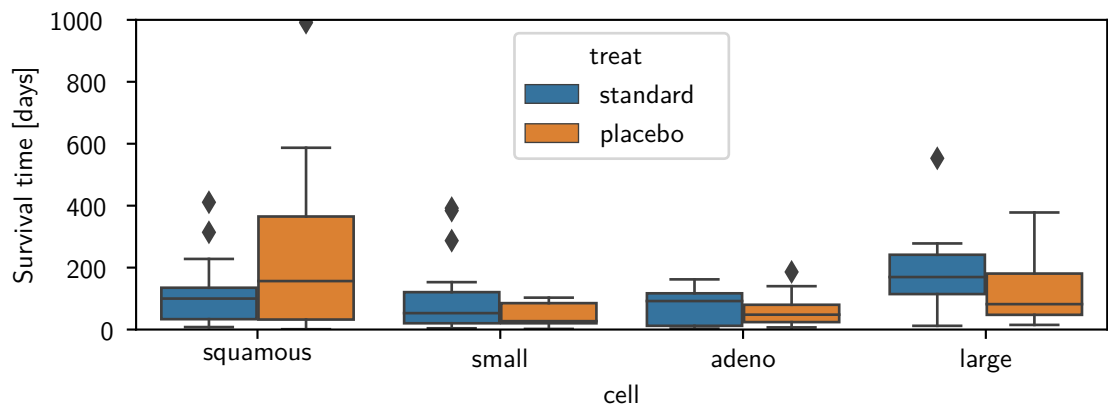
Tabulka 1: Základní statistiky numerických veličin.

Skupina		Četnost
celkem		135
cell	small	48
	squamous	35
	large	26
	adeno	26
treat	standard	68
	placebo	67
cens	died	126
	censored	9

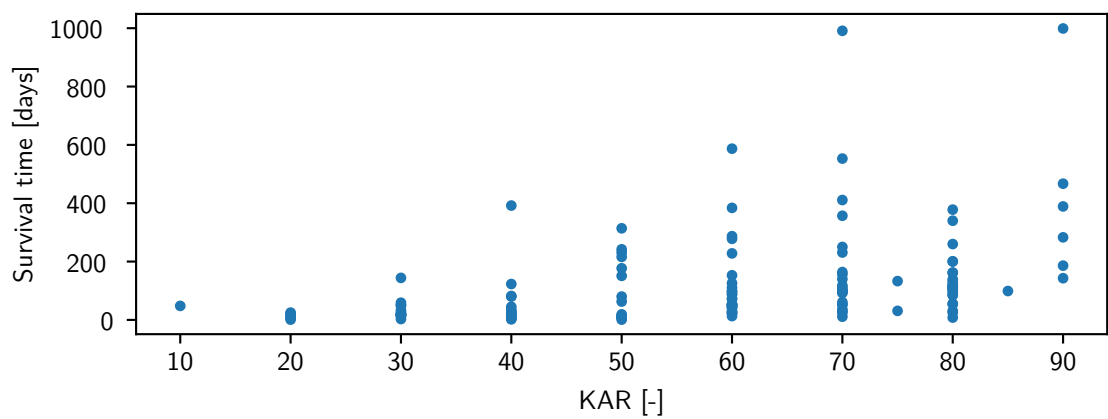
Tabulka 2: Četnosti faktorových veličin.



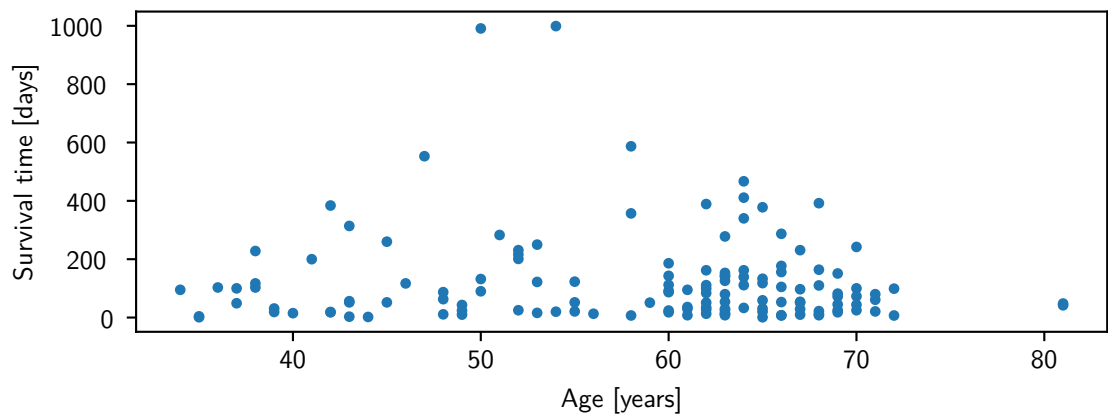
Obrázek 1: Časy přežití pacientů s rozlišením léčby, typu a cenzorování.



