

ODHADY RIZIKA SMRTI A NÁSLEDKŮ OTRAV PRO PACIENTY VZEŠLÉ Z ČESKÉ METANOLOVÉ AFÉRY V ROCE 2012

Jaromír Běláček, Sergej Zakharov, Martin Komarc, Daniela Pelclová

Anotace

Po vypuknutí aféry metanolových otrav v České republice v roce 2012 bylo na Klinice pracovního lékařství 1. LF UK a VFN v Praze shromážděno několik klinických datových souborů, které jsou následně podrobovány formálním statistickým analýzám. Cílem tohoto příspěvku je poskytnout základní informace o metodice statistického zpracování dvou datových souborů, které byly vnitřně diferencovatelné zejména typem aplikované hemodialýzy (intermitentní hemodialýza /IHD/ vs. kontinuální veno-venózní hemodialýza-hemodiafiltrace /CVVHD/), závažností metabolické acidózy (např. podle hladin pH, HCO₃ v krvi pacientů a jiných markerů), množství etanolu v krvi intoxikovaných pacientů a řadou zjevných či skrytých indikovaných (subklinických) následků akutních otrav.

Hlavním výsledkem formálních statistických analýz je simultánní vs multivariátní posouzení vlivu sledovaných markerů na poločas eliminace methanolu a kyseliny mravenčí z krve pacientů ve skupinách s odlišným typem hemodialýzy a odhady pravděpodobností rizika smrti a následků otrav prostřednictvím modelů ordinální multinomické regrese fitovaných na základě logaritmu věrohodnostního („log likelihood“) poměru LR R². Použitá metodika přímočaře zobecňuje dosud standardně užívané metodiky lineárních regresních analýz pro normálně a binomicky rozdělené závisle proměnné a poukazuje zejména na možnosti dobře interpretovatelných výsledků (včetně grafických prezentací statisticky nejvýznamnějších formálních závislostí).

Řešení úlohy bylo podpořeno z prostředků projektu PRVOUK P25/1LF/2/UK Praha.

Klíčová slova:

Intoxikace organismu metanolem, intermitentní hemodialýza /IHD/, kontinuální veno-venózní hemodialýza /CVVHD/, explorační faktorová analýza /EFA/, BW a FR regresní algoritmy, ordinální multinomická regrese

1. Úvod

Po vypuknutí aféry metanolových otrav v České republice v roce 2012 bylo na Klinice pracovního lékařství 1. LF UK a VFN v Praze vytvořeno několik datových souborů, které charakterizují zdravotní stav většiny postižených pacientů podle výsledků klinických a laboratorních vyšetření. Pro účely formálních statistických zpracování byly získané údaje o pacientech strukturovány postupně do dvou datových souborů diferencovatelných zejména – 1/ typem aplikované hemodialýzy (intermitentní hemodialýza /IHD/ vs. kontinuální veno-venózní hemodialýza/hemodiafiltrace /CVVHD/) a technickými parametry hemodialyzačních přístrojů; - 2/ závažností zdravotního postižení (v závislosti na stupni acidózy,

včasnosti zahájení medikace etanolem ev. dalšími charakteristikami) a řadou zjevných či skrytě indikovaných subklinických následků otrav.

Hlavní výsledky analýzy souboru indikátorů reprezentujících položky ad 1/ byly shrnuty již ve studii [1], kde jsou prezentovány m. j. i dva statisticky a věcně validní multivariátní lineární regresní modely prokazující závislost poločasu eliminace metanolu a kyseliny mravenčí z krve pacientů simultánně na několika technických parametrech použitých dialyzačních metod. Tyto modely vysvětlující více než 80% rozptylu (R^2) u obou závisle byly nalezeny s podporou automatického backward /BW/ a forward /FW/ výběru explanačních proměnných v normální regresi (viz. [4], str. 1632-1655). Protože ale statisticky plně validních modelů můžeme výše zmíněnou metodikou získat více (záleží na rozsahu, charakteru i struktuře seznamu vstupních proměnných), je vhodné kontrolovat formálně získané výsledky prostřednictvím nezávisle aplikované explorační faktorové analýzy /EFA/ (viz [4], str. 663-680 a dále ve stati 3.1).

Pro situaci ad 2/, kdy intoxikované pacienty máme rozdělené do skupin (finálně např. právě podle stupně závažnosti jejich postižení) a přejeme si hledat sestavy proměnných schopných predikovat tuto finální klasifikaci, s výše použitou metodikou vystačit nelze. Jednoduše již proto, že klasifikační závisle proměnné mají obvykle 3 nebo jiný omezený počet kategorií tzn. že nejsou normálně nebo jinak spojitě rozdělené. Pro účely modelování a predikcí závažnosti zdravotního postižení ad 2) jsme aplikovali metodiku ordinální multinomické regrese (viz. [4], str. 720-755), která respektuje charakter vstupních dat a de facto přímočaře zobecňuje alternativní multivariátní modely logistické regrese (viz v [4], str. 794-788). Lepšímu porozumění filozofii a smyslu v tomto textu spíše jen rámcově referovaných modelů logistické a ordinální multinomické regrese slouží Tabulka 3 a prezentace výsledků na Grafech 4 a 5 (ve stati 3.2), které přímočaře zobecňují metodiky standardně známé pro normálně rozdělené závisle proměnné.

Cílem tohoto příspěvku je shrnout základní informace o metodikách, které byly použity na oddělení BioStat (při ÚBI 1. LF UK Praha) v rámci statistického zpracování obou datových souborů. Přestože použité formální technologie vyhlížejí na první pohled značně odlišně, ve skutečnosti byly oba metodické postupy podřízeny analogickým zpracovatelským konceptům. Pokud máme k dispozici adekvátní softwarové nástroje, což by ovšem v dnešní realitě vyspělých informačních technologií neměl být zásadní problém, mohl by tento příspěvek sloužit i větší diseminaci a zvýšení zájmu o aplikace těchto ještě univerzálněji použitelných metod.

