

MOŽNOSTI VYUŽITÍ ANALÝZY PŘEŽÍVÁNÍ V BIOMEDICÍNĚ A TECHNICE

K. Langová, J. Zapletalová, L. Ličman

Ústav lékařské biofyziky, Lékařská fakulta, Univerzita Palackého v Olomouci, ČR

Souhrn

Cílem tohoto článku je představit méně známou oblast statistiky – analýzu přežívání. Hlavním znakem datových souborů, které zkoumáme metodami analýzy přežívání, je skutečnost, že kromě konečných dat mohou obsahovat i data cenzorovaná. Cenzorovaná data vznikají, sledujeme-li u daného objektu dobu do výskytu dané události a k této události v době sledování nedojde, máme tedy k dispozici pouze neúplný časový údaj. V článku je popsáno, jak je možné odhadnout a interpretovat pravděpodobnostní funkci přežití, odhadnout medián přežívání a průměr přežívání. Pozornost je zaměřena na porovnání dvou funkcí přežití.

Klíčová slova

analýza přežívání, cenzorovaná data, Kaplan-Meierova křivka, medián a průměr přežívání, log-rank test, obecný Wilcoxonův test

Abstract

The goal of this paper is to introduce a less known area of the statistics – survival analysis. The main feature of data files that are investigated using survival analysis methods is the fact, that they contain censored data besides the final data. The censored data arise when the time to occurrence of a given event at the object is observed and the event does not occur in the time of observation. We have only non-complete time data. In this paper there is described how it is possible to estimate and interpret the survival function, survival median and mean. The paper focuses on a comparison of two survival functions.

Keywords

survival analysis, censored data, Kaplan-Meier curve, survival median and survival mean, log-rank test, generalized Wilcoxon test

Úvod

Statistika je vědní obor, jehož výsledky můžeme při své práci používat prakticky denně. Neexistuje snad jediný odborný medicínský časopis, ve kterém bychom nenašli články obsahující odbornou statistickou terminologii, a standardní statistické metody jsou většinou v povědomí lékařů a techniků známy. Cílem tohoto článku je představit méně známou oblast statistiky – analýzu přežívání. Metody analýzy přežívání je možné aplikovat na živé organismy i na poruchy technických

zařízení. Hlavním znakem datových souborů, které zkoumáme metodami analýzy přežívání, je skutečnost, že kromě konečných dat mohou obsahovat i data cenzurovaná. Cenzorovaná (neúplná) data vznikají, sledujeme-li u daného objektu dobu do výskytu dané události a k této události v době sledování nedojde, máme tedy k dispozici pouze neúplný časový údaj. Událostí (event) se přitom rozumí výskyt libovolného definovaného jevu, chápaného ze subjektivního pohledu pozorovatele pozitivně, negativně nebo i neutrálně. I když je možno se v konkrétních případech setkat

s různými názvy pro událost, často bývá označována jako selhání.

Materiál a metody

Cenzorovaná data

Obecně označuje termín analýza přežívání soubor statistických metod, které zpracovávají proměnnou čas do výskytu nějakého jevu (události), který se často nazývá též doba přežití. Časem nebo dobou přežití rozumíme počet let, měsíců, týdnů nebo dní od začátku sledování jedince až do výskytu jevu. Tento jev můžeme zvolit různě, jako příklady lze uvést úmrtí, incidenci onemocnění, návrat choroby, selhání transplantovaného orgánu nebo v oblasti techniky selhání určité součástky průmyslového zařízení (sledujeme dobu životnosti).

Většina studií zabývajících se analýzou přežívání je v praxi ukončena dříve, než u všech pozorovaných objektů (osob) nastane sledovaná událost. Tuto situaci nazýváme v analýze přežívání cenzorováním. K cenzorování dochází tehdy, když je sledování objektu (např. pacienta) ukončeno dříve, než nastala sledovaná událost.

Cenzorování může nastat ze tří důvodů:

- sledovaná událost u objektu nenastane do konce studie,
- objekt je v průběhu sledování ze studie vyřazen nebo sám dobrovolně odstoupí,
- objekt je ztracen pro pozorování ještě před ukončením studie (např. pacient zemře a smrt není sledovanou událostí).

