

MASARYKOVA UNIVERZITA

Přírodovědecká fakulta

Institut biostatistiky a analýz

**Vliv rizikových faktorů na výskyt
zdravotních komplikací u pacientů
s domácí umělou výživou**

Týmový projekt

Vypracovali:

Jan Buďa

Tereza Jurková

Magdaléna Skácelová

Vedoucí práce:

Mgr. Jiří Šilar

Brno 2017

Obsah

Úvod	3
1 Parenterální výživa	4
1.1 Způsob zavedení.....	4
1.2 Typy výživových roztoků.....	4
1.3 Příčiny zavedení parenterální výživy (PV)	5
1.4 Komplikace	5
1.5 Domácí parenterální výživa (DPV).....	6
2 Datový soubor	7
2.1 Popis datového souboru	7
2.2 Kontrola, čištění a úprava dat.....	7
2.3 Popisná statistika	9
2.3.1 Kvalitativní data	10
2.3.2 Kvantitativní data	12
3 Formulace a statistické testování hypotéz	16
3.1 Formulace hypotéz	16
3.2 Statistické testování hypotéz	17
4 Analýza přežití pacientů s DPV	24
4.1 Základní pojmy	24
4.2 Neparametrický Kaplan-Meierův odhad funkce přežití.....	26
4.3 Aplikace Kaplan-Meierovy metody	27
Závěr	32
Seznam literatury a použitých zdrojů	34
Seznam tabulek	35
Seznam obrázků	36

Úvod

Zadání projektu

Parenterální výživou označujeme způsob dodávání živin mimo trávicí ústrojí přímo do krevního řečiště s cílem dlouhodobého udržení dobré výživy pacientů. V současnosti je parenterální výživa vyhrazena pro stavy, u kterých z důvodu dysfunkce trávicího ústrojí nelze použít varianty výživy enterální cestou. Jedná se především o pacienty s postiženou funkčností střev, trpících například syndromem krátkého střeva (stav po resekci střeva), střevní píštěle, malabsorpce či jinými poruchami trávení. Tento způsob nutriční podpory je však zatížen vyšším rizikem vzniku komplikací, ke kterým patří riziko vzniku infekce, výskyt žilní trombózy a jiné.

Byl dodán anonymizovaný datový soubor z reálné klinické praxe obsahující základní údaje o pacientech, kteří nejsou schopni přijímat živiny per os a kterým je z důvodu těžkých onemocnění podávána potrava cestou domácí parenterální výživy (DPV).

Cíle týmového projektu

- Popsat stav pacientů s DPV v přiloženém datovém souboru
- Vytipovat rizikové faktory a jejich vliv na výskyt zdravotních komplikací
- Nepovinný cíl: Nastudovat problematiku analýzy přežití pomocí Kaplan-Meierovy křivky a její použití na souboru pacientů s DPV

Zadání analýzy

Zahrnuje tyto úkony:

- Zorientovat se v přiloženém datovém souboru
- Zkontrolovat a vyčistit data před analýzou
- Prezentovat základní charakteristiky pacientů s DPV
- Vytipovat rizikové faktory, které mohou mít vliv na výskyt zdravotních komplikací a jejich statistické ověření
- Nepovinný úkol: Využít analýzu přežití dle metody Kaplan-Meiera na datech pacientů s DPV

1 Parenterální výživa

Parenterální výživa je způsob podávání živin pacientovi nefyziologickou cestou. Jedinec není schopen přijímat potravu per os nebo enterálně, proto mu výživa musí být podávána mimo zažívací trakt přímo do cévního systému. Pacientovi můžou být zjištěny přesné poměry živin, které mu byly podány a lékař může rychle reagovat na změny podáním pozměněného složení infuze. [6]

1.1 Způsob zavedení

Tato náhradní výživa je pacientovi podávána ve formě umělých substrátů a živin buď do centrální nebo periferní žíly.

Parenterální výživa se do centrální žíly zavádí v případech, kdy předpokládáme nutriční podporu po delší dobu, minimálně 1 týden. [6] Těmto pacientům se katétr nejčastěji zavede do *vena subclavia* nebo *vena jugularis*, velmi často se katétr vyvádí podkožním tunelem dále od místa primárního vpichu, což snižuje riziko možných infekčních komplikací, umožňuje měření centrálního žilního tlaku a dovoluje opětovné krevní odběry pouze za přísných podmínek, vždy se ale musí počítat s možností vzniku sepse. Další výhodou této cesty je, že pacientovi můžou být podávány roztoky malého objemu o velké koncentraci a nevzniká u nich riziko flebidit.

Parenterální výživa přes periferní řečiště se podává pacientům, u kterých se předpokládá, že tento způsob výživy bude jen krátkodobý, při přechodném odstranění katétru v centrální žíle nebo když není možné zavést trubici do centrální žíly. [3] Ke kanylaci se nejčastěji používají žíly horních končetin. Občas se využívají i žíly dolních končetin, ale to jen výjimečně, protože omezují pacienta v pohybu a navíc je zde výrazně vyšší riziko vzniku tromboflebitidy. Touto cestou nemocný přijímá roztoky o velkém objemu a vysoké osmolaritě, čímž je drážděna žilní stěna, což může vést až ke vzniku zánětu. [1]

1.2 Typy výživových roztoků

Výživové roztoky jsou ve formě infuzí přes infuzní pumpu podávány pacientovi především v noci, aby přes den mohl vykonávat běžné činnosti a nebyl nijak omezován. Přípravky jsou ve dvou formách, buď je to systém all-in-one nebo multi-bottle.

Systemem all-in-one jsou dodávány všechny živiny (lipidy, aminokyseliny i sacharidy) ve formě vícekomorových vaků současně, což je srovnatelné s běžným režimem příjmu potravy. Komory vaku se po přidání vitaminů, stopových prvků a iontů bezprostředně před podáním pacientovi smíchají. Tento systém má mnoho výhod. Patří mezi ně především nízká cena, delší doba trvanlivosti a jednoduchá aplikace, která snižuje výskyt infekčních komplikací.

Systemem multi-bottle jsou bílkoviny, tuky a cukry podávány v jednotlivých lahvích. To zapříčiňuje vznik řady komplikací. Ošetrovatelský personál musí pacientovi podávat 6-8 lahví denně. Častá manipulace s infuzemi způsobuje vyšší riziko porušení aseptického okruhu a s tím spojené vážné komplikace, především septické. Tento způsob podávání živin příliš neodráží běžný režim příjmu potravy a navíc hladiny jednotlivých substrátů kolísají. Další nevýhodou je vyšší cena a nízká doba trvanlivosti. Na rozdíl od all-in-one nemají standardizované složení, takže mohou být připraveny přesně na míru danému pacientovi. [3], [6]

1.3 Příčiny zavedení parenterální výživy (PV)

K hlavním příčinám zahájení parenterální výživy patří poškození trávicího traktu, zejména způsobenými malabsorpčními syndromy, syndromem krátkého střeva, rozsáhlými střevními operacemi, akutní fází ledvinového nebo jaterního selhání, mnohočetnými fistulacemi střeva, protrahovaným průjmovým onemocněním a dalšími. [3]

1.4 Komplikace

Jelikož se nejedná o fyziologické podávání živin, je tato cesta spjata s řadou komplikací. U pacientů jsou časté zejména infekční komplikace, které jsou spojeny se zaváděním katétru, porušením aseptického okruhu, manipulací s infuzními vaky, nebo s bakteriemi, které se množí ve střevě a přes jeho stěnu se dostávají až do krve. Další možnou komplikací je vznik trombózy, která je způsobena ucpáním žíly sraženinou, která se může dostat až do plic a způsobit plicní embolii. [3] Velmi časté jsou hepatobiliární komplikace, tedy onemocnění jater a žlučových cest. Patří mezi ně cholestáza, steatóza, fibróza a jiné. Při cholestáze není v těle tvořeno normální množství žluči, což vede ke zvýšení hladiny bilirubinu v krvi. Špatné trávení tuku a jeho hromadění v játrech označujeme termínem

steatóza. Fibróza se vyznačuje zmnožením vaziva v určitém orgánu, to může vést až k jeho ztuhnutí. [2] Pacienti přijímající umělé výživové roztoky jsou zatíženi i dalšími možnými komplikacemi, např. u nich mohou nastat metabolické komplikace (hyper/hypoglykémie, hypertriglyceridémie, hyperkapnie a jiné). [1]

1.5 Domácí parenterální výživa (DPV)

Domácí parenterální výživa patří mezi dlouhodobou výživu, kdy se předpokládá, že pacient bude muset přijímat živiny touto cestou minimálně několik týdnů nebo až do konce života a je ve stabilním stavu, takže není nutné o něj pečovat v nemocničním zařízení, ale mohou se o něj starat členové rodiny v domácím prostředí. [3] DPV může být zavedena už u dětí, které jsou starší tří měsíců. [1]

2 Datový soubor

V této kapitole nejprve zmíníme stručný popis datového souboru, který nám byl dodán jako reprezentativní vzorek pacientů s domácí parenterální výživou z národního registru. Dále se budeme zabývat kontrolou, čištěním a úpravou dat. Nakonec pro ilustraci a lepší přehlednost uvedeme vizualizace některých důležitých kvalitativních (kategoriálních) a zejména kvantitativních (měřitelných) proměnných v rámci popisné statistiky.

2.1 Popis datového souboru

Zadání týmového projektu bylo ve formátu Word, data jsme dostali v podobě datového souboru ve formátu Excel. Celkem se v souboru nacházelo 35 proměnných (sloupců tabulky), které představují základní informace o pacientech (např.: ID pacientů, datum narození, pohlaví, fyzická zdatnost, ekonomická aktivita pacientů, datum úmrtí a další), informace o DPV (datum zahájení výživy, kategorie DPV, datum ukončení výživy, příčina DVP a případné specifikace atd.) a nejrůznější typy zdravotních komplikací (infekční, hepatobiliární a jejich druhy, počet hospitalizací a mnoho dalších).

Data byla zaznamenána o 466 pacientech, kdy každé ID pacienta je unikátní – tedy celkový počet pacientů odpovídá počtu řádků (záznamů) tabulky.

Některé proměnné byly zadávány buď při vstupu pacienta do registru, nebo při následné kontrole záznamů v registru, kdy zadavatel mohl data upravit s relativně dlouhým odstupem.

2.2 Kontrola, čištění a úprava dat

Při kontrole záznamů v souboru jsme zjistili, že data byla už značně předpřipravena (např. unikátní ID pacientů nejsou vždy samozřejmostí) a naším úkolem tedy bylo najít a vyřešit nesrovnalosti, které by mohly hrát významnou roli při vizualizaci dat a zejména při testování námi formulovaných hypotéz.