2. Materiál a metody

První datový soubor obsahoval $N = 24$ intoxikovaných pacientů, z nichž 11 podstoupilo IHD a 13 CVVHD (typ hemodialýzy byl v našich datech identifikován prostřednictvím dichotomické proměnné 'Group2'). Vedle poločasu eliminace metanolu ('EthanolHT') a kyseliny mravenčí ('FormateHT') obsahoval soubor dalších cca 25 technických, laboratorních a dalších relevantních markerů vyšetřených k okamžiku přijetí intoxikovaných osob do nemocničního

ošetření. Většinu z nich jsme atestovali jako potenciální prediktory v normální tj. gaussovské regresi na 'EthanolHT' a 'FormateHT'. Vzhledem k malému počtu pacientů v tomto souboru bylo nutné seznam vstupních ukazatelů redukovat (prostřednictvím EFA aplikované na matici Pearsonových korelačních koeficientů). EFA umožnila finálně identifikovat rovněž několik významných hlavních komponent (interpretovatelných jako skryté latentní faktory v datech).

Statisticky nejvýznamnější výsledky na úrovni všech standardně provedených regresních analýz jsou shrnuty v Tabulkách 1 a 2, které obsahují hodnoty Pearsonových korelačních koeficientů R (pokud mělo smysl tyto hodnoty uvádět), procenta vysvětleného rozptylu (R^2 vyjádřená v [%]), P-hodnoty pro testy nulových hypotéz („o nesignifikanci“ jednotlivých vysvětlujících proměnných nebo simultánních multivariátních modelů) a síly testů OP („observed power“), což jsou pravděpodobnosti zamítnutí nulových hypotéz v případě, že tyto hypotézy neplatí (vypočteno prostřednictvím SPSS modulu GLM, viz [4] na str. 815-857). Pro výběr multivariátních sestav vysvětlujících proměnných pro 'EthanolOH' a 'FormateHT' byly použity (v rámci třídy aditivních lineárních modelů) standardní BW a FW optimalizační algoritmy založené na jedнокrokové ("stepwise") optimalizaci diferencních F-statistik (účelem je dosáhnout co nejvyšších hodnot R^2 pro vybranou sestavu signifikantních proměnných).

Druhý datový soubor reprezentoval rozšíření souboru výše na $N = 100$ meta-nolem intoxikovaných pacientů, kteří byli na základě detailních medicínských vyšetření klasifikováni a následně analyzováni ve struktuře závisle proměnné 'Sequelae' (49 pacientů v kategorii '1' - bez vážnějších následků, 29 klasifikováno jako '2' - s vážnými následky a 21 ve skupině '3' - zemřelí). Po eliminaci některých věcně duplicitních markerů (alternativně monitorujících kupř. acidobazickou rovnováhu v organismu pacientů) obsahoval tento soubor i některé ukazatele nové (jako např. 'FirstsAid' - indikátor, zda byla poskytnuta "první pomoc etanolem" pracovníky záchranné služby po příjezdu k postiženému pacientovi /v souladu s oficiálním doporučením brzy po vypuknutí metanolové aféry na podzim 2012/ anebo proměnnou 'Time' - čas od nahlášení události s podezřením na otravu metanolem a dopravením pacienta do nemocnice). Většina vstupních ukazatelů byla (z důvodů zlepšení konzistence explanačních multivariátních modelů a snadnější interpretace výsledků v počátečních stádiích analýz) kvalifikovaně přetransformována nejprve na ordinální proměnné (kupř. 'Sequelae' byla spojením sousedních kódů nebo vynecháním některého z nich převedena na binární markery - viz legenda pod Tabulkou 3, spojitě markery byly optimalizovány vhodnou volbou dělicích bodů /"cut points"/ na tříhodnotové).

Za podpory EFA aplikované na matici Spearmanových koeficientů pořadové korelace a testů X^2 -nezávislosti v kontingenčních tabulkách byly na proměnné v rámci druhého datového souboru následně aplikovány FW a BW algoritmy dostupné pro optimální výběr vysvětlujících proměnných v logistické regresi (v rámci SPSS modulu LOGISTIC). Až konečně pro tříhodnotovou 'Sequelae' byl použit model ordinální multinomické regrese (SPSS modul GENLIN), kde byly finálně verifikovány hypotézy pilotované v předchozích krocích analýzy.

Pro posouzení hierarchie významnosti sledovaných markerů z hlediska predikce rizika (pravděpodobností) následků na otravu metanolem jsme se – po důkladné rešerši dostupných možností – rozhodli opřít o hodnoty věrohodnostního $LR(R^2)$, který lze v nejjednodušší transkripci reprezentovat vzorcí (viz. [6] nebo [7], str. 11-17):

$$LR(R^2) = 1 - \{\ln(L(\Theta_M)) - \ln(L(\Theta_0))\} = 1 - \{\ln(L(\Theta_M)) - N [\sum_k p_k \ln(p_k)]\},$$

kde 1. sčítanec ve složené závorce vyjadřuje přirozený logaritmus věrohodnostní funkce pro model zahrnující M vysvětlujících proměnných (“log likelihood” z tabulky “Goodness of fit” zpracovatelského programu) a 2. člen ve složené závorce reprezentuje logaritmus věrohodnostního poměru v modelu zahrnujícím pouze intercept. Výraz v hranaté závorce posledního vyjádření měří diverzitu (nebo také informační míru entropie /se záporným znaménkem/) pro distribuční rozdělení dané závisle proměnné s pravděpodobnostmi p_k v jednotlivých kategoriích ($\sum_k p_k = 1$).

Hypotézu o vlivu vysvětlujících proměnných na výslednou entropii dané závisle proměnné lze založit na statistice $1-LR(R^2)$, kterou lze poměřovat – za předpokladu simultánního vlivu sestavy M signifikantních proměnných – s asymptotickým χ^2 -rozdělením o M stupni volnosti. Rovněž signifikance příspěvků jednotlivých vysvětlujících proměnných celkovému modelu lze založit na jednokrokových („stepwise“) diferenčních LR-statistikách porovnatelných s kritickými hodnotami $\chi^2(1)$. Při výše uvedené volbě kriteriálních funkcí jsou všechny úvahy založené na procentech vysvětlené $LR(R^2)$ diverzity v přímé analogii k aplikacím normálně používaného R^2 (procenta vysvětleného rozptylu) a BW a FW optimalizačních strategií referovaných výše pro normálně rozdělené závisle proměnné.

Všechny systematické statistické analýzy byly provedeny prostřednictvím programů SPSS (Ver. 22.0). Grafy 1-3 v programu byly vytvořeny v programu Statistica (Ver. 12.0), Tabulky 1-3 a Graf 5 resp. 4 v MS Excel resp. MS Publisher.