Obrázek 2: Krabicové diagramy času přežití dle léčby a typu.



Obrázek 3: Závislost doby přežití na Karnofsky score.



Obrázek 4: Závislost doby přežití na věku.

2 Srovnání doby přežití skupin pacientů

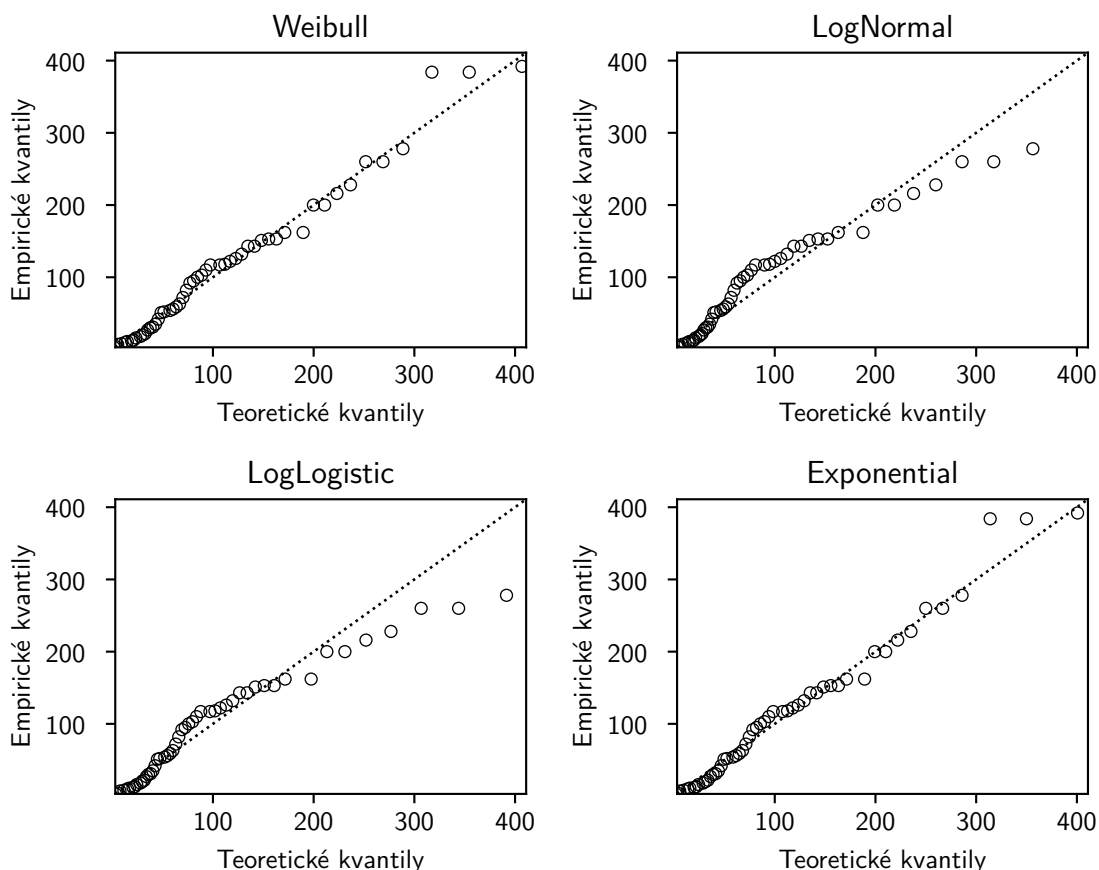
2.1 Postup

Pro srovnání různých skupin byl pro každou z nich sestavený model přežití pomocí neparametrických i parametrických metod.

Jako neparametrické metody byly použity Kaplanův–Meierův odhad k určení funkce přežití a Nelsonův–Aalenův odhad k učení kumulativní rizikové funkce.

Z parametrických rozdělení byly uvažovány čtyři distribuce – Weibullovo, exponenciální, log-normální a log-logistické. Pro každé z nich byly vždy odhadnuty parametry metodou maximální věrohodnosti, nejvhodnější model pro danou skupinu byl následně vybrán na základě kritéria AIC, grafů a kvantilových diagramů zobrazujících závislost teoretických kvantilů daného modelu a empirických určených z Kaplanova–Meierova odhadu. Příklad kvantilového diagramu pro prvního zkoumanou skupinu je na obr. 5. Vhodnost vybraných modelů je patrná i srovnáním grafů funkce přežití – parametrickou s Kaplanovým–Meierovým odhadem. Parametrizace použitých rozdělení jsou uvedeny v tab. 8 v dodatku A.

Pomocí každé ze tří použitých metod (Kaplan–Meier, Nelson–Aalen, parametrický model) byly spočteny střední a mediánové doby přežití.



Obrázek 5: Kvantilové diagramy pro výběr parametrických modelů (standardní léčba).