Nejprve bylo nutné zanalyzovat chybná nebo chybějící data o pacientech. Zde jsme byli nuceni odstranit záznamy s chybnými daty pacientů, u kterých předcházelo datum zahájení DPV datu narození pacienta anebo u kterých datum úmrtí předcházelo datu zahájení

DPV. Dále jsme se rozhodli smazat i pacienty s nevyplněným datem zahájení výživy, což je nezbytný údaj pro určení délky léčby. Museli jsme vyřadit i záznamy s nechronologickými daty, kdy například datum aktualizace záznamů v registru předcházelo datu zahájení DPV nebo datu úmrtí pacientů anebo datu ukončení výživy. Někdy bylo poněkud obtížné hledání nesmyslných a nelogických hodnot – kupříkladu, kdy byl uveden nulový počet hospitalizací pacientů za poslední rok a zároveň byl uveden také kladný počet hospitalizovaných dní za rok – tyto záznamy jsme samozřejmě také zavrhlí pro naše účely. Co bylo asi nejnáročnější – rozhodnout na základě nabytých znalostí o této problematice a po dohodě s vedoucím naší práce, jaké odlehle hodnoty považovat za opravdu extrémní, které nemohou nastat nebo mohou významně zkreslit naše výsledky a je nutné je tedy odstranit, a jaké ještě tolerovat i přesto, že se občas jednalo o významné odchylky v porovnání s ostatními pacienty. Nakonec jsme se rozhodli smazat tyto záznamy: pacienta, který měl uvedeno 77 infuzí týdně (viz téma „2.3 Popisná statistika“), což nám přišlo jako jev nemožný, dále pacienta s celkovým počtem 230 hospitalizací kvůli DPV za poslední rok, pacienta s nulovým věkem k datu úmrtí či aktualizace záznamů v registru, protože tato databáze zahrnuje pouze dospělé pacienty, a v neposlední řadě i pacienta, který měl shodné datum narození s datem zahájení DPV a jehož délka léčby byla dokonce 65 let. Celkem jsme odstranili ze souboru 40 pacientů, takže pro následnou popisnou statistiku a testování hypotéz bylo použito zbývajících 426 záznamů.

Dále pro sjednocení dat jsme upravili prázdné (nevyplněné) hodnoty různých proměnných. Do těchto hodnot bylo u většiny kategoriálních proměnných (např.: fyzická zdatnost, dominantní problém, příčina DPV – diagnóza, přítomnost hepatobiliárních komplikací a další) doplněno „Neuvedeno“, zatímco u všech typů hepatobiliárních komplikací (jako je např. cholestáza, steatóza, fibróza a jiné) jsme doplnili „Ne / Nezjišťováno / Neuvedeno“ podle přítomnosti hepatobiliárních komplikací. Nakonec jsme upravili nevyplněné hodnoty některých kvantitativních proměnných (např.: počet infekčních komplikací, žilních trombóz, počet hospitalizovaných dnů pouze v případě, že počet hospitalizací byl nulový,...) na „0“.

Pokud byly přítomny hepatobiliární komplikace, ale nebyl uveden konkrétní typ komplikace, tak jsme do všech prázdných proměnných doplnili „Neuvedeno“. Pokud pacient měl nulový počet infuzí týdně, tak jsme překódovali na prázdnou hodnotu ve významu neuvedeno, protože pacient by měl dostat nějakou infuzi, jestliže zahájil DPV (nesrovnalost mohla vzniknout při zadávání dat nebo při přípravě dat na projekt.)

Dále byly ještě doplněny některé proměnné pro lepší vizualizaci dat v rámci popisné statistiky i pro následné testování hypotéz. Například jsme přidali proměnnou „vek“ s věkem pacientů v letech k datu úmrtí či aktualizace záznamů v registru, abychom mohli vyřadit pacienty, kteří ani v průběhu léčby nedosáhli věku 18 let, a proto by se neměli vyskytovat v našem souboru, který zahrnuje pouze dospělé pacienty. Kvůli vizualizaci a pro účely testování hypotéz jsme standardně uvedli i proměnnou „vek2“ s věkem pacientů v letech k datu zahájení DPV. Samozřejmostí bylo i přidání proměnné „delka_lecby“ v letech s přesností na jedno desetinné místo. Jelikož máme v plánu aplikovat na náš soubor i analýzu přežití pomocí Kaplan-Meierovy křivky, tak jsme doplnili proměnnou „umrti“ s hodnotami 1/0 (ano/ne). A nakonec jsme zejména kvůli testování hypotéz vytvořili i proměnnou „komplikace“ udávající přítomnost jakéhokoli alespoň jednoho typu komplikace nebo nepřítomnost žádné komplikace (ano/ne). Tato proměnná je doplněna o „pocet_komplikaci“ udávající celkový počet zdravotních komplikací, avšak u mnoha pacientů tento počet nebylo možné určit.

Při tvorbě nových proměnných jsme narazili ještě na jeden problém – tím je délka léčby. Pro určení délky léčby jsme nezahrnovali pacienty, kteří umřeli a ukončili léčbu, ale neměli vyplněné datum ukončení výživy a ani datum úmrtí. Dále jsme byli nuceni pominout i pacienty, u kterých bylo uvedeno, že ukončili léčbu a žijí, ale nemají vyplněné datum ukončení DPV. Ze souboru vyřazeni nebyli, avšak pro testování hypotéz, kde je klíčová délka léčby, tyto pacienti nebudou zahrnuti.

2.3 Popisná statistika

Cílem této podkapitoly je přiblížit čtenáři strukturu našeho experimentálního vzorku v podobě nejrůznějších vizualizací a popisných statistik nejdůležitějších kategoriálních i numerických proměnných, které budou hrát významnou – klíčovou roli v následující analýze dat zahrnující statistické testování hypotéz.

Nejprve si představíme některé kvalitativní proměnné formou sloupcových (pruhových) grafů a tabulek četností, následovat budou proměnné kvantitativní, znázorněné pomocí krabicových grafů (boxplotů) i histogramů a doplněné tabulkou základních popisných statistik.

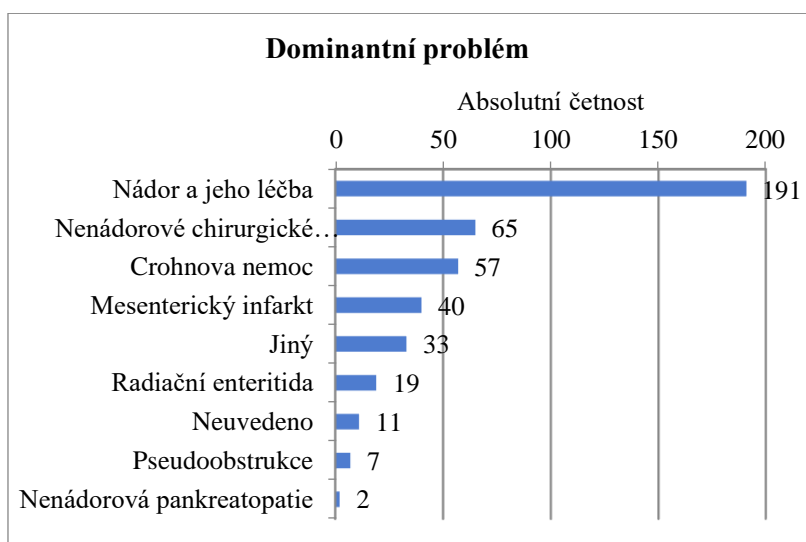
2.3.1 Kvalitativní data

V našich datech je velké množství kvalitativních dat. Tedy těch, která se dají roztrždit do několika kategorií. Jsou to různé základní informace o pacientovi (např. pohlaví), ale i o jeho zdravotním stavu a žilní výživě (např. příčina výživy, typ katetru, přítomnost hepatobiliárních komplikací).

Pohlaví není zastoupeno úplně rovnoměrně. Mužů máme jen 39 % oproti ženám, kterých je 61 %. Nejčastější problém našich pacientů je nádor a jeho léčba (u 45 % pacientů) (viz Obr. č. 1). Také je častá Crohnova nemoc, což je chronické zánětlivé onemocnění v některé části trávicí soustavy (ve 13 % případů). [4]

Diagnóza – příčina výživy		
	Četnost	Relativní četnost
Krátké střevo	171	40,2 %
Obstrukce	107	25,2 %
Malabsorbce	59	13,9 %
Píštěl	32	7,5 %
Anorexie	9	2,1 %
Dysfagie	3	0,7 %
Jiná	34	8,0 %
Neuvedeno	10	2,4 %

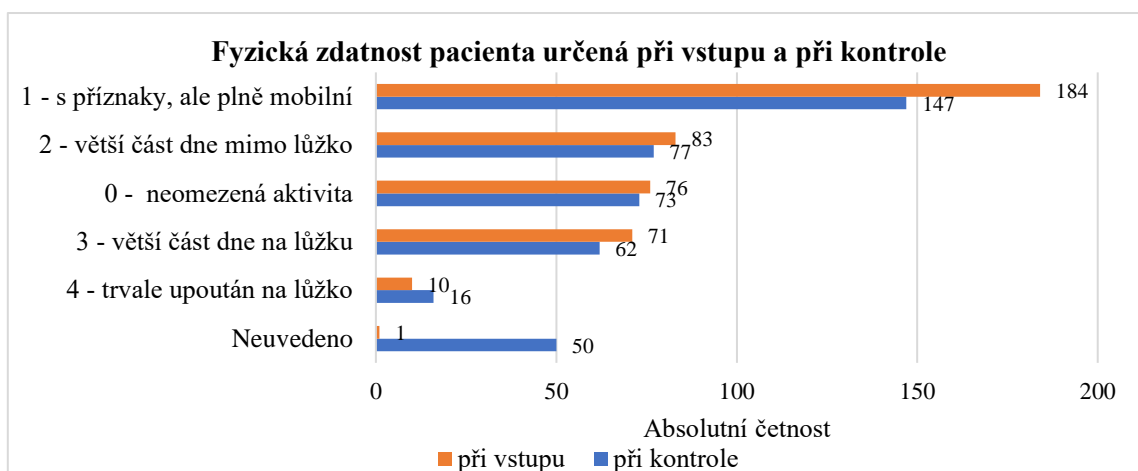
Tab. č. 1 Četnosti diagnóz pacientů



Obr. č. 1 Četnosti jednotlivých typů dominantních problémů pacientů

Dvě hlavní příčiny zavedení nitrožilní výživy jsou syndrom krátkého střeva (40 % případů) a obstrukce – zamezení průchodnosti střeva. Jen tyto dvě diagnózy jsou u více než 65 % pacientů (viz Tab. č. 1).