3. Výsledky

3.1 Regresní modely pro poločasy eliminace metanolu a kyseliny mravenčí

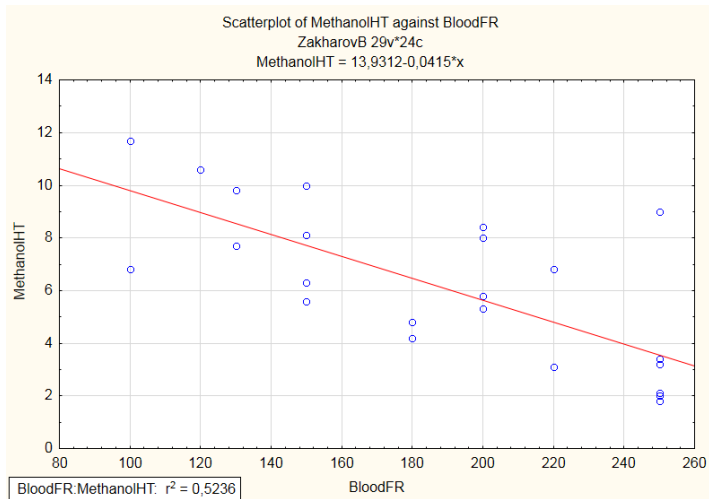
Nejvýznamnější výsledky na univariátní úrovni tj. když do analýzy vstoupila vždy pouze jedna ze seznamu vysvětlujících proměnných, jsou shrnuty v rámci Tabulky 1. Z údajů v tabulce vyplývá zřejmá závislost poločasů eliminace metanolu a kyseliny mravenčí na typu dialýzy (údaje pro ‘Group2’) a na rychlosti proudění krve pacientů (‘Blood FR’), což je demonstrováno i prostřednictvím Grafů 1 a 2. Vyšší hodnoty ‘Blood FR’, které byly u našeho vzorku pacientů lépe dosažitelné při aplikaci IHD tedy přispěly možností rychlejší rehabilitace zdravotního stavu intoxikovaných osob než tomu bylo u CVVHD ($P < 0.001$). Statisticky významně nižších poločasů eliminace u obou závisle proměnných mělo rovněž 8 žen ($P = 0.042$ resp. $P = 0.015$) na rozdíl od zbývajících 16ti mužů v rámci tohoto prvního souboru pacientů. Z Tabulky 1 je dále zřejmé, že na úrovni univariátních modelů byly další signifikantní závislosti indikovány již pouze

pro poločas rozpadu kyseliny mravenčí, a to pro proměnné: 'Sequelae' (P=0.005), 'ArterialPressure' (P=0.013), 'pH' (P=0.038), 'GCS' (P=0.013), 'Antidote2' (P=0.019) a 'pCO₂' (P=0.035) – popisky proměnných viz Legenda pro Tabulku 1. Indikované odhady síly testů (OP v posledním sloupci Tabulky 1) však s výjimkou 'Group2', 'BloodFR' a 'Sequelae' nevypovídají o dostatečné spolehlivosti výše uvedených signifikancí v Tabulce 1 uvedených P-hodnot (při daném relativně nízkém počtu pacientů N=24 v analyzovaném souboru).

Z důvodů zjištěných rozdílů mezi efekty aplikace dvou použitých typů hemodialýzy na poločasy eliminace metanolu a kyseliny mravenčí jsme ověřovali i možný systematictější vliv identifikační proměnné 'Group2' v bilaterální kombinaci (interakci) s dalšími proměnnými z Tabulky 1. Statisticky nejvýznamnější predikční model tohoto typu je vizualizován na Grafu 3 pro nezávisle proměnnou 'DialyzerSurface' (plocha dialyzátoru). V tomto případě byly všechny parametry modelu (včetně testů nenulovosti směrnic i atestace signifikance rozdílu mezi vyšším poklesem hodnot pro 'Methanol_HT' s nárůstem 'DialyzerSurface' u IHD) shledány jako statisticky významné. Kupodivu ale tato vlastnost v rámci našeho souboru již univerzálně neplatila pro univariátně nejvýznamnější proměnnou 'Blood FR'(?!).

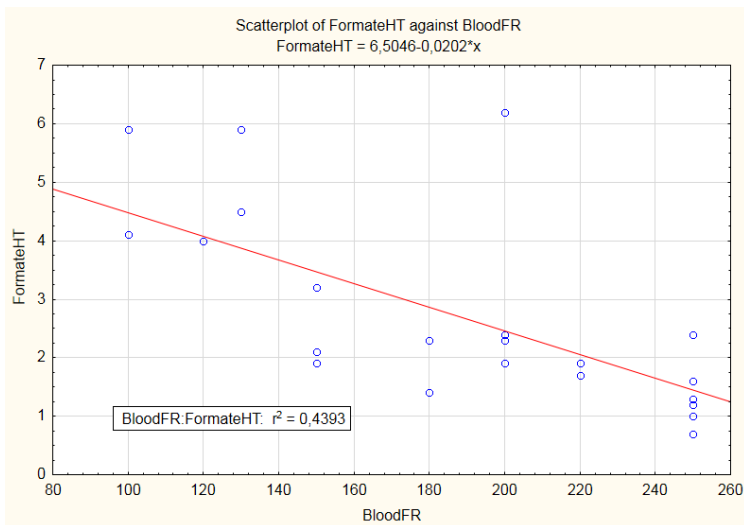
Složitost konceptu výběru optimálních sestav vysvětlujících proměnných v rámci multivariátních modelů (tak, aby byl založen na pouze statisticky významných vysvětlujících proměnných a aby přispěl co nejvýznamnějšímu navýšení procenta vysvětleného rozptylu dosaženého na úrovni univariátních analýz) demonstruje Tabulka 2. Pro obě závisle proměnné z univariátních modelů ('Methanol_HT' a 'Formate_HT') jsou prezentovány vždy 2 verze optimální sestavy vysvětlujících proměnných z Tabulky 1: jedna vzešlá z BW a druhá z FW aplikací formálního výběru proměnných. Pro 'Methanol_HT' jsou explanační/predikční modely založené dominantně na identifikátoru typu dialýzy ('Group2'), průtokovém objemu/ploše dialyzátoru ('DialyzerSurface') a proměnné 'AnionGap' (indikátor narušení acidobazické rovnováhy). Pro 'Formate_HT' vzešla prostřednictvím BW algoritmu jako vysoce signifikantní sestava založená dominantně na typu dialýzy ('Group2'), tepové frekvenci ('PulsRate') a pH; po aplikaci FW regrese pro 'Formate_HT' se optimalizovala sestava založená především na dialyzační teplotě ('DialyzeTemperature'), rychlosti proudění krve ('Blood FR') a arteriálním tlaku krve u pacientů ('ArterialPressure'). Všechny 4 modely uvedené v Tabulce 2 vysvětlily více než 80% rozptylu u obou závisle proměnných ('Methanol_HT' a 'Formate_HT').