Po ukončení klinické studie získáme vstupní údaje pro následné statistické zpracování. Obvykle máme k dispozici soubor pacientů se sledovanou diagnózou. Soubor můžeme rozdělit do tří skupin:

1. pacienti, u kterých nastala sledovaná událost (např. zemřeli) – konečná data
2. pacienti, u kterých dosud nenastala sledovaná událost (např. žijí) – cenzorovaná data
3. pacienti, kteří z nějakého důvodu byli vyřazeni ze studie nebo už prostě nemáme možnost je déle sledovat – rovněž cenzorovaná data.

U každého pacienta máme tedy dva údaje:

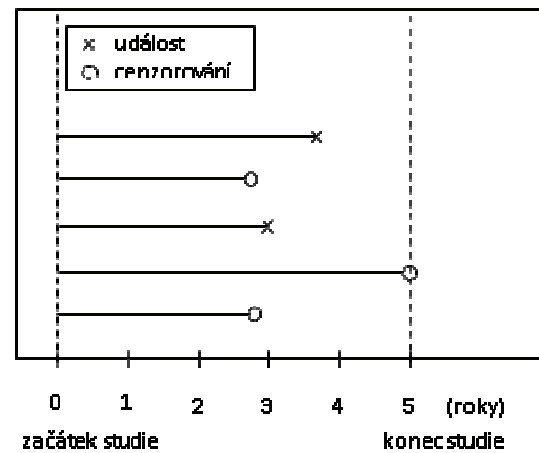
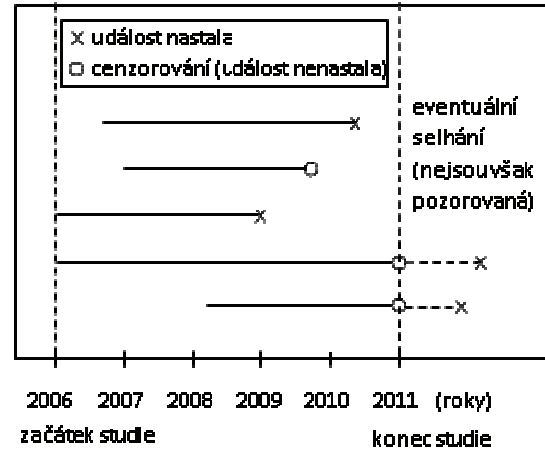
1. dobu sledování a
2. důvod ukončení sledování, který kódujeme číslem 1 pro případ, kdy sledovaná událost nastala (skupina 1) a číslem 0 pro případ, kdy sledovaná událost nenastala (skupina 2 nebo 3).

Cílem analýzy přežívání je

1. odhadnout a interpretovat pravděpodobnostní funkci přežití a rizikovou funkci
2. odhadnout medián přežívání a průměr přežívání
3. porovnat funkce přežití v případě, že studie pracuje s více skupinami pacientů

4. analyzovat vztah mezi sledovanými vysvětlujícími proměnnými a dobou přežití.

V článku se budeme zabývat pouze odhadem funkce přežití, mediánu a průměru přežívání a porovnáním více funkcí přežití.

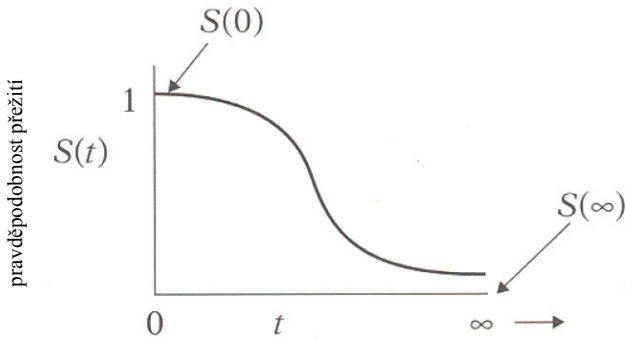


Obr. 1: Grafické znázornění přežívání pěti pacientů, kteří vstoupili do studie v různých časových okamžicích. Nahoře skutečný kalendářní čas, dole doba od vstupu do studie.