Fyzická zdatnost pacienta byla zaznamenána při začátku léčby a při ukončení nebo poslední kontrole. Lékaři zdatnost pacientů zaznamenali číslem od 0 do 4 (neomezená aktivita až trvale upoután na lůžko). Na vizualizaci si můžete všimnout, že během léčby vzrostl počet pacientů trvale upoutaných na lůžko. Nejvíce je ale patrný nárůst pacientů, u kterých lékař nevyplnil fyzickou zdatnost při kontrole (viz Obr. č. 2).



Obr. č. 2 Srovnání četností typů fyzické zdatnosti pacientů při vstupu a při kontrole

Podobně jako fyzická zdatnost se při vstupu i při kontrole zadávala také ekonomická aktivita. Také zde je velký nárůst pacientů, u kterých později lékař data nezadal (z 1 na 20 % pacientů). Pouze 21 % pacientů ukončilo výživu a přijímá potravu per os nebo domácí enterální výživou (DEV), 33 % pacientů umřelo (viz Tab. č. 2). V našich hypotézách nás také bude zajímat typ centrálního žilního katetru, protože může mít vliv na vzniklé komplikace. Více než polovina pacientů má tunelizovaný centrální žilní katetr (CŽK) (viz Tab. č. 3).

Současný stav		
	Četnost	Relativní četnost
Pokračuje v DPV	194	45,5 %
Ukončení DPV, úmrtí	140	32,9 %
Ukončení DPV, přijímá per os či DEV	88	20,7 %
Ztracen z evidence	4	0,9 %

Typ centrálního žilního katetru		
	Četnost	Relativní četnost
Tunelizovaný CŽK	238	55,9 %
Venózní port	91	21,4 %
PICC	68	16,0 %
Jiný	20	4,7 %
Neuvedeno	9	2,1 %

Tab. č. 2 Četnosti různých aktuálních stavů pacientů

Tab. č. 3 Četnosti jednotlivých typů centrálního žilního katetru

Z některých dat jsme museli udělat novou proměnnou. Zjišťovali jsme, jestli pacient zemřel (pro Kaplan-Meierovu křivku přežití). Také jestli se u pacientů vyskytly alespoň nějaké komplikace. U 27 % pacientů jsme toto nemohli rozhodnout, protože neměli vyplněné všechny údaje (viz Tab. č. 4).

Komplikace		
	Četnost	Relativní četnost
Ano	159	37,4 %
Ne	152	35,8 %
Nelze	114	26,8 %

Tab. č. 4 Četnosti přítomnosti/nepřítomnosti komplikací

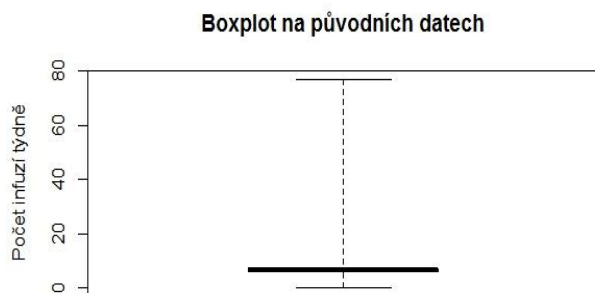
2.3.2 Kvantitativní data

Číselných proměnných se v našem datovém souboru vykytuje hned několik. Některé však nejsou vhodné k vizualizaci, jako jsou například počty různých typů komplikací, protože se vyskytují poměrně s malou frekvencí (např. infekční komplikace, žilní trombóza a další) anebo vykazují velmi malou variabilitu (např. počet hospitalizací a jiné). Proto v rámci identifikace odlehlých hodnot (viz podkapitola „2.2 Kontrola, čištění a úprava dat“) uvedeme pouze příklad počtu infuzí pacientů týdně na původním i pročištěném datovém souboru. Tuto proměnnou budeme navíc dále testovat, zda s rostoucím počtem infuzí pacientů stoupá i počet či výskyt jejich zdravotních komplikací (viz kapitola „3 Formulace a statistické testování hypotéz“), tedy jestli mezi nimi existuje závislost, což předpokládáme, že ano.

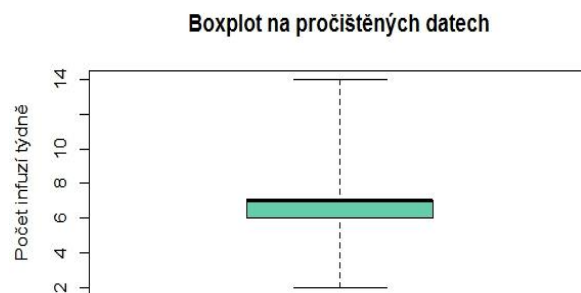
Dále se zaměříme ještě na dvě proměnné a těmi jsou věk pacientů k datu zahájení výživy, který opět budeme testovat s výskytem zdravotních komplikací, a zejména délka léčby pacientů v letech – ta bude mít klíčovou roli v analýze přežití pacientů.

Počet infuzí pacientů týdně je příkladem diskrétní proměnné a znázorňuje ji krabicový graf na původních a na pročištěných datech, pro kompletnost jsou oba grafy doplněné základními popisnými statistikami. Z boxplotu na nepročištěných datech si můžeme všimnout, že horní fousek dosahuje až po maximum (100% kvantil), kterým je odlehlá hodnota 77 infuzí (viz Obr. č. 3), a krabice značící 50 % hodnot je v tomto případě poněkud zmenšená. Naopak z grafu na pročištěných datech je vidět maximální a minimální (0% kvantil) hodnota 14 a 2 infuzí i poloha mediánu – středu rozdělení (50% kvantilu) na hodnotě 7, krabice v tomto případě i pěkně znázorňuje tzv. kvartilové rozpětí (25% až 75% kvantil), které je 6 až 7 infuzí (viz Obr. č. 4). Všechny hodnoty sumarizuje Tab. č. 6, z které je patrná poněkud nižší hodnota průměru oproti mediánu, což je způsobeno rozložením hodnot pod mediánem více k minimu, a proto je průměr menší. Při srovnání grafů i tabulek dojdeme k závěru, že maximální hodnota 77 infuzí týdně je opravdu velmi vzdálena od hodnot ostatních (viz maximum 14 na pročištěném souboru) a navíc je prakticky i hodně nereálná (průměrně 11 infuzí denně) a jelikož by mohla významně zkreslit naše výsledky, tak byla odstraněna. Dále si můžeme všimnout, že medián se odlehlou hodnotou nezměnil, ale průměr se mírně zvýšil, což značí, jak moc je průměr citlivý na hodně odlehlá pozorování, i když se vyskytovali v původním souboru pacienti s nulovým počtem infuzí (viz minimum), které jsme museli překódovat na prázdnou hodnotu.

Obr. č. 3 Krabíkový graf počtu infuzí týdně na původních datech



Obr. č. 4 Krabíkový graf počtu infuzí týdně na pročištěných datech



Popisné statistiky	
Průměr	6,6
Medián	7,0
Minimum	0,0
Maximum	77,0

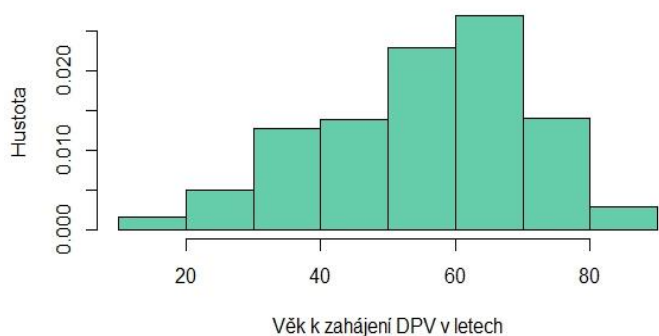
Tab. č. 5 Popisné statistiky počtu infuzí týdně na původních datech

Popisné statistiky	
Průměr	6,5
Medián	7,0
Minimum	2,0
Maximum	14,0

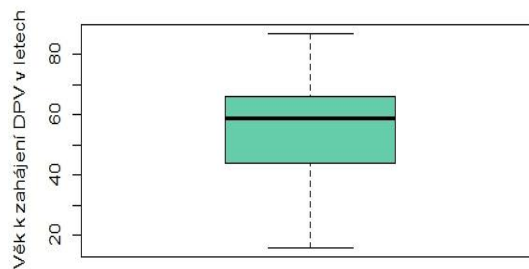
Tab. č. 6 Popisné statistiky počtu infuzí týdně na pročištěných datech

Věk pacientů k datu zahájení DPV je vizualizován na následujícím krabíkovém grafu i histogramu a je také doplněn tabulkou s popisnými statistikami. V histogramu je rozdělen do desetiletých věkových kategorií, z kterých můžeme vidět, že nejčastější je kategorie 60–70 let a nejméně častá do 20 let (viz Obr. č. 5). Vyskytují se i pacienti mladší 18 let (viz minimum), kteří dosáhli v průběhu léčby plnoletosti a byli tedy přeraženi do naší databáze dospělých. Histogram dále svým rozložením připomíná křivku hustoty symetrického normálního rozdělení. Boxplot pěkně vymezuje svými fousky rozmezí hodnot od minima (16) po maximum (87), znázorňuje medián (59) a krabice kvartilové rozpětí (cca od 45 do 65) (viz Obr. č. 6). Dále můžeme zmínit, že průměr se v případě symetrického rozdělení blíží k mediánu, v našem případě je trochu menší než medián (viz Tab. č. 7).