2.2 Srovnání standardní léčby a placebo

Skupina	Model	Mean surv. t.	Median surv. t.
standard	Kaplan–Meier	124,9	103,0
	Nelson–Aalen	133,3	103,0
	exponenciální, $\lambda = 125,2$	125,2	86,8
placebo	Kaplan–Meier	139,8	52,0
	Nelson–Aalen	146,7	52,0
	log-logistický, $\alpha = 59,5; \beta = 1,2$	283,2	59,5

Tabulka 3: Odhadnuté parametry modelů a střední a mediánové doby dožití pro dané skupiny.

- Je patrný rozpor mezi středními a mediánovými dobami přežití: zatímco střední hodnoty jsou vyšší u placebo, mediánové u standardní léčby. To je způsobeno sloučením velmi rozdílných podskupin – zřejmě typů buněk, kde u typu squamous byly v případě placebo pozorovány velmi dlouhé doby přežití zvyšující střední hodnotu, zatímco doby dožití ostatních typů jsou nižší. Bez rozlišení typů nemá smysl posuzovat.

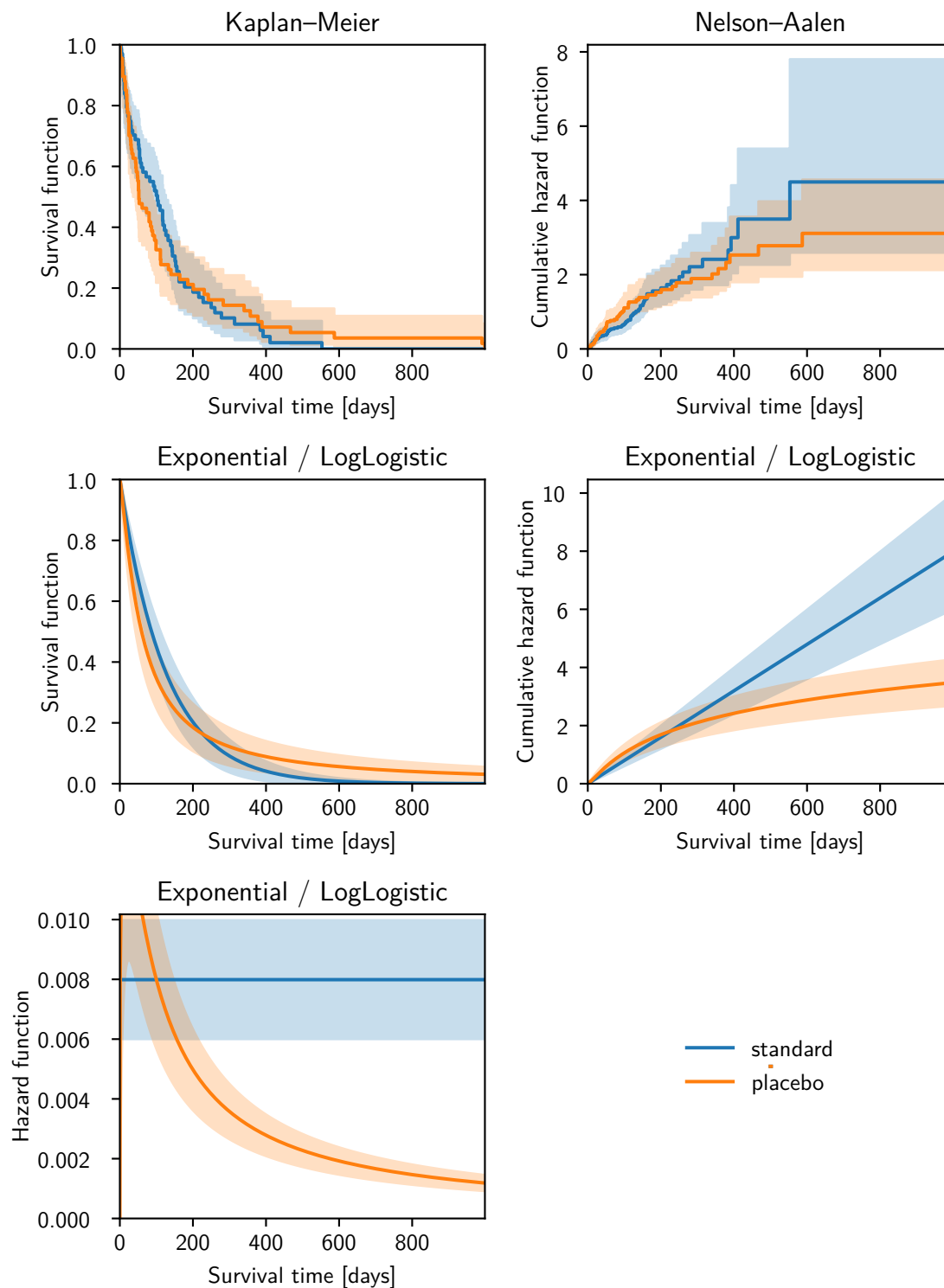
2.3 Srovnání různých typů buněk

Skupina	Model	Mean surv. t.	Median surv. t.
adeno	Kaplan–Meier	63,4	48,0
	Nelson–Aalen	82,5	51,0
	exponenciální, $\lambda = 63,4$	63,4	43,9
large	Kaplan–Meier	165,5	143,0
	Nelson–Aalen	182,0	156,0
	Weibull, $\lambda = 179,7; \rho = 1,3$	165,3	136,3
small	Kaplan–Meier	83,9	51,0
	Nelson–Aalen	101,2	51,0
	log-normální, $\mu = 3,8; \sigma = 1,2$	91,9	43,6
squamous	Kaplan–Meier	230,2	118,0
	Nelson–Aalen	243,0	118,0
	Weibull, $\lambda = 202,5; \rho = 0,8$	235,2	126,0

Tabulka 4: Odhadnuté parametry modelů a střední a mediánové doby dožití pro dané skupiny.