Odhad a interpretace funkce přežití

Jedním z hlavních cílů analýzy přežívání je odhadnout funkci přežití $S(t)$. Funkce přežití udává pravděpodobnost, že osoba žije déle než konkrétně specifikovaný čas t . $S(t) = P(T > t)$, kde T je nezáporná náhodná veličina, která vyjadřuje dobu uplynulou od zahájení pozorování jedince do výskytu sledovaného jevu (Obr. 2). V praxi se funkce přežití odhaduje metodou Kaplan a Meiera. V textu budeme symbolem $S(t)$ značit teoretickou funkci přežití platnou pro celou populaci a symbolem $\hat{S}(t)$ funkci přežití pro konkrétní soubor sledovaný studií. Hodnoty funkce $\hat{S}(t)$ se počítají ve všech časech t_j , kdy došlo u alespoň jednoho z pacientů ke sledovanému jevu, tedy pro všechna

konečná data. Výpočet vychází z teorie pravděpodobnosti a je založen na relativních četnostech.

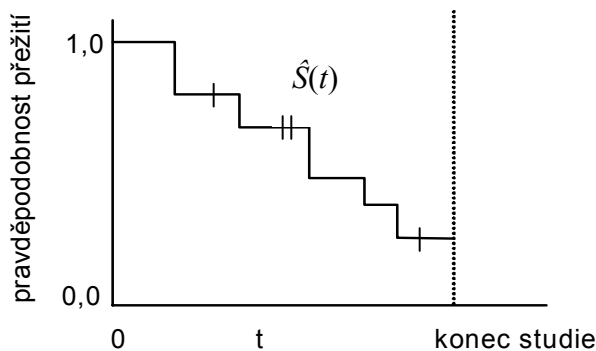


Obr. 2: Teoretický tvar funkce přežití.

Kaplan-Meierův odhad funkce přežití v čase t_j je definován následovně:

$$\hat{S}(t_j) = \prod_{i=1}^j P(T > t_i | T \geq t_i) = \hat{S}(t_{j-1}) \cdot P(T > t_j | T \geq t_j) \quad (1)$$

Skutečný tvar funkce přežití pro soubor sledovaný studií je znázorněn na obrázku 3. Na vodorovné ose je zaznamenán čas (tj. doba do výskytu sledované události ve dnech, týdnech, měsících nebo letech), na svislé ose pravděpodobnost přežití tohoto času, která nabývá hodnot od nuly do jedné. Funkce je schodovitá, každá událost (např. úmrtí) způsobí pokles pravděpodobnosti přežití (schod). Značky (svislé čárky) v grafu značí cenzorovaná (neúplná) data. Jinými slovy výskyt sledované události (např. úmrtí pacienta) vede ke snížení pravděpodobnosti přežití zbývajících subjektů.

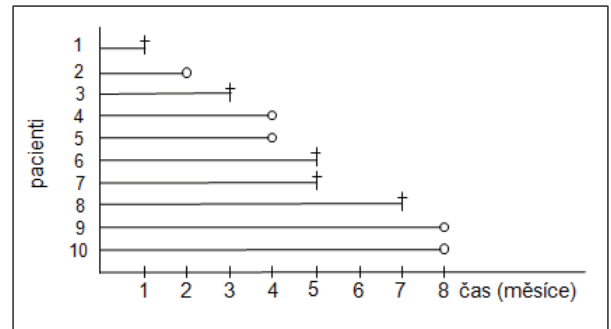


Obr. 3: Skutečný tvar funkce přežití, $\hat{S}(t)$.

Protože interval sledování klinické studie není nikdy nekonečný, je možné a v praxi časté, že sledovaná událost nenastane u všech pacientů a odhadovaná pravděpodobnost přežití neklesne na konci studie až k nule.

Postup odhadu Kaplan-Meierovy funkce přežití vysvětlíme na modelovém příkladu.

Na obrázku 4 jsou znázorněny údaje o 10 pacientech seřazené podle délky sledování (křížek značí výskyt sledované události např. úmrtí, kroužek je symbol pro cenzorovaný údaj).



Obr. 4: Vstupní data.

V čase $\hat{S}(t_0) = 0$ je pravděpodobnost přežití rovna 1.