Obr. č. 5 Histogram věku k datu zahájení DPV



Obr. č. 6 Krabicový graf věku k datu zahájení DPV

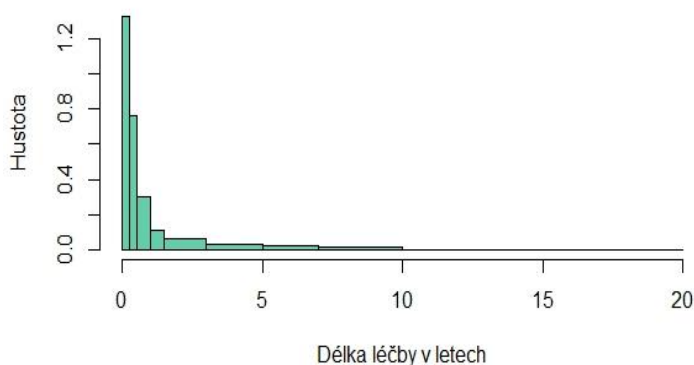


Popisné statistiky			
Průměr	Medián	Minimum	Maximum
55,7	59,0	16,0	87,0

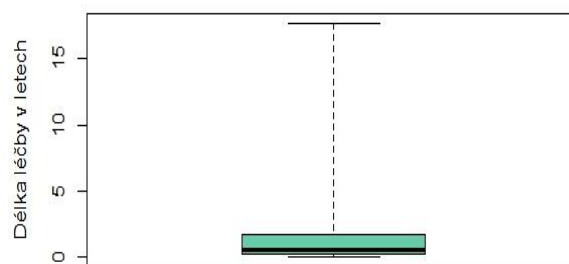
Tab. č. 7 Popisné statistiky věku k datu zahájení DPV v letech

Délka léčby pacientů v letech je vztažena k datu úmrtí, pokud pacient ukončil výživu a umřel, nebo k datu ukončení DPV, pokud ukončil DPV a pokračuje dále s tekutou stravou, anebo k datu aktualizace záznamů v registru, pokud pacient pokračuje s DPV. Vizualizaci této spojitě proměnné můžeme vidět na grafech níže s přiloženou tabulkou popisných statistik. Oproti věku má délka léčby značně asymetrické rozložení hodnot, kdy s rostoucí délkou léčby exponenciálně klesá četnost pacientů, jak pozorujeme v histogramu (viz Obr. č. 7). Vhodnější je tedy popis této proměnné pomocí mediánu (0,5) než pomocí průměru (1,72), který je více než 3krát větší než medián, protože je hodně ovlivněn málo četnými vyššími hodnotami věku. V histogramu je znázorněna relativní četnost pacientů na jednotku délky léčby v letech, jednotlivé kategorie se postupně zvětšují, ale jsou plně srovnatelné, obsah plochy histogramu je tedy roven 1 stejně jako pod hustotou rozdělení. Z boxplotu opět vidíme pěkně rozložení hodnot od minima (0) po maximum (17,7), kdy pouhá $\frac{1}{4}$ všech hodnot je větší než cca 1,5 let (viz Obr. č. 8). Jinými slovy asi $\frac{3}{4}$ pacientů se léčí DPV nejdéle do asi 1,5 let. Popisné statistiky opět shrnuje Tab. č. 8.

Obr. č. 7 Histogram délky léčby pacientů v letech



Obr. č. 8 Krabicový graf délky léčby pacientů v letech



Popisné statistiky			
Průměr	Medián	Minimum	Maximum
1,7	0,5	0,0	17,7

Tab. č. 8 Popisné statistiky délky léčby pacientů v letech

Počet komplikací pacientů jsme kvůli srovnatelnosti sumarizovali vzhledem k délce léčby, kdy jsme přepočítali celkový počet komplikací na jednotku délky léčby – tedy na jeden rok. Avšak logicky tento postup nebylo možné provést u všech pacientů, použili jsme jen záznamy s vyplněným počtem komplikací a s vyplněnou nenulovou délkou léčby. Výsledné popisné statistiky znázorňuje Tab. č. 9 níže. Největší vypovídací hodnotu má statistika průměr, která udává průměrný počet komplikací pacientů na 1 rok léčby.

Popisné statistiky				
Celkový počet	Průměr	Medián	Minimum	Maximum
233,0	2,0	0,0	0,0	43,3

Tab. č. 9 Popisné statistiky počtu komplikací vztahované na délku léčby

3 Formulace a statistické testování hypotéz

Tato kapitola vás seznámí s několika hypotézami, které jsme si zvolili k otestování v programu SPSS Statistics. Naším úkolem bylo vytipovat rizikové faktory, které by mohly ovlivňovat vznik komplikací. Zajímali jsme se především o to, jestli pohlaví, fyzická zdatnost pacientů, hlavní zdravotní problém, počet infuzí týdně a typ centrálního žilního katetru mají vliv na výskyt zdravotních komplikací u pacientů s DPV.

3.1 Formulace hypotéz

Hypotéza č. 1

H_0 : Výskyt zdravotních komplikací je nezávislý na pohlaví.

Zapsáno pomocí pravděpodobností: $p_{ij} = p_i * p_j$, kde p_{ij} je očekávaná sdružená pravděpodobnost nastání všech možných variant obou proměnných, zatímco p_i a p_j představují marginální pravděpodobnosti nastání jednotlivých kategorií (i – kategorie komplikací a j – kategorie pohlaví).

H_1 : Zdravotní komplikace jsou závislé na pohlaví.

$$p_{ij} \neq p_i * p_j$$

Hypotéza č. 2

H_0 : Výskyt komplikací je vůči kategoriím fyzické zdatnosti pacientů nezávislý, tedy pravděpodobnosti výskytu fyzické zdatnosti a komplikací se vzájemně neovlivňují.

$$p_{ij} = p_i * p_j$$

H_1 : Přítomnost zdravotních komplikací u pacientů ovlivňují kategorie fyzické zdatnosti. Jinými slovy: jsou závislé.

$$p_{ij} \neq p_i * p_j$$

Hypotéza č. 3

H_0 : Výskyt zdravotních komplikací není ovlivňován typem hlavního zdravotního problému. Oba jevy jsou vzájemně nezávislé.

$$p_{ij} = p_i * p_j$$

H_1 : Výskyt zdravotních komplikací je ovlivňován typem hlavního zdravotního problému.

$$p_{ij} \neq p_i * p_j$$

Hypotéza č. 4

H_0 : Výskyt zdravotních komplikací nesouvisí s použitím CŽK.

$$p_{ij} = p_i * p_j$$

H_1 : Výskyt zdravotních komplikací závisí na použitém CŽK.

$$p_{ij} \neq p_i * p_j$$

Hypotéza č. 5

H_0 : Výskyt infekcí nesouvisí s použitým CŽK.

$$p_{ij} = p_i * p_j$$

H_1 : Výskyt infekcí je ovlivněn použitým CŽK.

$$p_{ij} \neq p_i * p_j$$

Hypotéza č.6

H_0 : Počet infuzí týdně nesouvisí s výskytem komplikací. Jinými slovy, pacienti včetně nebo bez komplikací se statisticky neliší v počtu infuzí. Matematicky řečeno, jejich distribuční funkce se rovnají: $F(x) = F(y)$.

H_1 : Výskyt komplikací je ovlivněn počtem infuzí, tedy pacienti obou kategorií (přítomnost/nepřítomnost komplikací) se mezi sebou statisticky liší v počtu infuzí a tedy i odpovídající distribuční funkce jsou rozdílné: $F(x) \neq F(y)$.

3.2 Statistické testování hypotéz

Pro otestování nulové hypotézy č. 1, 2, 3, 4 a 5 jsme použili Pearsonův chí-kvadrát test. U poslední 6. hypotézy jsme použili neparametrický Mannův-Whitneyho test. Do testování jsme nezahrnovali neuvedené hodnoty. U určitých proměnných s nízkým zastoupením jsme kvůli splnění předpokladů daných testů logicky sloučili některé kategorie.

Ověření hypotézy č. 1

Pohlaví	Komplikace				Celkem
	Ano		Ne		
	Počet	Podíl	Počet	Podíl	Počet
Žena	96	60,0 %	92	60,5 %	188
Muž	64	40,0 %	60	39,5 %	124
Celkem	160	100,0 %	152	100,0 %	312

Tab. č. 10 Pozorované četnosti pohlaví v závislosti na komplikacích

	Hodnota testové statistiky	Stupně volnosti	P-hodnota
Pearsonův Chí-kvadrát test	0,009	1	0,924

Pozn.: všechny buňky mají očekávanou četnost větší než 5. Nejmenší očekávaná četnost je 60,41.

Tab. č. 11 Chí-kvadrát test pro hypotézu č. 1

Z uvedené Tab. č. 11 je patrné, že je splněn předpoklad Pearsonova chí-kvadrát testu, v našem případě má 100 % hodnot očekávanou četnost větší než 5. Hodnota testové statistiky je menší než kritická hodnota (95% kvantil chí-kvadrát rozdělení pro 1 stupeň volnosti), která je 3,84, a proto nezamítáme nulovou hypotézu o nezávislosti těchto dvou proměnných na 5% hladině významnosti. Nepřijetí alternativní hypotézy ukazuje i příslušná p-hodnota testové statistiky, která je podstatně větší než hladina významnosti (riziko) $\alpha = 0,05$.

Ověření hypotézy č. 2

Fyzická zdatnost se zadávala u pacientů při vstupu do registru a při kontrole (viz „2.3.1 Kvalitativní data“), proto jsme otestovali obě proměnné.

Fyzická zdatnost pacienta	Komplikace				Celkem	
	Ano		Ne		Počet	Podíl
	Počet	Podíl	Počet	Podíl		
0 – neomezená aktivita	37	23,3 %	30	19,7 %	67	21,5 %
1 – s příznaky, ale plně mobilní	71	44,6 %	65	42,8 %	136	43,7 %
2 – větší část dne mimo lůžko	34	21,4 %	32	21,1 %	66	21,2 %
3 – větší část dne na lůžku	13	8,2 %	21	13,8 %	34	10,9 %
4 – trvale upoután na lůžko	4	2,5 %	4	2,6 %	8	2,6 %
Celkem	159	100,0 %	152	100,0 %	311	100,0 %

Tab. č. 12 Pozorované četnosti fyzické zdatnosti pacientů při vstupu do registru v závislosti na komplikacích

	Hodnota testové statistiky	Stupně volnosti	P-hodnota
Pearsonův Chí-kvadrát test	2,783	4	0,595

20 % buněk má očekávanou četnost menší než 5. Nejmenší očekávaná četnost je 3,91.

Tab. č. 13 Chí-kvadrát test pro hypotézu č. 2 a.

Z Tab. č. 13 je zřejmé, že předpoklad pro použití chí-kvadrát testu je splněn hraničně. P-hodnota je větší než riziko $\alpha = 0,05$, proto nulovou hypotézu o nezávislosti opět na 5% hladině významnosti nezamítáme.

Fyzická zdatnost pacienta	Komplikace				Celkem	
	Ano		Ne		Počet	Podíl
	Počet	Podíl	Počet	Podíl		
0 – neomezená aktivita	35	23,0 %	30	21,2 %	65	22,1 %
1 – s příznaky, ale plně mobilní	60	39,5 %	51	35,9 %	111	37,8 %
2 – větší část dne mimo lůžko	31	20,4 %	33	23,2 %	64	21,7 %
3 – větší část dne na lůžku	20	13,2 %	22	15,5 %	42	14,3 %
4 – trvale upoután na lůžko	6	3,9 %	6	4,2 %	12	4,1 %
Celkem	152	100,0 %	142	100,0 %	294	100,0 %

Tab. č. 14 Pozorované četnosti fyzické zdatnosti pacientů při kontrole v závislosti na komplikacích

	Hodnota testové statistiky	Stupně volnosti	P-hodnota
Pearsonův Chí-kvadrát test	0,933	4	0,920

Pozn.: všechny buňky mají očekávanou četnost větší než 5. Nejmenší očekávaná četnost je 5,80.

Tab. č. 15 Chí-kvadrát test pro hypotézu č. 2 b.