Hlubšímu porozumění výsledkům vzešlým z formálních optimalizací multivariátních modelů výrazně napomohla formální aplikace EFA provedené na Pearsonově korelační matici pro 26 do analýzy zařazených proměnných. Analýza faktorových zátěží (po rotaci Varimax a redukci komunalit na hlavní diagonále) umožňuje interpretovat v zásadě 6 latentních faktorů (FA1-FA6, které dohromady vysvětlují 77.5% variability korelační matice); jejich faktorové zátěže (korelace faktorů s proměnnými) jsou uvedeny v 1. sloupci Tabulky 1. Čistě formální přepis identifikace FA1-FA6 do Tabulky 2 umožňuje nahlédnout, že



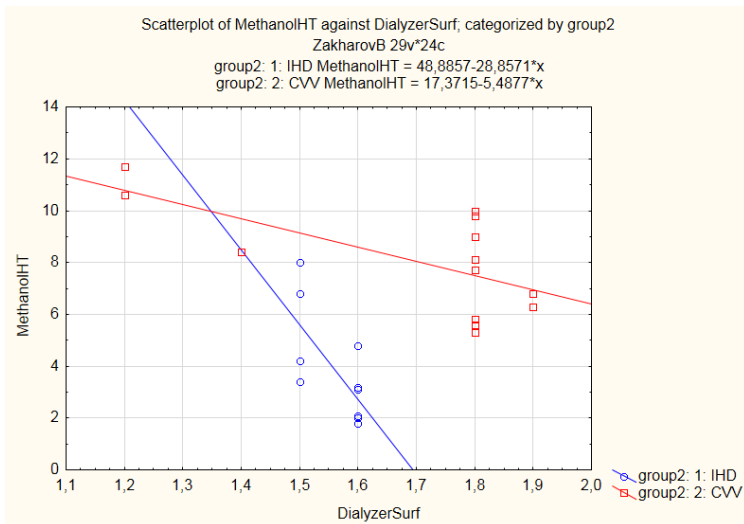
Graf 1 – Graf závislosti poločasu eliminace metanolu z krve intoxikovaných pacientů na krevním průtoku

Legenda (ke Grafu 1) – ‘Methano_HT’ – poločas eliminace metanolu z krve intoxikovaných pacientů; ‘BloodFR’ – rychlost průtoku krve; ‘r2’ – procento vysvětleného rozptylu



Graf 2 – Graf závislosti poločasu eliminace kyseliny mravenčí z krve metanolem intoxikovaných pacientů na krevním průtoku

Legenda (ke Grafu 2) – ‘Formate_HT’ – poločas eliminace kyseliny mravenčí z krve intoxikovaných pacientů; ‘BloodFR’ – rychlost průtoku krve; ‘r2’ – procento vysvětleného rozptylu



Graf 3 – Graf závislosti poločasu eliminace metanolu na ploše dialyzátoru podle typu hemodialýzy

Legenda (ke Grafu 3) – ‘Methanol_HT’ – poločas eliminace metanolu z krve intoxikovaných pacientů; ‘DialyzerSurf’ – plocha dialyzátoru; ‘r2’ – procento vysvětleného rozptylu

ve většině případů jsou v multivariátních modelech pro ‘Methanol_HT’ a ‘Formate_HT’ obsaženy sice ne zcela identické sestavy „optimálních“ vysvětlujících proměnných, ale že tyto vybrané proměnné v podstatě manifestují pouze tři (resp. čtyři v případě FW výběru pro ‘Formate_HT’) ze šesti extrahovaných hlavních faktorů. Pro výše uvedené úvahy je ovšem podstatný i fakt, že obě predikované závislé proměnné, ‘Methanol_HT’ a ‘Formate_HT’, jsou navzájem statisticky významně korelované ($R = 0.732$; $P < 0.001$) a že oba významně přispívají právě jednomu, jmenovitě nejsilnějšímu hlavnímu faktoru FA1.

		Methanol_HT				Formate_HT			
		R	R ²	P	OP	R	R ²	P	OP
FA1	Group2		54.4%	<0.001	.998		39.7%	<0.001	.953
FA1	BloodFR	-.724	52.4%	<0.001	.997	-.663	43.9%	<0.001	.977
FA2	PulsRate	.012	0.0%	0.957	.050	-.190	3.6%	0.375	.139
FA2	Age	.347	12.0%	0.097	.381	.300	9.0%	0.155	.292
FA2	AnionGap	-.041	0.2%	0.849	.054	.132	1.7%	0.540	.092
FA2	Lactate	.400	16.0%	0.053	.499	.364	13.2%	0.081	.417

	UNIVARIATE ANOVA:	Methanol_HT				Formate_HT			
		R	R ²	P	OP	R	R ²	P	OP
FA3	Sequelae	.330	10.9%	0.116	.347	.556	30.9%	0.005**	.850
FA3	ArterialPressure	-.348	12.1%	0.096	.384	-.500	25.0%	0.013*	.736
FA3	DialyzerSurface	-.115	1.3%	0.593	.081	-.197	3.9%	0.355	.148
FA3	pH	-.139	1.9%	0.518	.096	-.426	18.2%	0.038*	.560
FA3	GCS	-.399	15.9%	0.054	.496	-.498	24.8%	0.013*	.730
FA4	Fomepizol	.008	0.0%	0.970	.050	.093	0.9%	0.666	.070
FA4	HCO ₃	-.203	4.1%	0.342	.153	-.030	0.1%	0.890	.052
FA4	Ethanol	.250	6.3%	0.238	.213	.278	7.7%	0.188	.255
FA4	Antidote2		10.2%	0.129	.326		22.4%	0.019*	.674
FA5	S_MetOH	-.132	1.7%	0.540	.092	.033	0.1%	0.879	.053
FA5	OsmolalGap	.011	0.0%	0.960	.050	.250	6.3%	0.239	.213
FA5	S_Formate	.066	0.4%	0.761	.060	.211	4.5%	0.321	.163
FA6	Dialyze Temperature	.004	0.0%	0.984	.050	.135	1.8%	0.530	.094
FA6	pCO ₂	.280	7.9%	0.185	.258	.432	18.7%	0.035*	.575
FA6	Sex		17.5%	0.042*	.543		24.2%	0.015*	.718