V čase $t_1 = 1$ umírá první pacient, pravděpodobnost přežití tohoto časového okamžiku je $9/10$. Dle vzorce (1) dostáváme:

$$\hat{S}(t_1) = \hat{S}(t_0) \cdot P(T > t_1 | T \geq t_1) = 1 \cdot 9/10 = 0,9$$

V čase $t_2 = 2$ žádná událost nenastala, pravděpodobnost přežití tohoto časového okamžiku je tedy stejná $0,9$.

V čase $t_3 = 3$ sledujeme už jen 8 pacientů a jeden pacient umírá. Přirozeným odhadem přežití 3. měsíce je $7/8$, ale pouze pro pacienty, kteří se dožili 3. měsíce. Pro celkový odhad musíme násobit $7/8$ pravděpodobností přežití 2. měsíce, tedy:

$$\hat{S}(t_2) = \hat{S}(t_1) \cdot P(T > t_2 | T \geq t_2) = 0,9 \cdot 7/8 = 0,788$$

V čase $t_4 = 4$ žádná událost nenastala, pravděpodobnost přežití tohoto časového okamžiku je tedy stejná $0,788$.

V čase $t_5 = 5$ sledujeme 5 pacientů a dva pacienti umírají.

$$\hat{S}(t_3) = \hat{S}(t_2) \cdot P(T > t_3 | T \geq t_3) = 0,788 \cdot 3/5 = 0,473$$

V čase $t_6 = 6$ žádná událost nenastala, pravděpodobnost přežití tohoto časového okamžiku je tedy stejná $0,473$.

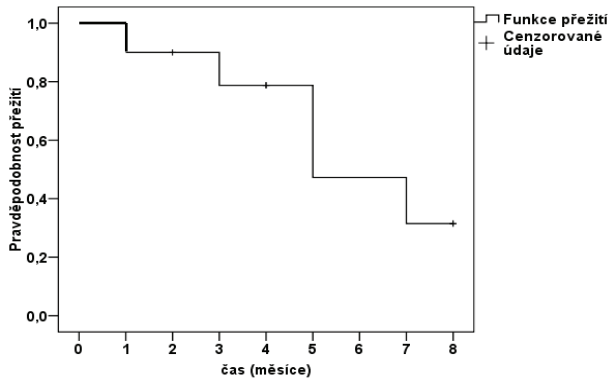
V čase $t_7 = 7$ sledujeme 3 pacienty a jeden pacient umírá.

$$\hat{S}(t_4) = \hat{S}(t_3) \cdot P(T > t_4 | T \geq t_4) = 0,473 \cdot 2/3 = 0,315$$

V čase $t_8 = 8$ žádná událost nenastala, pravděpodobnost přežití tohoto časového okamžiku je tedy stejná $0,315$.

Z vypočítaných hodnot funkce přežití můžeme funkci sestavit (obr. 5), cenzorované údaje jsou znázorněny pomocí krátkých úseček.

Obr. 5: Kaplan-Meierův odhad funkce přežití.



Odhad a interpretace mediánu a průměru přežívání

V analýze přežívání jsou důležitými údaji medián přežívání a průměrné přežívání.

Medián přežití je časový údaj, ve kterém je pravděpodobnost přežití rovna 0,5.

Průměrná doba přežívání se definuje jako plocha pod křivkou přežití. Pokud tedy soubor dat obsahuje cenzorovaná data, nelze pro výpočet mediánu a průměru použít klasické postupy známé z popisné statistiky pro konečná data.

Jestliže by při práci se souborem dat, který obsahuje i cenzorované údaje, byl výpočet průměru a mediánu proveden klasickým způsobem, znamenalo by to, že všechna data jsou konečná, tedy že u všech subjektů došlo ke sledované události.

V takovém případě by odhad průměru i mediánu byl nižší než při použití správného postupu, při kterém je rozlišováno, zda u daného objektu došlo k události či nikoliv.