Z Tab. č. 15 je zřejmé, že podmínka pro použití tohoto testu je splněna. P-hodnota příslušná testové statistice je větší než riziko $\alpha = 0,05$, proto nulovou hypotézu stejně jako v předchozím případě na 5% hladině významnosti nezamítáme. Tedy jinými slovy jsme neprokázali, že by úroveň fyzické aktivity pacienta zaznamenaná při kontrole i při vstupu do registru nějak ovlivňovala výskyt komplikací, což je v rozporu s tím, co jsme si mysleli.

Ověření hypotézy č. 3

Dominantní problém pacienta	Komplikace				Celkem	
	Ano		Ne			
	Počet	Podíl	Počet	Podíl	Počet	Podíl
Nádor a jeho léčba	47	29,9 %	79	52,3 %	126	40,9 %
Nenádorové stavy	28	17,8 %	24	16,0 %	52	16,9 %
Crohnova nemoc	25	15,9 %	24	16,0 %	49	15,9 %
Mesenterický infarkt	23	14,6 %	6	3,9 %	29	9,4 %
Radiační enteritida	13	8,3 %	6	3,9 %	19	6,2 %
Pseudoobstrukce	3	1,9 %	3	1,9 %	6	1,9 %
Jiné	18	11,6 %	9	6,0 %	27	8,8 %
Celkem	157	100,0 %	151	100,0 %	308	100,0 %

Tab. č. 16 Pozorované četnosti dominantního problému v závislosti na komplikacích

	Hodnota testové statistiky	Stupně volnosti	P-hodnota
Pearsonův Chí-kvadrát test	23,892	6	0,001

Pozn.: 14,3 % buněk má očekávanou četnost menší než 5. Nejmenší očekávaná četnost je 2,94.

Tab. č. 17 Chí-kvadrát test pro hypotézu č. 3

Předpoklady u zvoleného Pearsonova testu jsou splněny, protože všechny hodnoty mají očekávanou četnost větší než 2 a 14,3 % očekávaných hodnot je menší než 5. Původně toto splněno nebylo, a proto jsme k jiné připojili i nenádorovou pankreatopatii. Po spojení má již test splněné předpoklady (viz Tab. č. 17). P-hodnota je zde menší než hladina významnosti $\alpha = 0,05$, a tedy zamítáme nulovou hypotézu o nezávislosti na 5% hladině významnosti. Jak jsme předpokládali, hlavní zdravotní problém pacienta má vliv na přítomnost komplikací. Při porovnání jednotlivých typů dominantního problému (viz Tab. č. 16) si můžeme všimnout, že větší podíl výskytu komplikací v porovnání s ostatními problémy se objevuje například u pacientů s mesenterickým infarktem nebo s radiační enteritidou.

Ověření hypotézy č. 4

Typ CŽK	Komplikace				Celkem	
	Ano		Ne			
	Počet	Podíl	Počet	Podíl	Počet	Podíl
Tunelizovaný CŽK	102	64,1 %	81	53,7 %	183	59,1 %
Venózní port	27	17,0 %	40	26,5 %	67	21,6 %
PICC	26	16,4 %	23	15,2 %	49	15,8 %
Jiný	4	2,5 %	7	4,6 %	11	3,5 %
Celkem	159	100,0 %	151	100,0 %	310	100,0 %

Tab. č. 18 Pozorované četnosti typů CŽK v závislosti na komplikacích

	Hodnota testové statistiky	Stupně volnosti	P-hodnota
Pearsonův Chí-kvadrát test	5,731	3	0,125

Pozn.: všechny buňky mají očekávanou četnost větší než 5. Nejmenší očekávaná četnost je 5,36.

Tab. č. 19 Chí-kvadrát test pro hypotézu č. 4

Pod Tab. č. 19 si můžeme všimnout, že předpoklad pro použití Pearsonova testu je splněn, protože 100 % hodnot má očekávanou četnost větší než 5. P-hodnota je vyšší než hladina významnosti $\alpha = 0,05$, proto nulovou hypotézu o nezávislosti nezamítáme na 5% hladině významnosti.

Ověření hypotézy č. 5

Typ CŽK	Infekční komplikace				Celkem	
	Ano		Ne			
	Počet	Podíl	Počet	Podíl	Počet	Podíl
Tunelizovaný CŽK	21	47,7 %	217	58,2 %	238	57,1 %
PICC	16	36,4 %	52	13,9 %	68	16,3 %
Venózní port	5	11,4 %	86	23,1 %	91	21,8 %
Jiný	2	4,5 %	18	4,8 %	20	4,8 %
Celkem	44	100,0 %	373	100,0 %	417	100,0 %

Tab. č. 20 Pozorované četnosti typů CŽK v závislosti na infekčních komplikacích

	Hodnota testové statistiky	Stupně volnosti	P-hodnota
Pearsonův Chí-kvadrát test	15,360	3	0,002

Pozn.: 12,5% buněk má očekávanou četnost menší než 5. Nejmenší očekávaná četnost je 5,36.

Tab. č. 21 Chí-kvadrát test pro hypotézu č. 5

Z výše uvedené Tab. č. 21 je zřejmé, že podmínky pro použití chí-kvadrát testu jsou splněny. P-hodnota je menší než riziko $\alpha = 0,05$, proto nulovou hypotézu v tomto případě o nezávislosti typu centrálního žilního katetru a vzniku infekčních komplikací zamítáme na 5% hladině významnosti. Největší statistický rozdíl podílu přítomnosti infekčních komplikací u pacientů s různým typem CŽK se vyskytuje mezi katetrem PICC a Venózním portem, kdy podíl výskytu infekčních komplikací u pacientů s PICC je více jak čtyřnásobně větší (viz Tab. č. 20). Musíme ale poznamenat, že infekční komplikace se vyskytují pouze u 11 % pacientů a závislost proměnných zde byla prokázána na rozdíl od předchozí hypotézy, když jsme sledovali přítomnost různých druhů zdravotních komplikací u pacientů s různým typem CŽK. Z toho plyne, že pokud někdy blíže specifikujeme určitou proměnnou (komplikace => infekční komplikace), může se zvětšit diference neagregovaných četností (máme více informace), stejně jako v tomto případě.

Ověření hypotézy č. 6

Komplikace	Počet pacientů s infuzemi	Součet pořadí počtu infuzí	Medián
Ano	156	23873	7
Ne	141	20380	7
Celkem	297		

Tab. č. 22 Sumarizace počtu infuzí týdně u pacientů s/bez komplikací

	Testová statistika	P-hodnota
Mann-Whitneyho test	10369	0,293

Tab. č. 23 Mannův-Whitneyho test pro hypotézu č. 6

Tuto hypotézu jsme otestovali pomocí Mann-Whitneyho testu. Tento neparametrický test porovnává distribuční funkce dvou skupin – v našem případě pacienty včetně a bez komplikací, za účelem odhalení statistického rozdílu v počtu infuzí mezi těmito kategoriemi. Výsledná testová statistika se počítá pomocí součtu pořadí hodnot počtu infuzí v jednotlivých kategoriích (viz Tab. č. 22). Jediným předpokladem tohoto testu je stejné rozdělení dat v obou souborech, což je mnohem slabší předpoklad než normalita dat v případě např. parametrických t-testů. Použili jsme neparametrický test, protože jsme zjistili, že počet infuzí týdně nemá normální rozdělení (Kolmogorovův-Smirnovův test, kde p-hodnota byla menší než 0,001, a grafické ověření pomocí histogramu, boxplotu i Q-Q grafu). Jelikož je p-hodnota příslušná testové statistice větší jak riziko $\alpha = 0,05$, tak nulovou hypotézu o rovnosti distribučních funkcí v obou kategoriích nezamítáme na 5% hladině významnosti. Neprokázali jsme tedy statistický rozdíl v počtu infuzí u pacientů včetně a bez komplikací pravděpodobně v důsledku nedostatku informací v našich datech. Tento závěr je poměrně překvapivý, jelikož jsme předpokládali, že u pacientů se zdravotními komplikacemi se bude vykytovat také větší počet infuzí a naopak, což se nám ale nepodařilo ve výběrovém vzorku pacientů s DPV dokázat.

4 Analýza přežití pacientů s DPV

V této kapitole budeme hodnotit čas (délku léčby) do výskytu sledované události, kterou bude v našem případě úmrtí pacientů s DPV. Nezajímá nás tedy pouhá četnost výskytu úmrtí pacientů, ale také časový průběh tohoto výskytu.

Nejprve probereme nejdůležitější pojmy analýzy přežití v kontextu s naším datovým souborem. Dále se zaměříme na funkci přežití a popíšeme si základní informace o mediánu přežití a průměrném přežití. Následovat bude popis nejznámějšího a nejpoužívanějšího neparametrického odhadu funkce přežití, kterým je Kaplan-Meierův odhad. Tuto metodu nakonec aplikujeme na náš experimentální vzorek v software SPSS Statistics a uvedeme výstupy v podobě sumarizace, odhadů popisných statistik přežití (včetně intervalu spolehlivosti pro odhad průměru přežití) a křivky přežití pacientů s DPV.

4.1 Základní pojmy

Klíčový termín analýzy přežití je **sledovaná událost**, ta je v našem případě dána jednoznačně – úmrtí pacientů s DPV. Tuto proměnnou jsme si již definovali v podkapitole „2.2 Kontrola, čištění a úprava dat“ a pro připomenutí je výskyt úmrtí pacienta vyjádřen hodnotou 1. Dále je poměrně důležitou složkou v hodnocení času přežití také **počáteční bod sledování**, od kterého se počítá výsledný čas přežití. Jelikož čas přežití bude představovat délka léčby pacientů v letech, tak počátečním bodem sledování je samozřejmě datum zahájení léčby – DPV.

Cenzorování je časová složka hrající velmi důležitou roli v analýze přežití, protože nám působí určité výpočetní potíže. Cenzorovaný je takový čas přežití pacientů, u nichž v průběhu léčby nedošlo k úmrtí. Tedy přesněji, nevíme, kdy a jestli vůbec pacient umřel v čase po ukončení jeho sledování, víme jen, že žije před ukončením jeho sledování (datum ukončení DPV nebo datum aktualizace záznamů v registru). Naopak, pokud pacient v průběhu léčby umřel – došlo u něj ke sledované události, tak jeho čas přežití je necenzorovaný (kompletní). Výše popsáný typ cenzorování se nazývá **cenzorování zprava**, existují i další druhy, avšak pro naše účely je zmiňovat nebudeme.