Tabulka 1 – Korelace a závislost poločasů eliminace metanolu a kyseliny mravenčí na vysvětlujících proměnných – výsledky aplikací univariálních ANOVA modelů

Legenda (k Tabulce 1) – 'Methanol_HT' resp. 'Formate_HT' – poločas eliminace metanolu resp. kyseliny mravenčí z krve intoxikovaných pacientů; 'R' – Pearsonův korelační koeficient; 'R²' – procento vysvětleného rozptylu; 'P' /p-value/ resp. 'OP' (observed power) – chyba 1. resp. 2. druhu; 'Group2' – formální identifikátor typu hemodialýzy (IHD vs CVVHD); 'BloodFR' – rychlost průtoku krve; 'PulsRate' – tepová frekvence; 'AnionGap' – aniontové okno; 'OsmolalGap' – osmolární okno; 'Lactate' – koncentrace laktátů v krvi; 'Sequelae' – následky intoxikací (viz legenda k Tabulce 3); 'ArterialPressure' – arteriální krevní tlak, 'DialyzerSurface' – plocha dialyzátoru; 'GCS' /Glasgow Coma Scale/ - škála kvantifikující úroveň vědomí či bezvědomí pacienta vůči komatu; 'Fomepizol' resp. 'Ethanol' – binární indikátor poskytnutého antidota; 'Antidote2' - (s hodnotami: '1' - aplikace pouze jednoho antidota /'Ethanol' nebo 'Fomepizol'; '2' – současná aplikace obou antidot); 'S_MetOH' resp. 'S_Formate' – koncentrace metanolu resp. kyseliny mravenčí v okamžiku přijetí pacienta do nemocnice; 'DialyzeTemperature' – teplota dialyzátoru; 'pCO₂' – parciální tlak oxidu uhličitého v krvi; 'pH' resp. 'HCO₃' – indikátory acidobazické rovnováhy v krvi; FA1-FA6: formální příslušnost k latentnímu faktoru (podle statistické významnosti faktorových zátěží po aplikaci EFA a rotace Varimax po extrakci šesti faktorů).

Methanol_HT				Formate_HT			
BW:	R ²	P	OP	BW:	R ²	P	OP
Intercept	51.7%	<0.001	.990	Intercept	55.5%	<0.001	.996
Group2 (FA1)	73.4%	<0.001	1.000	Group2 (FA1)	73.4%	<0.001	1.000
DialyzerSurf (FA3)	47.2%	<0.001	.974	PulsRate (FA2)	46.6%	<0.001	.971
AnionGap (FA2)	39.3%	0.002**	.914	pH (FA3)	36.3%	0.004**	.877
Lactate (FA2)	20.2%	0.041*	.548	DialyzerSurf (FA3)	19.2%	0.047*	.523
CELKEM	81.4%	<0.001	1.000	CELKEM	80.6%	<0.001	1.000
Ethanol_HT				Formate_HT			
FW	R ²	P	OP	FW	R ²	P	OP
Intercept	54.2%	<0.001	.994	Intercept	35.2%	0.006**	.841
Group2 (FA1)	54.2%	<0.001	.994	Dialyz Temper (FA6)	54.3%	<0.001	.992
DialyzerSurf (FA3)	40.3%	0.002**	.925	BloodFR (FA1)	54.1%	<0.001	.991
AnionGap (FA2)	30.1%	0.010**	.775	Arterial Pressure (FA3)	41.3%	0.002**	.920
BloodFR (FA1)	20.6%	0.039*	.558	pCO2 (FA6)	23.3%	0.031*	.601
				PulsRate (FA2)	24.4%	0.027*	.627
CELKEM	81.5%	<0.001	1.000	CELKEM	81.5%	<0.001	1.000

Tabulka 2 – Závislost poločasů eliminace metanolu a kyseliny mravenčí na vysvětlujících proměnných – výsledky aplikací BW a FW algoritmů (multivariátní ANOVA modely)

Legenda (k Tabulce 2) – ‘Methanol_HT’ resp. ‘Formate_HT’ – poločas eliminace metanolu resp. kyseliny mravenčí z krve intoxikovaných pacientů; ‘R²’ – parciální resp. celkové procento vysvětleného rozptylu (v rámci multivariátního modelu); ‘P’ /p-value/ resp. ‘OP’ (observed power) – chyba 1. resp. 2. druhu; FA1–FA6: formální příslušnost k latentnímu faktoru (převzato z prvního sloupce Tabulky 1).

3.2 Predikční modely pro rizika následků otrav

Pro klasifikaci následků otrav jsme měli možnosti využít několik specificky konstruovaných ukazatelů, z nichž některé, např. APACHE II (zkr. pro Acute Physiology and Chronic Health Evaluation), mohou mít rovněž charakter normálně rozdělených veličin. Vzhledem k zájmu a potřebě vytvořit predikční model/y na základě vlastních sestav dostupných vysvětlujících markerů (pro soubor velikosti N=100 pacientů) jsme se rozhodli modelovat riziko metanolových otrav pro tříhodnotovou proměnnou ‘Sequelae’. Jelikož standardní aplikace modelů

logistické regrese jsou speciální případy modelů ordinální multinomické regrese (jak jsme v průběhu analýz ověřili i na numerické úrovni), bylo na místě vzájemně porovnat výsledky získané paralelně pro několik smysluplně interpretovatelných alternativ odvozených z proměnné 'Sequelae' (jejich definice viz v legendě pod Tabulkou 3).

Z Tabulky 3 je zřejmá především dominantní pozice následujících predičních markerů (v hierarchii od nejvyšší statistické významnosti): množství etanolu v krevním séru (proměnná 'S-EtOH') – $LR(R^2) = 29.4\%$; koncentrace pH resp. množství HNO_3 v krvi postižených osob – $LR(R^2) = 27.5\%$ resp. 18.5% ; a proměnná 'FirstAid' (indikátor monitorující, zda byla "první pomoc etanolem" poskytnuta pracovníky záchranné služby ještě před dopravením pacienta do nemocničního zařízení /v souladu s oficiálním doporučením brzy po vypuknutí metanolové aféry na podzim roku 2012/) – $LR(R^2) = 12.4\%$.