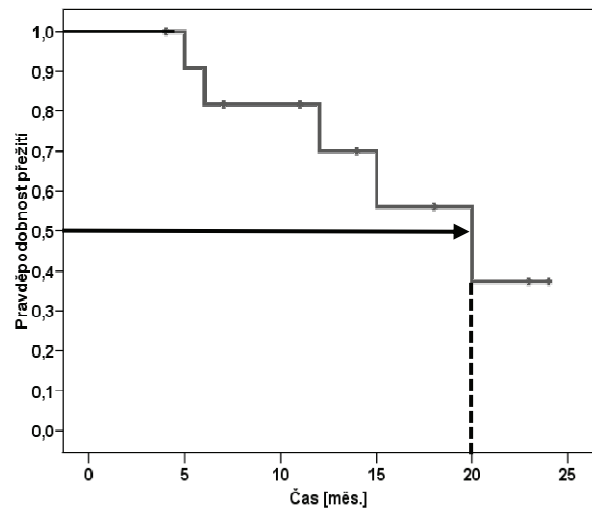
V tabulce 1a jsou zadány časy přežití 12 pacientů, ve sloupci status je kódováno, zda u pacienta nastala (=1) sledovaná událost či nenastala (=0). Tvar Kaplan-Meierovy křivky přežití s vyznačenou hodnotou mediánu je znázorněn na obrázku 6. Plocha pod křivkou přežití odpovídá průměrné hodnotě přežívání. Výpočet průměru a mediánu přežívání pomocí statistického programu SPSS i se standardní chybou odhadu (SE) a 95% intervalem spolehlivosti (95% CI) je ukázán v tabulce 1b.

Pokud by se s daty pracovalo pomocí klasických metod popisné statistiky, znamenalo by to, že všem objektům je přiřazen status 1 (tab. 2a), tvar Kaplan-Meierovy křivky (obr. 7) i odhady průměru a mediánu přežití (tab. 2b) by byly jiné.

Tab. 1a: Příklad souboru dat, který obsahuje konečné i cenzorované údaje.

Pacient	Přežití (mės.)	Status
1	15	1
2	23	0
3	18	0
4	20	1
5	5	1
6	11	0
7	4	0
8	14	0
9	6	1
10	7	0
11	12	1
12	24	0

Obr. 6: Kaplan-Meierova křivka přežití pro data z tab. 1a se znázorněnou hodnotou mediánu přežití.

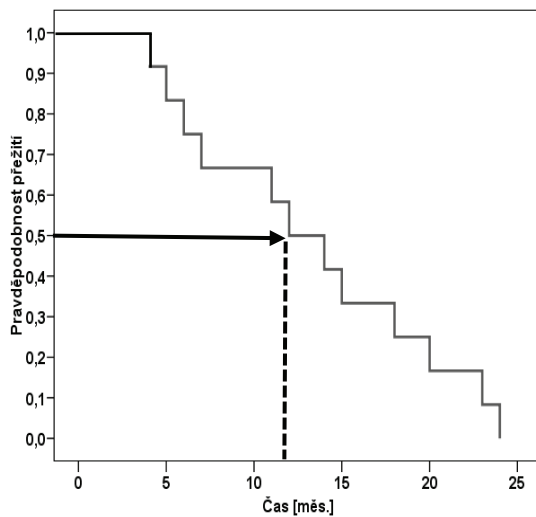


Tab. 1b: Výpočet průměru a mediánu přežívání pro data z tab. 1a pomocí statistického software SPSS.

		Průměr		Medián			
		95% CI		95% CI			
Odhad	SE	Dolní hranice	Horní hranice	Odhad	SE	Dolní hranice	Horní hranice
17,2	2,2	12,8	21,6	20,0	5,1	10,0	30,0

Tab. 2a: Příklad souboru dat, který obsahuje chybně pouze konečné údaje (tj. cenzorované údaje jsou považovány za konečné).

Pacient	Přežití (měs.)	Status
1	15	1
2	23	1
3	18	1
4	20	1
5	5	1
6	11	1
7	4	1
8	14	1
9	6	1
10	7	1
11	12	1
12	24	1



Obr. 7: Kaplan-Meierova křivka přežití pro data z tab. 2a se znázorněnou hodnotou mediánu přežití.

Tab. 2b: Výpočet průměru a mediánu přežívání pro data z tab. 1a pomocí statistického software SPSS.