Cenzorování způsobuje vlastně ztrátu určité informace, protože nemáme kompletní údaje ze sledování pacientů se zaznamenaným úmrtím, takže nelze použít standardní

statistické testy, které berou každou hodnotu za konstantní. Procento cenzorovaných bývá měřítkem kvality sledování daného souboru. Vysoké procento cenzorovaných časů znamená, že délka sledování byla nedostatečná. Na našem souboru by šla ovlivnit například dalšími kontrolami záznamů v registru, kdy by se s rostoucí délkou léčby k tomuto datu zvyšovala i pravděpodobnost úmrtí, takže by u více pacientů došlo ke sledované události. Naopak nízké procento cenzorovaných časů přežití pak odráží dostatečnou dobu sledování. Dostatečný počet kompletních časů přežití se dá také ovlivnit větší velikostí souboru. [5]

Funkce přežití (*Survival function*) $S(t)$ je definovaná pomocí distribuční funkce $F(t)$ a vyjadřuje pravděpodobnost, že čas přežití daného pacienta T bude větší, než je pevně zvolený čas t . Naproti tomu distribuční funkce znamená pravděpodobnost, že čas přežití daného pacienta T bude menší nebo roven hodnotě t . Níže je uvedena matematická definice funkce přežití – viz (1). Je zřejmé, že funkce přežití nabývá pouze hodnot mezi 1 a 0 a je vždy funkcí nerostoucí. [5]

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

Medián přežití $t_{0,5}$ je jednou z hlavních číselných charakteristik v analýze přežití a je definován jako čas, ve kterém má funkce přežití hodnotu 0,5 – viz (2). Jinými slovy je to 50% kvantil času přežití. V analýze přežití je právě většinou preferována hodnota mediánu přežití kvůli snadné interpretaci. Lze definovat samozřejmě i další kvantily. [5]

$$S(t_{0,5}) = 0,5; S(t_{0,25}) = 0,75 \text{ a } S(t_{0,75}) = 0,25 \quad (2)$$

Průměrná doba přežití μ představuje střední hodnotu náhodné veličiny T , která udává délku přežití. Po několika úpravách a s využitím vztahů výše lze zjistit, že průměrné přežití je definováno jako integrál z funkce přežití na intervalu od nuly do nekonečna – viz (3). Pro existenci střední hodnoty však musí platit, aby funkce přežití s rostoucím časem nabyla nulové hodnoty – tedy podíl žijících pacientů je nulový. Tento předpoklad však v mnoha případech není splněn, proto se snažíme odhadnout střední hodnotu pomocí podmíněné pravděpodobnosti (hustoty a funkce přežití), kdy alespoň někdy v budoucnu předpokládáme výskyt sledované události (úmrtí). [5]

$$\mu = E(T) = \int_0^{\infty} tf(t)dt = \int_0^{\infty} S(t)dt \quad (3)$$

4.2 Neparametrický Kaplan-Meierův odhad funkce přežití

Neparametrické odhady, na rozdíl od parametrických, nevyžadují žádné zvláštní předpoklady ohledně rozdělení pravděpodobnosti náhodné veličiny, proto je jejich aplikace široká v analýze přežití, kdy je znalost rozdělení velmi omezená. Avšak parametrické metody jsou obecně jednodušší a při správně identifikovaném rozdělení i přesnější. Podmínka korektní specifikace rozdělení je ale velmi důležitá a při jejím nesplnění může dojít ke značnému zkreslení výsledků. [5]

Kaplan-Meierův odhad $\hat{S}(t)$ je standardní neparametrická metoda pro hodnocení přežití v klinických studiích. Metoda výpočtu je poměrně jednoduchá, nejprve musíme odhadnout podmíněnou pravděpodobnost přežití \hat{p}_i pro každý časový interval, což znamená podělit počet pacientů, kteří zůstávají žít ($R_i - d_i$), kde d_i je počet úmrtí v čase t_i , počtem pacientů, kteří zůstali žít před tímto intervalem v čase t_{i-1} – jsou tedy aktuálně v riziku výskytu úmrtí v čase t_i a značí se R_i . Následuje výpočet odhadu kumulativní pravděpodobnosti přežití $\prod_{t_i \leq t} \hat{p}_i$, která nám udává, jaký podíl pacientů ze všech v určitém čase t stále žije, což je vlastně odhad funkce přežití $\hat{S}(t)$.

Kaplan-Meierův odhad řeší cenzorování tak, že tito pacienti vypadávají ze skupiny pacientů v riziku ihned po zaznamenaném čase cenzorování. [5] Je-li tedy nějaký čas cenzorovaný, pak je daný pacient v určitém čase, kdy je cenzorován, započítán do skupiny pacientů v riziku R_i (víme, že k této době žije), ale v následujícím čase ho již do skupiny v riziku R_{i+1} nezahrnujeme (nevíme, jestli v této době ještě pořád žije nebo jestli umřel).

Výsledný vzorec lze po úpravě pro Kaplan-Meierův odhad funkce přežití zapsat takto

$$\hat{S}(t) = \prod_{t_i \leq t} \hat{p}_i = \prod_{t_i \leq t} \frac{R_i - d_i}{R_i} = \prod_{t_i \leq t} \left(1 - \frac{d_i}{R_i}\right). \quad (4)$$

Pro odhad rozptylu Kaplan-Meierova odhadu funkce přežití $\text{var}(\hat{S}(t))$ se využívá tzv. **Greenwoodův vzorec** – viz (5), který je zabudován ve statistických softwarech. Pro kompletnost této práce je uveden níže. Rozptyl potřebujeme získat proto, abychom mohli zkonstruovat interval spolehlivosti pro Kaplan-Meierův odhad funkce přežití $\hat{S}(t)$.

$$\text{var}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{t_i \leq t} \frac{d_i}{R_i(R_i - d_i)} \quad (5)$$

Pro konstrukci $100(1-\alpha)\%$ **intervalu spolehlivosti** se využívá aproximace normálním rozdělením z platnosti centrální limitní věty (CLV), kde $z_{1-\alpha/2}$ je $100(1-\alpha/2)\%$ kvantil $N(0,1)$, $\sqrt{\text{var}(\hat{S}(t))}$ se obvykle ve statistických softwarech označuje jako standardní chyba odhadu $SE(\hat{S}(t))$. Výsledný $100(1-\alpha)\%$ interval spolehlivosti pro Kaplan-Meierův odhad pravděpodobnosti přežití v čase t můžeme zformulovat takto

$$\left(\hat{S}(t) - z_{1-\frac{\alpha}{2}} \sqrt{\text{var}(\hat{S}(t))}; \hat{S}(t) + z_{1-\frac{\alpha}{2}} \sqrt{\text{var}(\hat{S}(t))} \right). \quad (6)$$

Musíme však konstatovat, že tato podoba intervalu spolehlivosti má určité nevýhody. Je sice početně poměrně jednoduchý, ale je také symetrický, což znamená, že může dojít k situaci, kdy bude nabývat hodnoty přežití větší než 1 nebo naopak menší než 0, což je samozřejmě nežádoucí. Tomu se můžeme vyhnout tím, že transformujeme odhad $\hat{S}(t)$ (např. pomocí logaritmické transformace) na hodnoty z intervalu $(-\infty, \infty)$. [5] Tuto konstrukci intervalu spolehlivosti však pro účely této práce už uvádět nebudeme.

4.3 Aplikace Kaplan-Meierovy metody

Výše popsaný odhad funkce přežití podle Kaplan-Meiera jsme použili na náš soubor pacientů s DPV a zde jsou uvedeny výstupy této analýzy.

Nejprve shrneme, že celkem jsme pro analýzu přežití nepoužili všechny pacienty. Jak bylo řečeno v předchozí podkapitole „2.2 Kontrola, čištění a úprava dat“, museli jsme vynechat pacienty, u kterých nebylo možné určit délku léčbu z důvodu nedostatku informací (pacient „Ztracen z evidence“, nevyplněná data ukončení výživy či úmrtí). Dále máme asi třetinu pacientů, kteří umřeli, tedy u nich došlo ke sledované události – jsou necenzorovaní. Podíl cenzorovaných pacientů činí v našem vzorku asi 67 %, což je poměrně velký podíl žijících pacientů. Všechna tato data sumarizuje Tab. č. 24 níže.

Celkový počet pacientů	Počet úmrtí	Cenzorovaní pacienti	
		Počet	Podíl
416	136	280	67 %

Tab. č. 24 Sumarizace dat pro analýzu přežití pomocí Kaplan-Meierovy křivky

Tabulku přežití pacientů s odhady kumulativních pravděpodobností přežití v určitém čase a se standardními chybami uvádět nebudeme. Jen poznamenáme, že pomocí této tabulky a kvantilů standardizovaného normálního rozdělení by se dal zkonstruovat interval spolehlivosti pro jakýkoliv Kaplan-Meierův odhad funkce přežití v čase.