- Jednotlivé typy buněk způsobují rozdílné parametrické modely pro přežití pacientů.
- Zatímco typ adeno má konstantní riziko, typy squamous a small jej mají klesající a naopak large jej má rostoucí.

Srovnání standardní léčby a placeba



Obrázek 6: Odhady funkce přežití, rizikové funkce a kumulativní rizikové funkce pro dané skupiny – Kaplanův–Meierův, Nelsonův–Aalenův a parametrické modely.

- Typ adeno má téměř po celou dobu nejvyšší riziko, tomu odpovídají nízké střední a mediánové doby přežití. Odhady potvrzují pozorování patrné na obr. 1.
- Nejnižší riziko a zároveň nejvyšší funkce přežití odpovídá typu squamous, u kterého jsou vysoké doby přežití (střední, mediánové).

2.4 Srovnání věkových skupin

Skupina	Model	Mean surv. t.	Median surv. t.
≤ 60	Kaplan–Meier	168,5	92,0
	Nelson–Aalen	176,8	95,0
	Weibull, $\lambda = 142,5; \rho = 0,8$	165,4	88,8
> 60	Kaplan–Meier	107,4	61,0
	Nelson–Aalen	114,2	72,0
	exponenciální, $\lambda = 107,1$	107,1	74,2

Tabulka 5: Odhadnuté parametry modelů a střední a mediánové doby dožití pro dané skupiny.

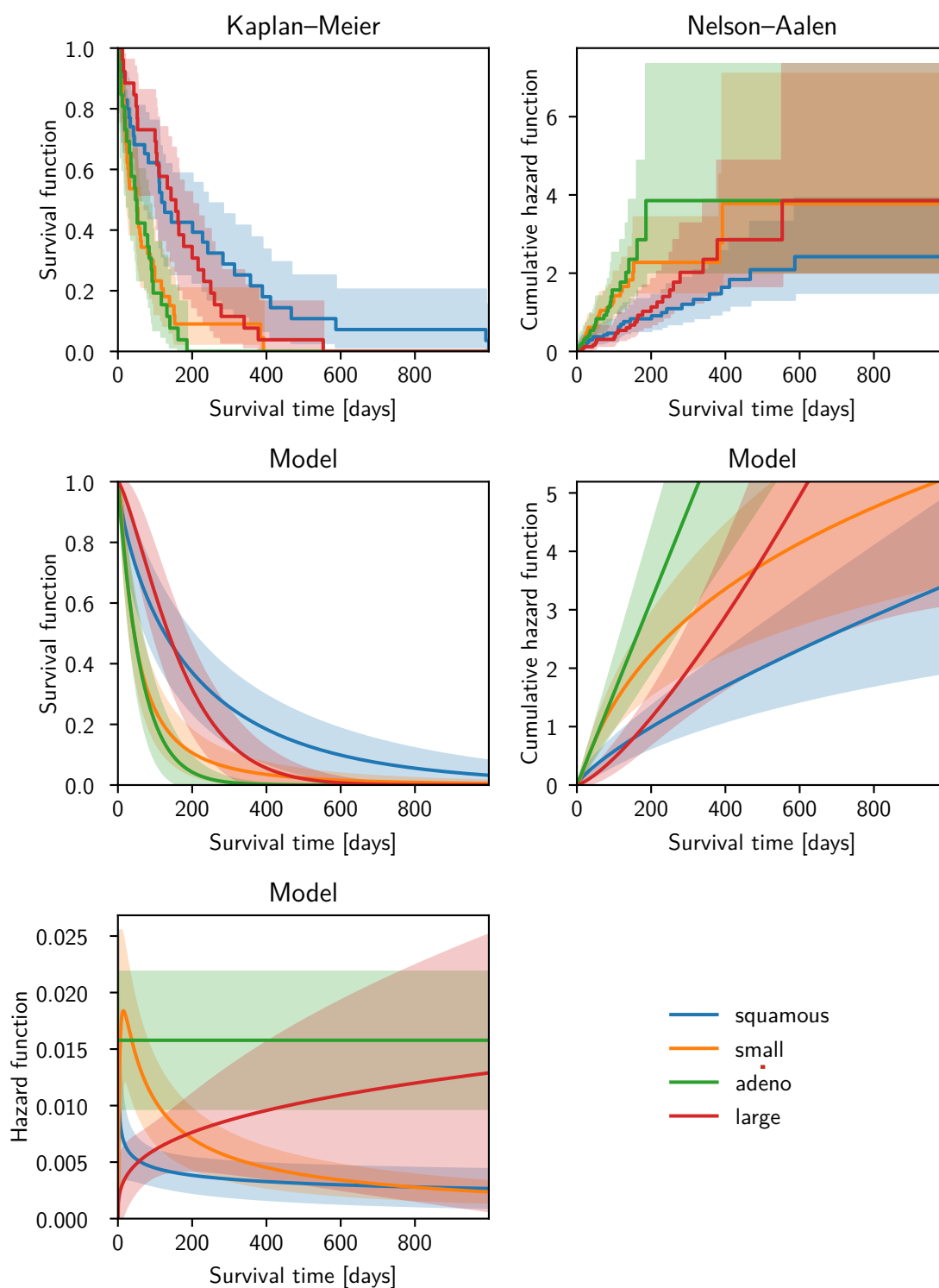
- Věková skupina nad 60 let s konstantním rizikem má téměř po celou dobu vyšší riziko než skupina do 60 let s klesající rizikovou funkcí. Tomu odpovídá funkce přežití, která je vyšší pro mladší pacienty.
- Také střední a mediánové doby přežití mají mladší pacienti vyšší (všechny odhady).