Průměr				Medián			
Odhad	SE	95% CI		Odhad	SE	95% CI	
		Dolní hranice	Horní hranice			Dolní hranice	Horní hranice
13,3	2,0	9,3	17,2	12,0	2,6	6,9	17,1

Porovnání funkce přežití pro více vstupních souborů

Běžným problémem v klinických studiích je srovnání dvou nebo více typů léčby z hlediska jejich schopnosti prodloužit život pacientů nebo udržet pacienty v remisi. Vystává tedy problém porovnat dvě nebo více křivek přežití. Pro toto srovnání existuje několik

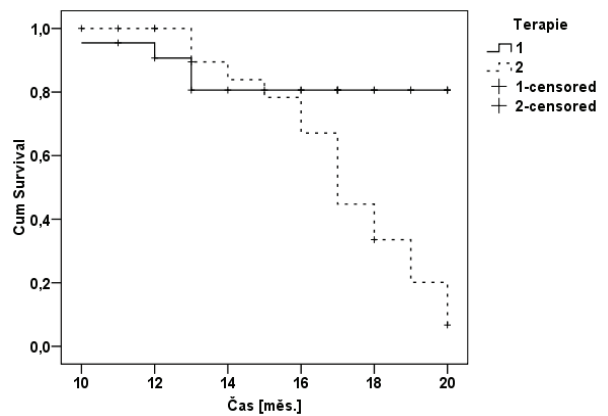
statistických testů. Nejpoužívanější jsou log-rank test a obecný Wilcoxonův test. Tyto testy nám odpovídají na otázku, zda jsou Kaplan-Meierovy křivky statisticky ekvivalentní.

Log-rank test je v podstatě chí-kvadrát test pro velký výběr. Log-rank statistika porovnává pozorovaný počet jevů v každém časovém okamžiku, kdy došlo ke sledovanému jevu, s očekávaným počtem za platnosti nulové hypotézy, která předpokládá, že porovnávané křivky přežití se neliší.

Obecný Wilcoxonův test nepracuje s konkrétními časy do výskytu sledovaného jevu, ale pouze s pořadím výskytu událostí a cenzorováním v jednotlivých souborech. Každý takto uspořádaný údaj v jednom souboru se porovnává se všemi upořádanými údaji ve druhém souboru. Tím vyloučíme vliv extrémních časových údajů ve vstupních souborech.

Je-li dosažená hladina signifikance příslušného testu menší než 0,05, považujeme posuzované typy léčby za rozdílné z hlediska jejich schopnosti prodloužit život pacientů.

Zabýváme-li se otázkou, kdy použít log rank test a kdy použít Wilcoxonův test, je třeba vědět, že hlavní rozdíl mezi testy je v tom, jakou kladou během doby přežití váhu na jednotlivé události. Log-rank test klade během doby sledování na všechny události stejnou váhu. Wilcoxonův test se od předchozího liší tím, že čím je doba pozorování delší, tím dává menší váhu rozdílům mezi pozorovaným a očekávaným počtem sledovaných událostí. Jednoduchou pomůckou, jak zjistit, který z testů použít, je vykreslit si obě funkce přežití pro jednotlivé skupiny do jednoho grafu a podle jejich průběhu vybrat vhodnější test. Pokud se například tyto dvě funkce kříží (obr. 8), je vhodnější zvolit k testování Wilcoxonův test. Pokud se funkce nekříží, je lepší zvolit k testování, zda jsou funkce přežití pro jednotlivé skupiny rozdílné či nikoliv, log-rank test. Křivky přežití, které jsou znázorněny na obrázku 8, byly porovnány pomocí statistického software SPSS dvěma statistickými testy log-rank testem a Wilcoxonovým testem (tab. 3).



Obr. 8: Funkce přežití pro dvě skupiny pacientů léčených dvěma terapiemi.

Tab. 3: Výsledek dvou statistických testů, výstup z programu SPSS.