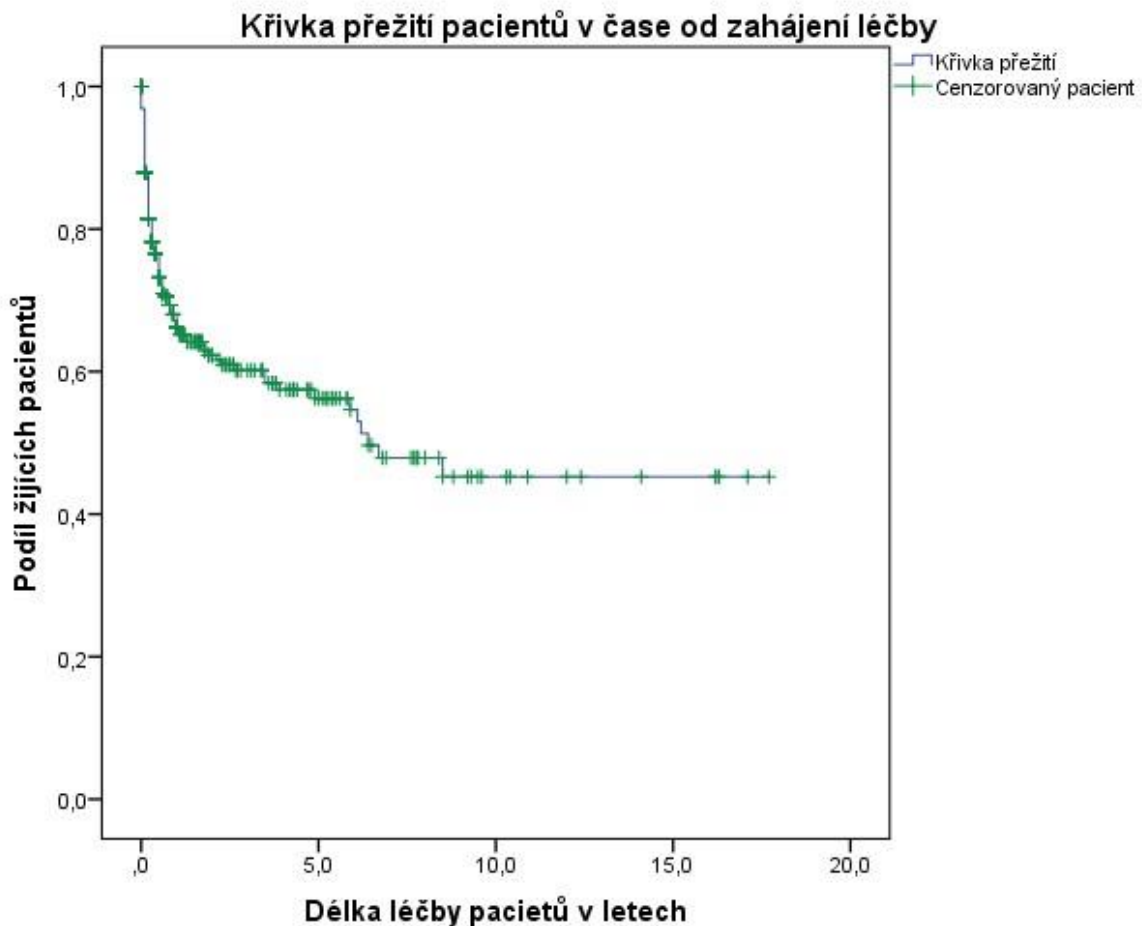
Dvě nejdůležitější charakteristiky analýzy přežití – průměr a medián přežití sumarizuje následující Tab. č. 25. Odhad průměru se standardní chybou je doplněn i o 95% interval spolehlivosti, avšak odhad mediánu není doplněn o standardní chybu a ani o interval spolehlivosti, protože pravděpodobně z nedostatku informací je nebylo možné určit (hodnota mediánu byla hraniční – je tedy málo věrohodná) – nejnižší podíl přežití byl pouze 0,45, tedy těsně pod hodnotou odpovídající mediánu přežití (0,5). Můžeme si všimnout, že hodnota mediánu přežití je menší oproti průměru přežití. Pro shrnutí můžeme z hodnoty mediánu přežití říci, že v čase délky léčby 6,4 let umřelo méně než polovina pacientů, tedy v tomto čase je pravděpodobnost přežití rovna 0,5.

Průměr přežití (v letech)				Odhad mediánu přežití (v letech)
Odhad	Standardní chyba	95% interval spolehlivosti		
		Spodní mez	Horní mez	
9,134	0,655	7,851	10,418	6,400

Tab. č. 25 Číselné charakteristiky analýzy přežití – průměr přežití a medián přežití

Nakonec uvádíme vizualizaci této analýzy – křivku přežití pacientů jako Kaplan-Meierův odhad funkce přežití (viz Obr. č. 9). Křivka přežití odráží podíl žijících pacientů v daném čase od zahájení DPV a je nerostoucí (klesající i konstantní). Cenzorovaní pacienti jsou znázorněni v grafu příslušným symbolem, pokles (schod) křivky přežití nastává v čase koncové události vždy u necenzorovaného (mrtvého) pacienta a velikost tohoto schodu je dána počtem pacientů, kteří v daném časovém intervalu zůstávají v tzv. riziku. S rostoucím časem se snižuje počet pacientů v riziku, a tudíž se každý další výskyt úmrtí výrazněji promítne v podílu žijících pacientů, a proto je schod křivky přežití větší – z toho plyne i nízká věrohodnost konce křivky přežití. Maximální a minimální délka sledování je shodná s délkou léčby sumarizovanou v podkapitole „Kvantitativní data“ od 0 do 17,7 let. Celkově můžeme zmínit, že křivka je poměrně strmá, což svědčí o poměrně vysokém riziku úmrtí při rostoucí délce léčby. Už ročnímu přežití odpovídá podíl žijících pacientů 66 %, dvouletému přežití 62 % a 75 % pacientů se dožilo délky léčby 0,5 roku. Při rostoucí délce léčby k maximu je vidět, že podíl žijících pacientů se ustálí na hodnotě těsně pod polovinou přežití – 0,45. Tento fakt poměrně vysoké úmrtnosti pacientů během prvních několika let délky léčby je zapříčiněn

zejména tím, že máme v našem experimentálním vzorku i pacienty s paliativní DPV, u kterých se přímo předpokládá brzké úmrtí, protože jejich neléčitelné onemocnění je téměř v konečném stavu. V důsledku toho dále ukážeme i sumarizaci a vizualizaci dat pomocí této metody po vyřazení právě těchto pacientů s paliativní DPV a porovnáme dosažené výsledky.



Obr. č. 9 Kaplan-Meierova křivka přežití pacientů v čase od data zahájení DPV

Pro kompletnost uvádíme i **95% intervaly spolehlivosti (IS)** pro výše uvedené Kaplan-Meierovy odhady funkce přežití. První z nich pro $\hat{S}(t = 0,5) = 0,73$ vypočteme podle zmíněného vzorce (6) s rizikem $\alpha = 0,05$:

$$\begin{aligned} & \left(\hat{S}(t) - z_{0,975} \text{SE}(\hat{S}(t)); \hat{S}(t) + z_{0,975} \text{SE}(\hat{S}(t)) \right) \\ & = (0,73 - 1,96 * 0,024; 0,73 + 1,96 * 0,024) = (0,68; 0,78). \end{aligned}$$

Interpretace tohoto intervalu je následující: S 95% pravděpodobností pokryje interval neznámou hodnotu funkce přežití v čase $t = 0,5$, kterou jsme bodově odhadli podle Kaplan-Meiera. Analogicky se dají spočítat i IS pro další odhady $\hat{S}(1) = 0,66$ a $\hat{S}(2) = 0,62$:

$$(0,66 - 1,96 * 0,027; 0,66 + 1,96 * 0,027) = (0,61; 0,71)$$

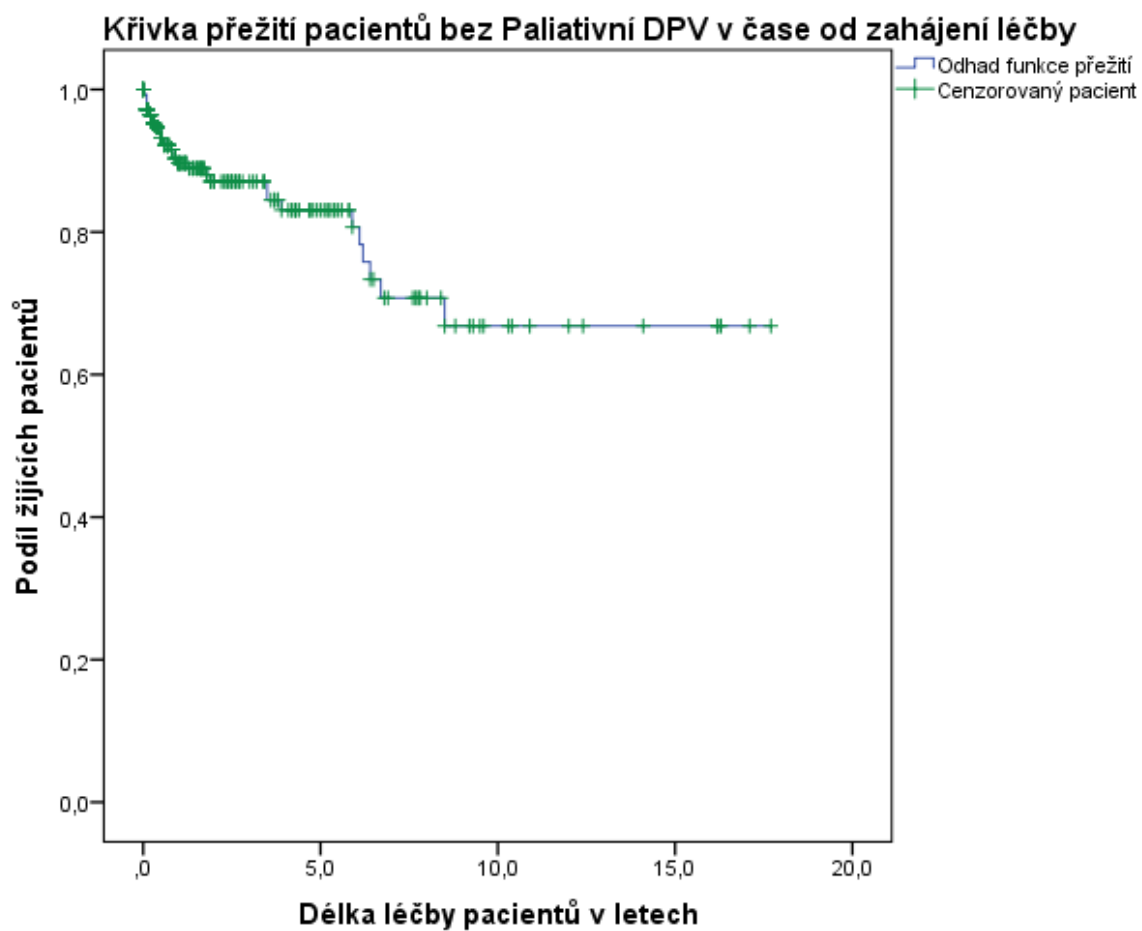
$$(0,62 - 1,96 * 0,029; 0,62 + 1,96 * 0,029) = (0,56; 0,68).$$

Nyní použijeme stejnou metodu, avšak z dat vyfiltrujeme pacienty s **paliativní DPV** a uvidíme, jak výrazně se tito pacienti podílejí na strmosti křivky přežití v prvních několika letech. Sumarizace vstupních dat pro Kaplan-Meierův neparametrický odhad funkce přežití znázorňuje Tab. č. 26, z které je vidět, že bylo vypuštěno více než 100 pacientů, takže se nám výrazně snížila velikost souboru. Dále si můžeme všimnout výrazného poklesu necenzorovaných pacientů, u kterých došlo k úmrtí a hlavně se zvýšil i podíl cenzorovaných pacientů na hodnotu blížíící se 90 %. Můžeme tedy konstatovat, že paliativní pacienti se výrazně podílí na četnosti výskytu sledované události (úmrtí) a délka sledování pacientů bez paliativní DPV je poměrně nedostatečná.