2.5 Srovnání vybraných typů buněk a léčby

Skupina	Model	Mean surv. t.	Median surv. t.
squamous + standard treat	Kaplan–Meier	136,8	110,0
	Nelson–Aalen	173,4	118,0
	exponenciální, $\lambda = 138,4$	138,4	95,9
squamous + placebo	Kaplan–Meier	293,4	201,0
	Nelson–Aalen	313,2	242,0
	Weibull, $\lambda = 248,7; \rho = 0,7$	312,9	147,8
small + standard treat	Kaplan–Meier	103,0	54,0
	Nelson–Aalen	129,0	54,0
	log-normální, $\mu = 4,0; \sigma = 1,2$	114,8	53,1
small + placebo	Kaplan–Meier		29,0
	Nelson–Aalen	122,2	29,0
	Weibull, $\lambda = 51,8; \rho = 1,2$	49,3	37,6

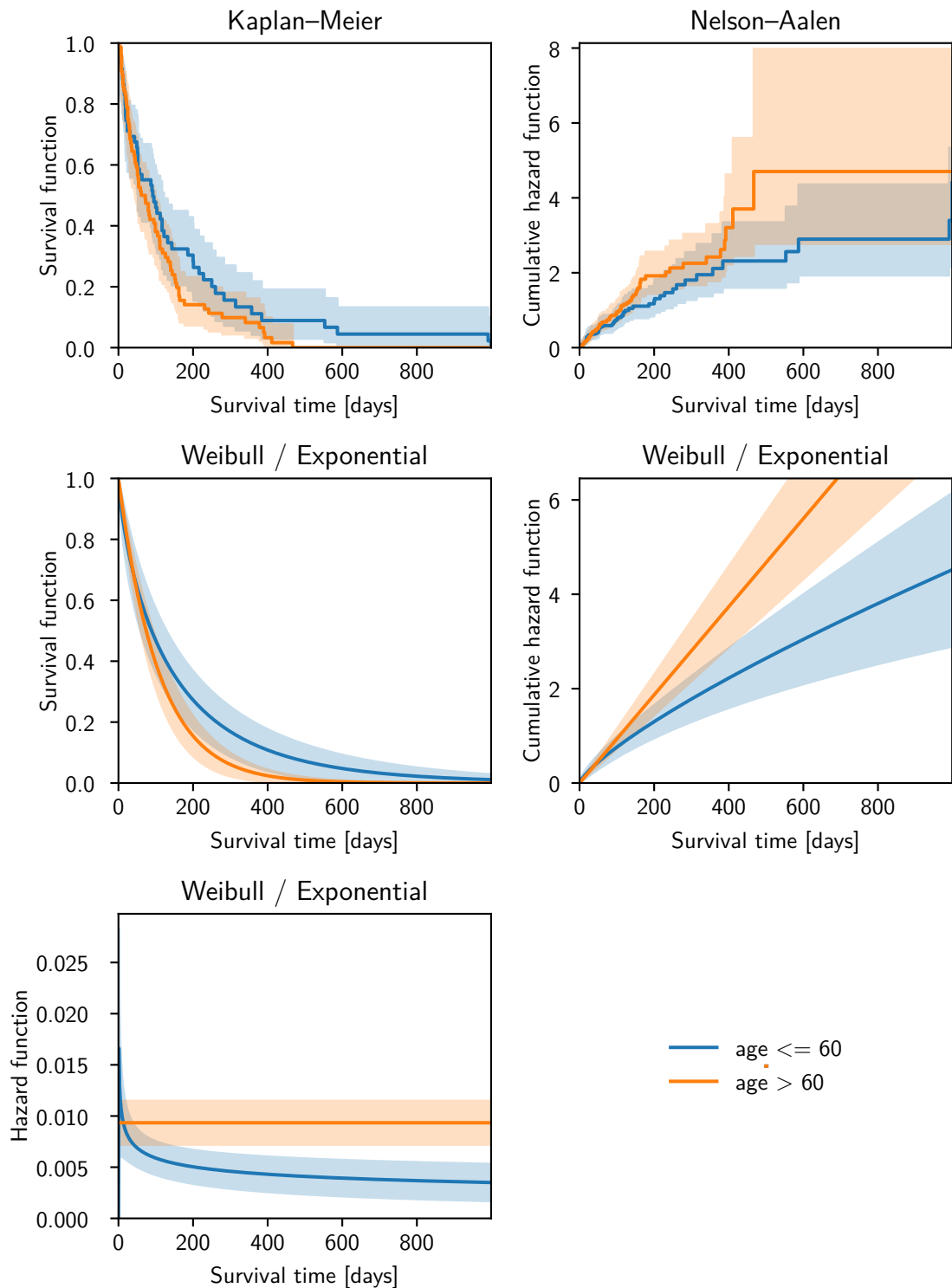
Tabulka 6: Odhadnuté parametry modelů a střední a mediánové doby dožití pro dané skupiny.