	Chi-kvadrát	df	Sig.
Log-rank test	7,115	1	,008
Obecný Wilcoxonův test	1,914	1	,166

Z tabulky 3 je zřejmé, že dosažená hladina signifikance 0,008 u log-rank testu je nižší než obecně uvažovaná hladina $\alpha = 0,05$, výsledek bychom tedy mohli považovat za statisticky významný a rozdíly v přežívání pacientů léčených dvěma terapiemi za skutečné. Naopak dosažená hladina signifikance u Wilcoxonova testu je 0,116, v tomto případě by byl výsledek považován za statisticky nevýznamný. Použijeme-li pomůcku o křížení křivek přežití, rozhodneme se použít Wilcoxonův test.

Diskuse

Na příkladech bylo ukázáno, jak je možné špatně interpretovat výsledky statistického šetření, pokud je použita nevhodná metoda. V případě výpočtu průměru a mediánu u souboru dat, ve kterém jsou obsaženy i cenzorované údaje, dochází k podhodnocení odhadu obou statistik. Hodnoty průměru i mediánu vycházejí nižší ve srovnání s hodnotami, které dostaneme, pokud správně aplikujeme metody analýzy přežívání.

Při porovnání funkcí přežívání statistickými testy, je třeba zvolit vhodný test. Pokud se nezamýšlíme nad volbou správného testu a vybereme například první možnost, kterou nám statistický software nabízí, může také dojít k chybné interpretaci závěrů statistického zpracování.

Závěr

Při statistickém zpracování dat je třeba zvážit charakter dat a zvolit správnou statistickou metodu. Kromě klasických statistických metod, které pracují s konečnými daty, existují také metody analýzy přežívání, které berou do úvahy skutečnost, že informace u všech subjektů nejsou úplné. Při nevhodně zvolené metodě může dojít ke špatné interpretaci výsledků statistického zpracování.

Poděkování

Tato práce byla zpracována za podpory projektu CZ.1.07/2.4.00/17.0058 Prohloubení odborné spolupráce a propojení ústavů lékařské biofyziky na lékařských fakultách a projektu CZ.1.05/2.1.00/01.0030.

Literatura

- [1] Akritas G. Michael: *Nonparametric Survival Analysis*, Statistical Science, 2004, vol. 19, no. 4, p. 615–623.
- [2] Alling D. *Early decision in the Wilcoxon two-sample test*. J. Amer. Statist. Ass. 1963; vol. 58, p. 713–20.
- [3] Altman D. G. *Practical statistics for medical research*. Chapman & Hall, UK 1995.
- [4] Anděl J. *Matematická statistika*. SNTL – Nakladatelství technické literatury, Praha, 1985.
- [5] Gehan, E. A. *Statistical methods for survival time studies*. In: *Cancer Therapy: Prognostic Factors and Criteria*, M. J. Staquet, eds. New York: Raven Press; 1975.
- [6] Kalbfleisch, J. D., Prentice P.L. *The statistical analysis of failure time data*. New York: John Wiley & Sons, Inc., 1980.
- [7] Kaplan E. L., Meier P. *Nonparametric estimation from incomplete observations*. J Amer Statist Ass, 1958, vol. 53, p. 457–481.
- [8] Klein J. P., Goel P. K. *Survival Analysis: State of the Art*. Kluwer Academic Publisher, Dordrecht, The Netherlands, 1992.
- [9] Klein J. P., Moeschberger M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, USA, 1997.
- [10] Kleinbaum G. David, Klein Mitchell: *Survival Analysis, A Self-Learning Text*, Springer New York, 2005.
- [11] Langová K.: *Survival analysis for clinical studies*, Biomed Pap Med Fac Univ Palacky, 2008, vol. 52, no.2, p. 303–307.
- [12] Lawless R. F. *Statistical models and methods for lifetime data*. John Wiley & Sons, Inc., New York, USA, 1982.
- [13] Zheng Z. *Two methods of estimating the mean survival time from censored samples*. The Indian Journal of Statistics, 1995, vol. 57, no. 1, p. 126–136.

Kateřina Langová, Ph.D.
Ústav lékařské biofyziky,
Lékařská fakulta
Univerzita Palackého v Olomouci
Hněvotínská 3, CZ-775 15 Olomouc

E-mail: langova@tunw.upol.cz
Tel.: +420 585 632 120