Celkový počet pacientů	Počet úmrtí	Cenzorovaní pacienti	
		Počet	Podíl
292	35	257	88 %

Tab. č. 26 Sumarizace dat bez pacientů s paliativní DPV pro Kaplan-Meierův odhad

Pokud se podíváme na graf s křivkou přežití těchto pacientů (viz Obr. č. 10) tak na první pohled zjistíme, že strmost křivky je mnohem více pozvolná a neklesne ani pod hodnotu 0,5 – proto nemůžeme odhadnout medián přežití jako čas, ve kterém je pravděpodobnost přežití 0,5. Větší pokles žijících pacientů nastává až po pěti letech od zahájení výživy. Pro matematické srovnání, půl roku délky léčby se dožilo asi 93 % pacientů ($\hat{S}(t = 0,5) = 0,93$), ročnímu přežití odpovídá podíl žijících pacientů 90 %, tedy $\hat{S}(1) = 0,9$, což je téměř o 25 % více než v předchozím případě, kde jsme zahrnovali všechny pacienty s DPV. S rostoucím časem přežití se podíl žijících pacientů ustálí na hodnotě 0,67, která je poměrně hodně vzdálena od podílu 0,5. Tuto hodnotu konce křivky ale musíme brát jako orientační z již zmíněných důvodů, kdy s rostoucím časem přežití klesá spolehlivost (věrohodnost) tohoto odhadu. Závěrem poznamenejme, že riziko úmrtí s rostoucí délkou léčby pacientů bez paliativní DPV se poměrně výrazně snížilo od situace, kdy jsme pracovali s celým souborem pacientů, jak jsme předpokládali. Tedy tímto jsme ukázali, jak může v některých případech dojít ke zkreslení prezentovaných výsledků, které mnohdy vyžadují různou interpretaci.



Obr. č. 10 Kaplan-Meierova křivka přežití pacientů bez paliativní DPV v čase od data zahájení DPV

Závěr

Tato práce je rozdělena do několika částí. V úvodní části jsme se zabývali teoretickým pozadím umělé (parenterální) výživy, kde jsme se zaměřili především na zdravotní komplikace, které se vyskytovaly v našem experimentálním vzorku pacientů s domácí parenterální výživou (DPV). Tento specifický typ umělé výživy spočívá v dlouhodobém podávání výživových roztoků přímo do krevního řečiště a zavádí se zejména u pacientů se závažným onemocněním trávicí soustavy (krátké střevo, neprůchodnost střev, záněty, nádory a mnoho dalších).

Další část byla věnována práci s datovým souborem pacientů s DPV a přípravě dat pro následné statistické testování hypotéz a analýzu přežití. Při kontrole a čištění těchto dat jsme se museli potýkat s mnoha problémy, jejichž řešení mnohdy nebylo příliš intuitivní – například jsme polemizovali nad odlehlými hodnotami kvantitativních proměnných, museli jsme kvůli zkreslení výsledků vyřadit pacienty s chybnými nebo nelogickými hodnotami, nechronologií dat a sjednotili jsme občas i neuvedené hodnoty. Zde jsme využili SQL skriptování. V následné popisné statistice jsme sumarizovali a zvizualizovali některé kategoriální a kvantitativní proměnné, zejména ty, které by mohly mít vliv na výskyt zdravotních komplikací pacientů. Pro tvorbu grafů jsme použili programy Excel, R a SPSS.

Jednou z klíčových částí bylo testování námi vybraných statistických hypotéz. Pomocí Pearsonova chí-kvadrát testu jsme ověřovali, zda vytipované kategoriální proměnné jsou závislé s výskytem zdravotních komplikací pacientů. Na 5% hladině významnosti se nám podařilo prokázat závislost pouze hlavního zdravotního problému na výskytu komplikací, jak jsme předpokládali. U centrálního žilního katetru jsme prokázali závislost až po specifikaci komplikací na infekční, kdy se zvýšil rozdíl neagregovaných četností, protože jsme měli více informace. Neprokázali jsme vliv fyzické aktivity pacientů (při vstupu i při kontrole) na výskyt zdravotních komplikací. Nakonec jsme pomocí neparametrického Mannova-Whitneyho testu porovnávali počet infuzí týdně u pacientů s komplikacemi a bez komplikací, avšak ani zde jsme neshledali významný statistický rozdíl.

Poslední nepovinnou, ale podle našeho názoru klíčovou, částí práce byla analýza přežití pacientů s DPV pomocí neparametrického Kaplan-Meierova odhadu funkce přežití, který udává podíl žijících pacientů v čase od data zahájení výživy. Zde bylo ale nutno nejprve definovat nejdůležitější pojmy této problematiky jako je třeba cenzorování, funkce přežití či odhad mediánu přežití. Po vysvětlení základního principu této metody, který spočívá

v odhadu kumulativních pravděpodobností přežití spočtených pomocí podmíněných pravděpodobností pro každý časový interval, jsme použili tuto analýzu i na náš experimentální vzorek pacientů s DPV. Kvůli zkreslení výsledků v podobě prudké strmosti křivky přežití v prvních několika letech od zahájení léčby (DPV) jsme uvedli dva výstupy této metody, kdy jsme porovnali a náležitě interpretovali dosažené výsledky. Hlavním faktorem zvyšující riziko úmrtí pacientů s rostoucím časem přežití byli pacienti s paliativní DPV, u kterých se přímo předpokládá brzké úmrtí, proto byla bez jejich vyřazení strmost křivky přežití poměrně velká. Po vyřazení těchto pacientů sice kleslo riziko úmrtí, snížila se také věrohodnost tohoto odhadu, protože se nám zvýšil podíl cenzorovaných pacientů, kteří indikují nedostatečnou dobu sledování.

Závěrem, shrneme-li zjištěné výsledky, některé předpoklady se nám podařilo z našich dat statisticky ověřit, některé hypotézy však nikoli z nejrůznějších příčin (nedostatek informací, velikost vzorku a jiné), proto rozhodně nemůžeme zobecnit závěry na celou populaci pacientů s DPV. Vždy je třeba zohlednit chybu, jaké se při testování dopouštíme, pojetí dat a reprezentativnost vzorku pro naše účely.

Seznam literatury a použitých zdrojů

- [1] CHARVÁT, Jiří; KVAPIL, Milan a kol. *Praktikum umělé výživy*: Učební texty k praktickým cvičením z umělé výživy. Praha: Karolinum, 2006, ISBN 80-246-1303-4.
- [2] KASPER, Heinrich; překlad: PROCHÁZKA, Karel. *Výživa v medicíně a dietetika*: Překlad 11. vydání. 1. české vydání. Praha: Grada Publishing, a.s., 2015, 592. ISBN 978-80-247-4533-6.
- [3] KOHOUT, Pavel; KOTRLÍKOVÁ, Eva. *Základy klinické výživy*. 1. vydání. Praha: KRIGL, 2005, ISBN 80-86912-08-6.
- [4] KOHOUT, Pavel; RUŠAVÝ, Zdeněk; ŠERCLOVÁ, Zuzana. *Vybrané kapitoly z klinické výživy I.* 1. vydání. Praha: Forsapi, s.r.o., 2010, ISBN 978-80-87250-08-2.
- [5] PAVLÍK, Tomáš. *Aplikovaná analýza přežití* [online]. [cit. 5.4.2017]. Dostupný na WWW: <http://portal.matematickabiologie.cz/index.php?pg=aplikovana-analyza-klinickych-a-biologickych-dat--aplikovana-analyza-preziti--neparametricke-odhady--kaplanuv-meieruv-odhad-funkce-preziti>
- [6] URBÁNEK, Libor; URBÁNKOVÁ, Pavla; MARKOVÁ, Jaroslava. *Klinická výživa v současné praxi*. druhé upravené. Brno: NCO NZO, 2010, 97. ISBN 978-80-7013-525-9.

Seznam tabulek

Tab. č. 1 Četnosti diagnóz pacientů	10
Tab. č. 2 Četnosti různých aktuálních stavů pacientů	11
Tab. č. 3 Četnosti jednotlivých typů centrálního žilního katetru	11
Tab. č. 4 Četnosti přítomnosti/nepřítomnosti komplikací	11
Tab. č. 5 Popisné statistiky počtu infuzí týdně na původních datech.....	13
Tab. č. 6 Popisné statistiky počtu infuzí týdně na pročištěných datech	13
Tab. č. 7 Popisné statistiky věku k datu zahájení DPV v letech	14
Tab. č. 8 Popisné statistiky délky léčby pacientů v letech	15
Tab. č. 9 Popisné statistiky počtu komplikací vztahované na délku léčby.....	15
Tab. č. 10 Pozorované četnosti pohlaví v závislosti na komplikacích	18
Tab. č. 11 Chí-kvadrát test pro hypotézu č. 1	18
Tab. č. 12 Pozorované četnosti fyzické zdatnosti pacientů při vstupu do registru v závislosti na komplikacích.....	18
Tab. č. 13 Chí-kvadrát test pro hypotézu č. 2 a.....	19
Tab. č. 14 Pozorované četnosti fyzické zdatnosti pacientů při kontrole v závislosti na komplikacích.....	19
Tab. č. 15 Chí-kvadrát test pro hypotézu č. 2 b.	19
Tab. č. 16 Pozorované četnosti dominantního problému v závislosti na komplikacích	20
Tab. č. 17 Chí-kvadrát test pro hypotézu č. 3	20
Tab. č. 18 Pozorované četnosti typů CŽK v závislosti na komplikacích.....	21
Tab. č. 19 Chí-kvadrát test pro hypotézu č. 4	21
Tab. č. 20 Pozorované četnosti typů CŽK v závislosti na infekčních komplikacích.....	21
Tab. č. 21 Chí-kvadrát test pro hypotézu č. 5	22
Tab. č. 22 Sumarizace počtu infuzí týdně u pacientů s/bez komplikací	22
Tab. č. 23 Mannův-Whitneyho test pro hypotézu č. 6	22
Tab. č. 24 Sumarizace dat pro analýzu přežití pomocí Kaplan-Meierovy křivky.....	27
Tab. č. 25 Číselné charakteristiky analýzy přežití – průměr přežití a medián přežití.....	28
Tab. č. 26 Sumarizace dat bez pacientů s paliativní DPV pro Kaplan-Meierův odhad	30

Seznam obrázků

Obr. č. 1 Četnosti jednotlivých typů dominantních problémů pacientů.....	10
Obr. č. 2 Srovnání četností typů fyzické zdatnosti pacientů při vstupu a při kontrole	11
Obr. č. 3 Krabíkový graf počtu infuzí týdně na původních datech	13
Obr. č. 4 Krabíkový graf počtu infuzí týdně na pročištěných datech.....	13
Obr. č. 5 Histogram věku k datu zahájení DPV	14
Obr. č. 6 Krabíkový graf věku k datu zahájení DPV	14
Obr. č. 7 Histogram délky léčby pacientů v letech	15
Obr. č. 8 Krabíkový graf délky léčby pacientů v letech.....	15
Obr. č. 9 Kaplan-Meierova křivka přežití pacientů v čase od data zahájení DPV	29
Obr. č. 10 Kaplan-Meierova křivka přežití pacientů bez paliativní DPV v čase od data zahájení DPV	31