Srovnání různých typů buněk



Obrázek 7: Odhady funkce přežití, rizikové funkce a kumulativní rizikové funkce pro dané skupiny – Kaplanův–Meierův, Nelsonův–Aalenův a parametrické modely.

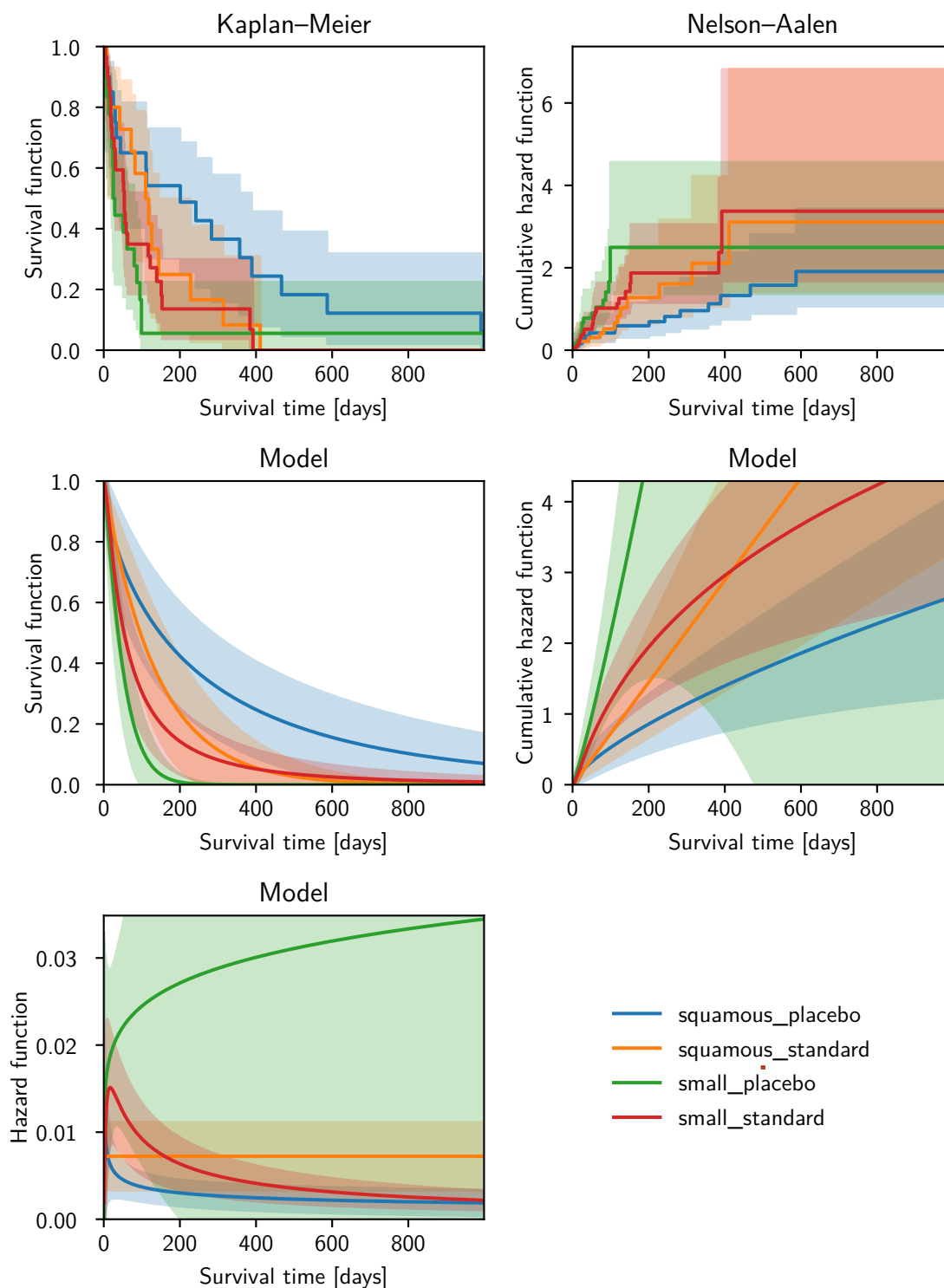
Srovnání různých věkových skupin



Obrázek 8: Odhady funkce přežití, rizikové funkce a kumulativní rizikové funkce pro dané skupiny – Kaplanův–Meierův, Nelsonův–Aalenův a parametrické modely.

- V případě typu small léčba prodlužuje dobu dožití (střední, mediánovou). Léčba má klesající rizikovou funkci a vyšší funkci přežití, zatímco placebo rostoucí a nižší.
- Léčba u typu squamous snižuje dobu přežití a zvyšuje rizikovou funkci – ta je u léčby konstantní, u placebo klesá.

Srovnání různých typů buněk a léčby



Obrázek 9: Odhady funkce přežití, rizikové funkce a kumulativní rizikové funkce pro dané skupiny – Kaplanův–Meierův, Nelsonův–Aalenův a parametrické modely.