Jako statisticky významných pro klasifikaci následků otrav se projevila ještě řada dalších exogenních markerů; pro některou ze zbývajících tří binárních závisle proměnných však můžeme v Tabulce 3 vysledovat i jejich odlišnou statistickou významnost (oproti syntetické tříhodnotové 'Sequelae'). Tak například proměnná 'GCS' (posouzení stavu vědomí) nebo 'Lactate' (koncentrace laktátů v krevním řečišti) se jeví jako 3. a 4. nejméně významný predictor pro 'Sequelae' [12;3]. To znamená, že tyto dvě proměnné mají vyšší schopnost diskriminovat budoucí zemřelé od později přeživších osob, než měly již výše zmíněné HNO_3 nebo indikátor 'FirstAid'. V obdobném smyslu i na příjmu zjištěná koncentrace metanolu v krvi pacientů (proměnná 'S-MeOH') má větší význam než 'GCS', 'Lactate' nebo 'TimeMet' (doba mezi příjezdem k osobě s methanolem v krvi a příjezdem do nemocničního zařízení) z hlediska pozdější klasifikace míry postižení u přeživajících pacientů (viz 12.5% vysvětlené diversity v předposledním sloupci Tabulky 3). Úvahy tohoto typu jsou ale korektní pouze z hlediska hierarchie pořadí hodnot $LR(R^2)$ v rámci seznamu univariátních vysvětlujících proměnných, protože již samotná procenta diversity vysvětlovaná ordinálním multinomickým modelem jsou (ze své definice) závislá na vstupní diverzitě kategoriální závisle proměnné.

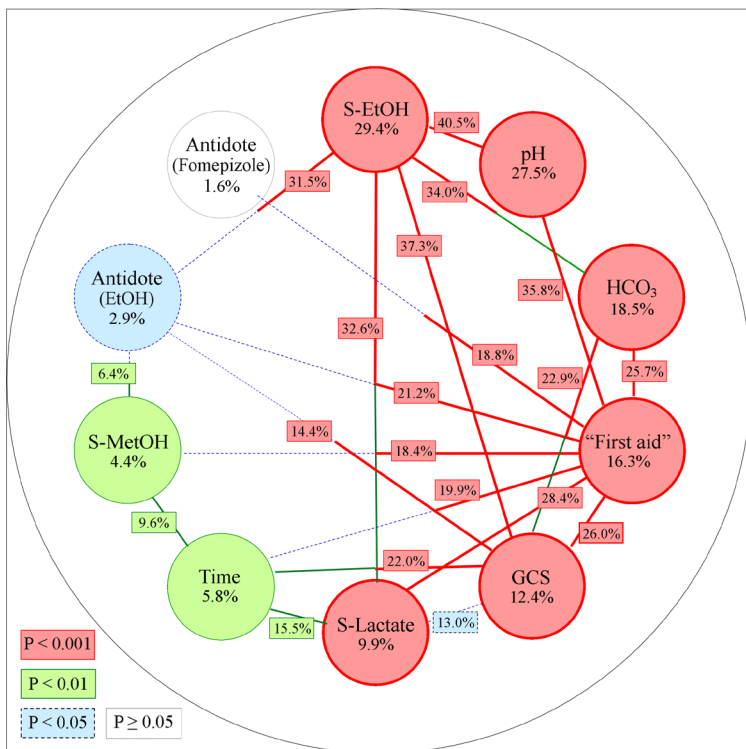
Rovněž v případě ordinální multinomické regrese jsme se pokusili o formální analýzy na bázi multivariátních sestav vysvětlujících proměnných. FW i BW automatizované algoritmy založené na statistické významnosti diferencí LR odhadů (v rámci modulu LOGISTIC programu SPSS) však zastavilo všechny automatické výběry již na úrovni bivariátních modelů (do modelů se tedy dostaly vždy nejvýše dvě vysvětlující proměnné). Všechny výsledky signifikantní pro tříhodnotovou 'Sequelae' a nezávisle proměnné z Tabulky 3 jsou vyjádřeny prostřednictvím schematu na Grafu 4 (viz opět Legenda pod Grafem 4). Statisticky nejméně významný (bivariátní ordinální trinomický) model příslušel vysvětlující proměnné 'S-EtOH' (množství etanolu v séru v okamžiku přijetí) v kombinaci s 'pH' ($LR(R^2) = 40.5\%$; 'S-EtOH' i 'pH' signifikantní pro $P < 0.001$). Pro ilustraci formálního i věcného smyslu tohoto modelu prezentujeme nejdůležitější teoretické křivky založené na odhadu parametrů jednoho z alternativních modelů stejného typu prostřednictvím Grafu 5. Na Grafu 5 je zobrazen model

pro spojitě hodnoty 'pH' a jim příslušné odhady pravděpodobností úmrtí a pravděpodobností těžkých následků otrav pro tři kategorizované úrovně hladin etanolu: '1: EtOH = 0'; '2: 0 < EtOH ≤ 500mg/l'; '3: 500mg/l < EtOH'. (Procenta vysvětlené diversity i signifikance parametrů pro všechny modely odvozené ze stejných vysvětlujících proměnných byly ve všech námi odhadovaných modelech takřka stejné jako na schématu Grafu 4, který je založený na tříhodnotových reprezentacích většiny původně spojitých vysvětlujících proměnných.)

UNIVARIATE	Sequela[1;2;3]		Sequela[1;23]		Sequela[12;3]		Sequela[1;2]	
ORDINAL	N=[49; 30; 21]		N=[49; 51]		N=[79; 21]		N=[49; 30]	
MULINOMIAL	LR(R ²)	P	LR(R ²)	P	LR(R ²)	P	LR(R ²)	P
S-EtOH	29.4%	<0.001	41.7%	<0.001	26.1%	<0.001	<0.001	<0.001
pH	27.5%	<0.001	37.5%	<0.001	23.7%	<0.001	<0.001	<0.001
HCO ₃	18.5%	<0.001	27.9%	<0.001	15.9%	<0.001	<0.001	<0.001
FirstAid	16.3%	<0.001	23.0%	<0.001	16.8%	<0.001	<0.001	<0.001
GCS	12.4%	<0.001	11.0%	<0.001	23.3%	<0.001	<0.001	0.072
Lactate	9.9%	<0.001	8.2%	0.003**	22.8%	<0.001	<0.001	0.286
Time	5.8%	0.001**	8.8%	0.001**	4.5%	0.060	0.060	0.008**
S-MetOH	4.4%	0.003**	9.4%	<0.001	0.8%	0.382	0.382	<0.001
Ethanol	2.9%	0.013*	4.4%	0.014*	2.8%	0.091	0.091	0.062
Fomepizole	1.6%	0.069	2.8%	0.047*	0.9%	0.332	0.332	0.084
Group2	1.1%	0.176	0.5%	0.448	3.6%	0.072	0.072	0.922
Folates	0.8%	0.193	0.5%	0.409	2.6%	0.101	0.101	0.965

Tabulka 3 – Závislost následků otravy metanolem (v různých variantách klasifikace) na vysvětlujících proměnných – výsledky pro univariátní modely ordinální multinomické regrese

Legenda (k Tabulce 3) – 'Sequela[1;2;3]' – následky otrav (v klasifikaci: '1' - bez vážnějších následků; '2' - s vážnými následky; '3' – zemřelí); 'Sequela[1;23]', 'Sequela[12;3]', 'Sequela[1;2]' – binomické proměnné definované z tříhodnotové 'Sequela[1;2;3]' parciálním spojením kódů: '2+3' - pacienti s následky včetně zemřelých; '1+2' – přeživší pacienti; vynecháním kódu '3' – zemřelí. Např. posledně jmenovaná binomická proměnná nabývá tedy hodnoty '1' pro 49 osob bez vážnějších následků a '2' - s vážnými následky pro 30 dosud přeživajících pacientů; LR(R²) - procento vysvětlené diversity; 'P' /p-value/ – chyba 1. druhu; 'FirstAid' - indikátor, zda byla poskytnuta "první pomoc etanolem" pracovníky záchranné služby po příjezdu k postiženému pacientovi; 'Time' - čas od nahlášení události s podezřením na otravu metanolem a dopravením pacienta do nemocnice; pro ostatní vysvětlující proměnné viz legenda k Tabulce 1.

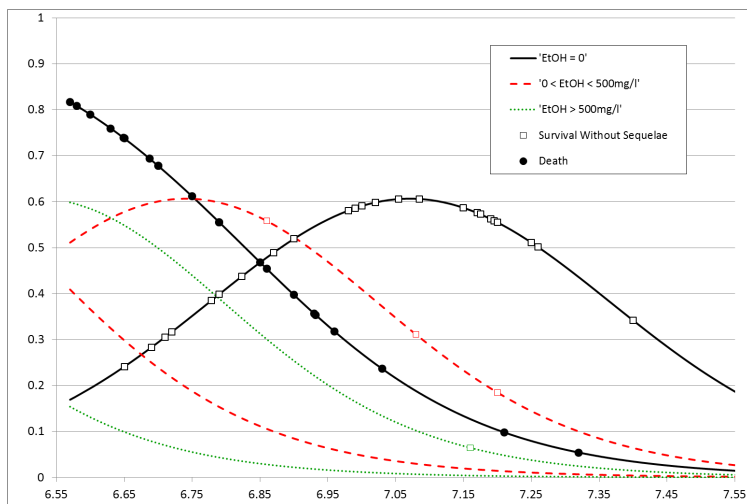


Graf 4 – Schema závislosti následků otrav (podle tříhodnotové klasifikace proměnné ‘Sequelae’ (‘1’ - bez vážnějších následků; ‘2’ - s vážnými následky; ‘3’ – zemřeli) na nejvýznamnějších vysvětlujících proměnných – statisticky významné hodnoty procenta vysvětlené diverzity/entropie LR(R²) v modelech ordinální multinomické regrese

Legenda (ke Grafu 4) – Hodnoty LR(R²) pro bivariátní modely jsou zobrazeny na spojnicích mezi jednotlivými vysvětlujícími proměnnými, které jsou uspořádány do kruhu podle hierarchie významnosti LR(R²) na univariátní úrovni t. j. podle 2. sloupce v Tabulce 3. Individuální signifikance proměnných v rámci každého bivariátního modelu je odstupňována tloušťkou resp. kružnic korespondující se statistickou významností LR odhadů parametrů modelů na hladinách významnosti 0.05, 0.01 a 0.001. Na grafu jsou prezentovány pouze statisticky významné proměnné v modelech, s výjimkou indikátoru aplikace ‘Fomepizolu’ s hodnotou LR(R²) = 1.6% (v rámci univariátního modelu), která byla shledána jako nesignifikantní.

4. Diskuse a závěry

Z výsledků prezentovaných v tomto příspěvku jednoznačně vyplývá, že jak počasy eliminace metanolu a kyseliny mravenčí, tak budoucí následky otrav lze predikovat prostřednictvím hned několika alternativních nezávisle měřených markerů. Úspěšnost predikce měřená procenty vysvětleného rozptylu



Graf 5 – Graf závislosti rizika úmrtí nebo vážných následků otravy metanolem na 'pH' a na množství etanolu v krvi pacientů v okamžiku přijetí do nemocničního ošetření

Legenda (ke Grafu 5) – Na grafu jsou zobrazeny teoretické křivky rizika/pravděpodobnosti úmrtí a pravděpodobnosti přežití s vážnými následky (na ose Y) ve vztahu k spojitě naměřeným hodnotám pH a podle množství etanolu zjištěného v krvi pacientů (kategorizované do tří skupin podle legendy zobrazené uvnitř grafu). Za pozornost stojí modelem odhadnutá hodnota pH ~ 6.85 (na ose X), která modeluje rozhraní, kde se mění nerovnováha mezi vyšší pravděpodobností úmrtí (pro pH < 6.85) vůči pravděpodobnosti holého přežití (s následky) v rámci skupiny pacientů, kteří neměli v okamžiku přijetí do nemocnice žádný etanol v krvi. Pro pH > 6.85 je tomu u pacientů bez etanolu již obráceně. Ve skupinách osob, kteří měli v krvi aspoň nepatrné množství etanolu (předpokládá se ale v zásadě větší než 100 mg/l) je modelovaná pravděpodobnost přežití pacientů vyšší než pravděpodobnost úmrtí pro jakékoli reálně zjištěné pH.

(jmenovitě u normálně rozdělených veličin) ale formálně závisí na schopnosti jednotlivých vysvětlujících proměnných numericky reprezentovat algebraickou/vektorovou strukturu dané závisle proměnné. Pro optimalizaci sestav statisticky nejvýznamnějších prediktorů (v rámci třídy aditivních lineárních regresních modelů) je relevantní informace obsažena v matici Pearsonových korelačních koeficientů. Hodnota této matice, což je vlastně počet lineárně nezávislých sloupců/proměnných v tzv. design matici každého konkrétního lineárního regresního modelu, de facto určuje i maximální počet vysvětlujících proměnných, které můžeme najít prostřednictvím formálních (v našem případě FW nebo BW) optimalizačních/vyhledávacích algoritmů.