3 Coxův model proporcionálních rizik

Pro zhodnocení vlivu jednotlivých proměnných na přežití pacienta byl sestavne Coxův model proporcionálních rizik. Protože jsou patrné výrazné odlišnosti mezi typy buněk (už z grafů v části 1), byl vytvořený model pro každý z typů zvlášť. Alternativou by byl jeden model s interakcemi proměnných a typu buněk, který by byl zbytečně složitý, vhodnější je použití čtyř jednoduchých modelů.

Pro každý z typů buněk byly uvažovány vlivy veličin Karnofsky score v desítkách, věku v desítkách let, nějaké předchozí péče oproti žádné a placebo oproti standardní léčbě:

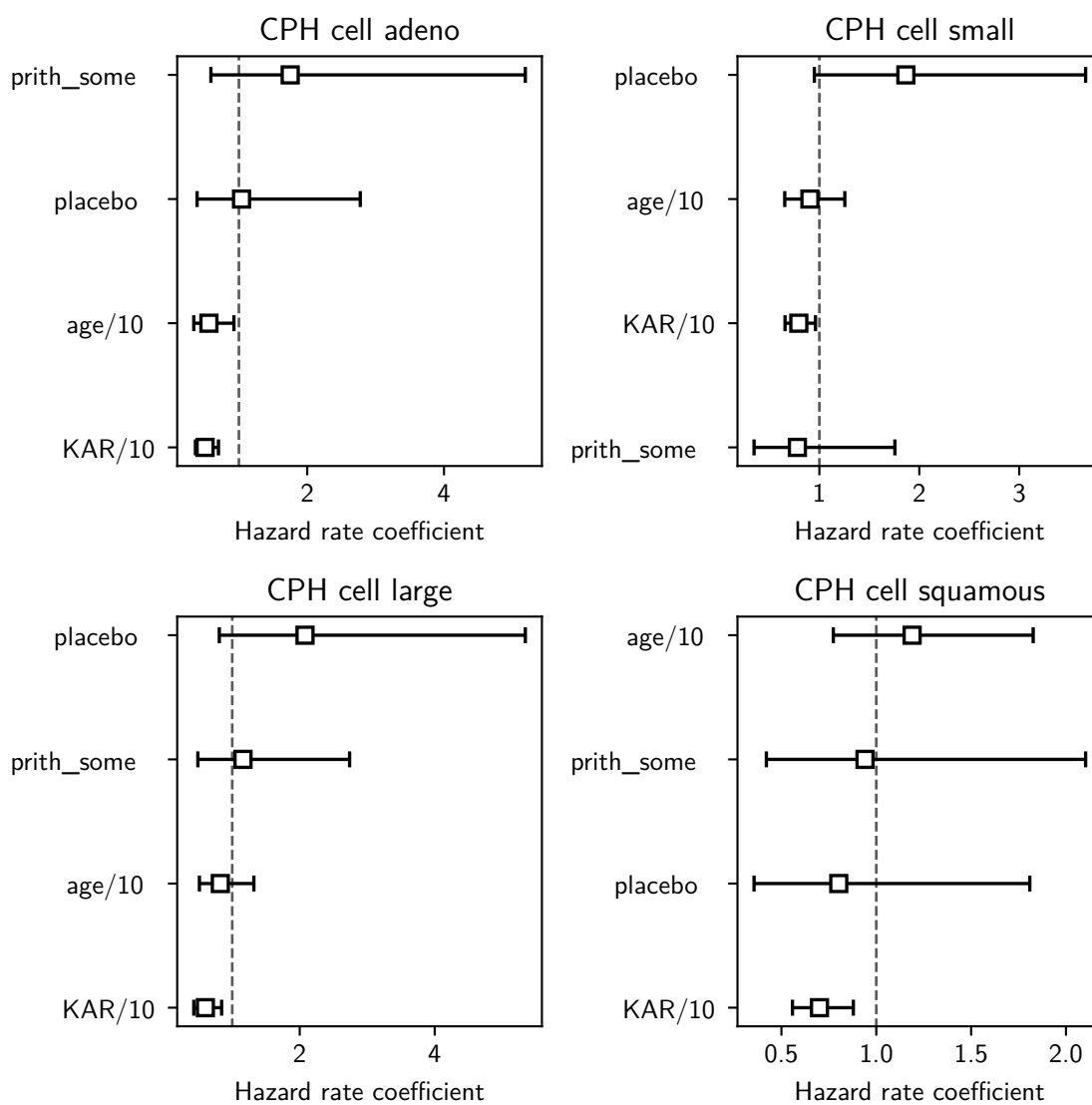
$$h(t; x) = h_0(t) \exp \left(c_1 \frac{KAR}{10} + c_2 \frac{age}{10} + c_3 [prith = some] + c_4 [treat = placebo] \right).$$

Odhadnuté regresní koeficienty jsou uvedeny v tab. 7 a znázorněny v grafech na obr. 10. Z jednotlivých modelů vyplývá, že

- zatímco u typu squamous užívání placebo riziko snižuje (léčba je škodlivá), u ostatních typů jej zvyšuje (výrazné rozdíly u typů small a large), jak je zřejmé z grafů v části 1,
- u typů adeno, large a small riziko s věkem klesá, u squamous roste,
- u prvních dvou typů předchozí péče riziko snižuje, u zbylých jej zvyšuje,
- vyšší stupeň hospitalizace (Karnofsky score) výrazně snižuje riziko,
- kromě veličiny Karnofsky score a age (u typu adeno) není vliv veličin statisticky významný (na hladině významnosti 5%).

Skupina	Veličina	Koeficient	exp koeficient	95% konf. int.
adeno	<i>KAR/10</i>	-0,68	0,51	-1,01; -0,36
	<i>age/10</i>	-0,58	0,56	-1,08; -0,08
	<i>prith_some</i>	0,56	1,75	-0,53; 1,65
	<i>placebo</i>	0,04	1,04	-0,95; 1,02
large	<i>KAR/10</i>	-0,51	0,60	-0,84; -0,17
	<i>age/10</i>	-0,19	0,82	-0,67; 0,28
	<i>prith_some</i>	0,15	1,16	-0,71; 1,01
	<i>placebo</i>	0,73	2,08	-0,21; 1,68
small	<i>KAR/10</i>	-0,23	0,79	-0,42; -0,04
	<i>age/10</i>	-0,10	0,91	-0,43; 0,23
	<i>prith_some</i>	-0,25	0,78	-1,06; 0,56
	<i>placebo</i>	0,62	1,87	-0,05; 1,30
squamous	<i>KAR/10</i>	-0,36	0,70	-0,58; -0,13
	<i>age/10</i>	0,17	1,19	-0,26; 0,60
	<i>prith_some</i>	-0,06	0,94	-0,87; 0,74
	<i>placebo</i>	-0,22	0,80	-1,03; 0,59

Tabulka 7: Koeficienty Coxova modelu pro jednotlivé typy buněk a jejich konfidenční intervaly.



Obrázek 10: Koeficienty (exponenciální) Coxova modelu jednotlivých typů buněk.

A Parametrizace rozdělení

Rozdělení	Survival function	Hazard rate	Parametry
Weibull	$S(t) = \exp\left(-\left(\frac{t}{\lambda}\right)^\rho\right)$	$h(t) = \frac{\rho}{\lambda} \left(\frac{t}{\lambda}\right)^{\rho-1}$	$\lambda, \rho > 0$
exponenciální	$S(t) = \exp\left(-\frac{t}{\lambda}\right)$	$h(t) = \frac{1}{\lambda}$	$\lambda > 0$
log-logistický	$S(t) = \left(1 + \left(\frac{t}{\alpha}\right)^\beta\right)^{-1}$	$h(t) = \frac{\left(\frac{\beta}{\alpha}\right) \left(\frac{t}{\alpha}\right)^{\beta-1}}{1 + \left(\frac{t}{\alpha}\right)^\beta}$	$\alpha, \beta > 0$
log-normální	$S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$		$\mu \in \mathbb{R}, \sigma > 0$

Tabulka 8: Parametrizace použitých rozdělení.