Pokud formálně optimalizovaná sestava proměnných pro FW i BW není identická, je třeba chápat tento rozpor jako určitou „daň poplatnou zejména nižšímu rozsahu výběru N“, kterou můžeme s jistým nadhledem kompenzovat experimentální záměnou/substitucí některé formálně vybrané proměnné

za jinou. Pokud chceme při této substituci ale zachovat počet signifikantních proměnných i přibližně stejné procento vysvětleného rozptylu závisle proměnné, je nutné (ale nikoli postačující) volit novou/nahrazující proměnnou ze stejného latentního faktoru, jako byla proměnná, kterou bychom si přáli právě substituuovat.

Naše dosavadní empirické zkušenosti však naznačují, že obdobné závěry by mohly platit i v rámci obecnějších tříd tzv. zobecněných lineárních modelů, když matici Pearsonových korelací alternujeme kupř. maticí Spearmanových koeficientů pořadové korelace. V rámci výše provedených aplikací modelu ordinálními multinomické regrese (za účelem modelování pravděpodobností rizika úmrtí nebo vážných následků metanolových otrav) se nám nepodařilo nalézt jediný multivariátní predikční model, který by obsahoval více než dvě statisticky významné vysvětlující proměnné. Právě tak jako rozklad matice Spearmanových pořadových koeficientů nevedl k identifikaci více než dvou zjevně významných hlavních komponent. Univariátně nejvýznamnější vysvětlující proměnné se tedy seskupily v zásadě do dvou „latentních faktorů - FA1: 'S-EtOH', 'FirstAid', 'Time'; a FA2: 'pH', 'HCO3', 'GCS', 'S-MetOH'. Tato klasifikace v podstatě vysvětluje i strukturu bilaterálních vazeb v rámci schématu na Grafu 4, neboť proměnné významně korelující v rámci téhož faktoru – byť ve „spearmanovském“ smyslu - asi ani nemohou mít schopnost navyšovat statistiky založené na bázi $LR(R^2)$ – s asymptotickým χ^2 rozdělením – nezávislými signifikantními přírůstky.

Parametry odhadované v rámci modelů ordinální multinomické regrese umožňují zakreslit do grafů formálně rovnocenných Grafu 5 spojitě křivky modelující jednak nárůst pravděpodobností „přežití bez vážných následků“ (s rostoucími hodnotami hladin 'pH'), a také pravděpodobností „rizika smrti nebo přežití s vážnými následky“ nebo pravděpodobností „přežití /bez následků nebo s následky/“, které můžeme odhadnout zcela separátně v modelech logistické regrese pro závisle proměnné 'Sequelae [1;23]' nebo 'Sequelae [12;3]' (z Tabulky 3). Na rozdíl od separátních logistických modelů však aplikace modelu ordinální multinomické regrese umožňuje i testování hypotéz o parametrech těchto modelů (např. na bázi kontrastů) a dokonce modelovat hypotetické křivky i v situaci úplné absence dat v rámci některých skupin (jmenovitě na Grafu 5 se to týká odhadů teoretických křivek pravděpodobností „rizika smrti“ ve skupinách pacientů s nenulovým množstvím etanolu v krvi). V zásadě jde ale opět o analogickou situaci jako v případech dvou lineárních přímků zobrazených na Grafu 3, jejichž parametry lze standardně odhadnout buď separátně tj. pro každý typ dialýzy samostatně, anebo v rámci jediného modelu ANCOVA s dvouhodnotovým faktorem 'Group2', kovariátou 'DialyzerSurf' a jejich vzájemným interakčním faktorem. V tomto modelu lze pak snadno testovat např. hypotézy o rovnoběžnosti obou regresních přímků, o hodnotách jejich absolutních členů apod.

5. Poděkování

Kvalifikované zvládnutí výše představených aplikací ordinální multinomické regrese bylo umožněno soustředěnou pozorností spoluautorů tohoto příspěvku metodice analýzy dat získaných z vyšetření pacientů metanolové aféry v ČR a materiálně-technickou podporou zpracovatelskému oddělení BioStatu z prostředků projektu PRVOUK P25/1LF/2/UK Praha.

Literatura:

- [1.] Zakharov S, Pelcova D, Navratil T, Belacek J, Kurcova I, Komzak O, Salek T, Latta J, Turek R, Bocek R, Kucera C, Hubacek JA, Fenclova Z, Petrik V, Cermak M, Hovda KE: Intermittent hemodialysis is superior to continuous veno-venous hemodialysis/hemodiafiltration to eliminate methanol and formate during treatment for methanol poisoning. *Kidney International* 86/1: 199-207, JUL 2014
- [2.] Zakharov S, Pelcova D, Navratil T, Belacek J, Kurcova I, Komzak O, Hovda KE: Methanol and formate elimination half-life during treatment for methanol poisoning: Intermittent hemodialysis versus continuous hemodialysis /hemodiafiltration. *Clinical Toxicology* 52/4, 399-400, Meeting Abstract: 243, Published: APR 2014
- [3.] Zakharov S, Navratil T, Belacek J, Hovda KE, Pelcova D: Enhanced elimination methods in treatment of acute methanol poisonings: Continuous hemodialysis/hemodiafiltration versus intermittent hemodialysis. *Clinical Toxicology* 52/4, 403-403, Meeting Abstract: 250, Published: APR 2014
- [4.] SPSS Command Syntax Reference (2008). IBM SPSS Statistics for Windows, Ver. 17.0. Chicago
- [5.] StatSoft, Inc. (2013). STATISTICA (data analysing software system), Ver. 12.0
- [6.] Hosmer DW, Lemeshow S (1989). *Applied logistic regression*. New York: Wiley. 1989
- [7.] www.wikipedia.org/wiki/Logistic_regression

Kontakt:

RNDr. Běláček Jaromír, CSc.

Mgr. Komarc Martin

Ústav biofyziky a informatiky

1. LF UK + VFN Praha

Salmovská 1, 121 08 Praha 2

tel: +420 224 965 706

e-mail: jaromir.belacek@lf1.cuni.cz;

<http://biof.lf1.cuni.cz/biostat.html>

MUDr. Zakharov Sergej, Ph.D.

Prof. MUDr. Pelclová Daniela, CSc.

Klinika pracovního lékařství 1. LF UK

+ TIS VFN Praha