



POUŽITÍ LOGISTICKÉ REGRESE PRO DIAGNOSTIKU VÝSKYTU RAKOVINY PROSTATY

Kamila Fačevicová

kamila.facevicova@upol.cz

Katedra matematické analýzy a aplikací matematiky,
Přírodovědecká fakulta Univerzity Palackého v Olomouci



SHRNUTÍ

V příspěvku je nalezen vhodný model pro odhad pravděpodobnosti, že bude při rebiopsii odhalen karcinom prostaty. Pro tuto predikci se jako významné faktory ukázaly hladina celkového a volného PSA v krvi pacienta, objem jeho prostaty a také výsledek předchozí biopsie, přičemž platí, že rizikovými kombinacemi jsou malý objem prostaty a nízká hladina fPSA nebo naopak velký objem a vysoká hladina fPSA v krvi, a to bez ohledu na předchozí diagnózu.

ÚVOD

Tento příspěvek se věnuje problematice včasného odhalení karcinomu prostaty prostřednictvím správného vyhodnocení hodnot ukazatelů, měřených při standardních urologických prohlídkách. Zatímco publikace [2] se zabývala situací při prvním odběru vzorku tkáně (biopsii), nyní se zaměřujeme na odběr následný, tedy první rebiopsii. Pomocí modelu logistické regrese chceme nalézt faktory, které mají významný vliv na výsledek rebiopsie, tento vliv kvantifikovat a porovnat s vlivem těchto ukazatelů na výsledek biopsie první. Dále nás také zajímá, zda popř. jak souvisí výsledek rebiopsie s diagnózou stanovenou při první biopsii.

K dispozici byly záznamy z vyšetření celkem od 126 pacientů, z nichž 18ti bylo při následné rebiopsii odhaleno zhoubné onemocnění prostaty. Tato vyšetření byla provedena v období od června 2006 do konce roku 2010 na Urologické klinice Fakultní nemocnice Olomouc a zahrnovala jen pacienty ve věku od 45 do 80 let s hladinou prostatického specifického antigenu (PSA) v krvi menší než 20 ng/ml a objemem prostaty do 150 ml. Pro hledání vhodného modelu byly k dispozici tyto vysvětlující proměnné: věk pacienta, hodnota PSA, volné frakce PSA (fPSA) a PSA indexu (poměr volného a celkového PSA), celkový objem prostaty (všechny měřeny při první rebiopsii) a výsledek předchozí biopsie, kdy jsme rozlišovali jen zda byla pacientovi diagnostikována hyperplazie či nikoliv, četnosti zbylých diagnóz totiž nebyly dostatečné. Případy, kdy byla již při první biopsii odhalena rakovina, nebyly do souboru zařazeny. Vysvětlovanou proměnnou pak byl výsledek rebiopsie, ten byl považován za pozitivní (značíme 1) v případech, kdy odhalil karcinom prostaty a za negativní (značíme 0) ve zbylých případech.

biopsie \ rebiopsie	negativní	pozitivní	celkem
hyperplazie	63	13	76
zbylé	45	5	50
celkem	108	18	126

Table 1: Rozdělení pozitivních a negativních výsledků rebiopsie v závislosti na výsledku první biopsie.

MODEL LOGISTICKÉ REGRESE

Obecný zápis modelu logistické regrese je

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k,$$

kde π je pravděpodobnost úspěchu, tedy že vysvětlovaná proměnná bude rovna jedné a regresory x_1, x_2, \dots, x_k pak zastupují jednotlivé vysvětlující proměnné. Hledané parametry $\beta_1, \beta_2, \dots, \beta_k$ lze interpretovat jako logaritmus poměru šancí na úspěch v případě, kdy příslušný regresor zvýšíme o jedničku a zbylé zůstanou neměnné. Absolutní člen β_0 je pak logaritmus šance, že nastane sledovaná událost v situaci, kdy jsou všechny regresory nulové.

Vztah mezi výsledkem rebiopsie a výše uvedenými regresory nejlépe popisuje model

$$\ln \frac{\pi}{1-\pi} = -2,14 + 0,44 \cdot \text{PSA} - 2,70 \cdot \text{fPSA} + 3,34 \cdot \text{hp} - 0,09 \cdot \text{ob} - 2,03 \cdot \text{fPSA} \cdot \text{hp} + 0,05 \cdot \text{fPSA} \cdot \text{ob}.$$

V tomto modelu π značí pravděpodobnost, že bude výsledek rebiopsie pozitivní a bude tedy odhalen karcinom prostaty. Regresor 'hp' je roven 1, pokud byla pacientovi při první biopsii zjištěna hyperplazie prostaty, ve zbylých případech je nulový. Za všechny zbylé regresory dosazujeme hodnoty naměřené pacientovi při vyšetření, jež předcházelo rebiopsii.

Bodové i intervalové odhady jednotlivých parametrů jsou uvedeny v následující tabulce:

Parametr	Regresor	Bodový odhad	Intervalový odhad
β_0		-2,1435	$\langle -6,1931; 1,9061 \rangle$
β_1	PSA	0,4390	$\langle 0,2252; 0,6529 \rangle$
β_2	fPSA	-2,7036	$\langle -5,5837; 0,1764 \rangle$
β_3	hyperplazie	3,3443	$\langle 0,5539; 6,1348 \rangle$
β_4	objem prostaty	-0,0884	$\langle -0,1655; -0,0113 \rangle$
β_5	interakce fPSA a hyperplazie	-2,0247	$\langle -4,0313; -0,0181 \rangle$
β_6	interakce fPSA a objemu	0,0467	$\langle 0,0081; 0,0853 \rangle$

Table 2: Bodové a intervalové odhady parametrů.

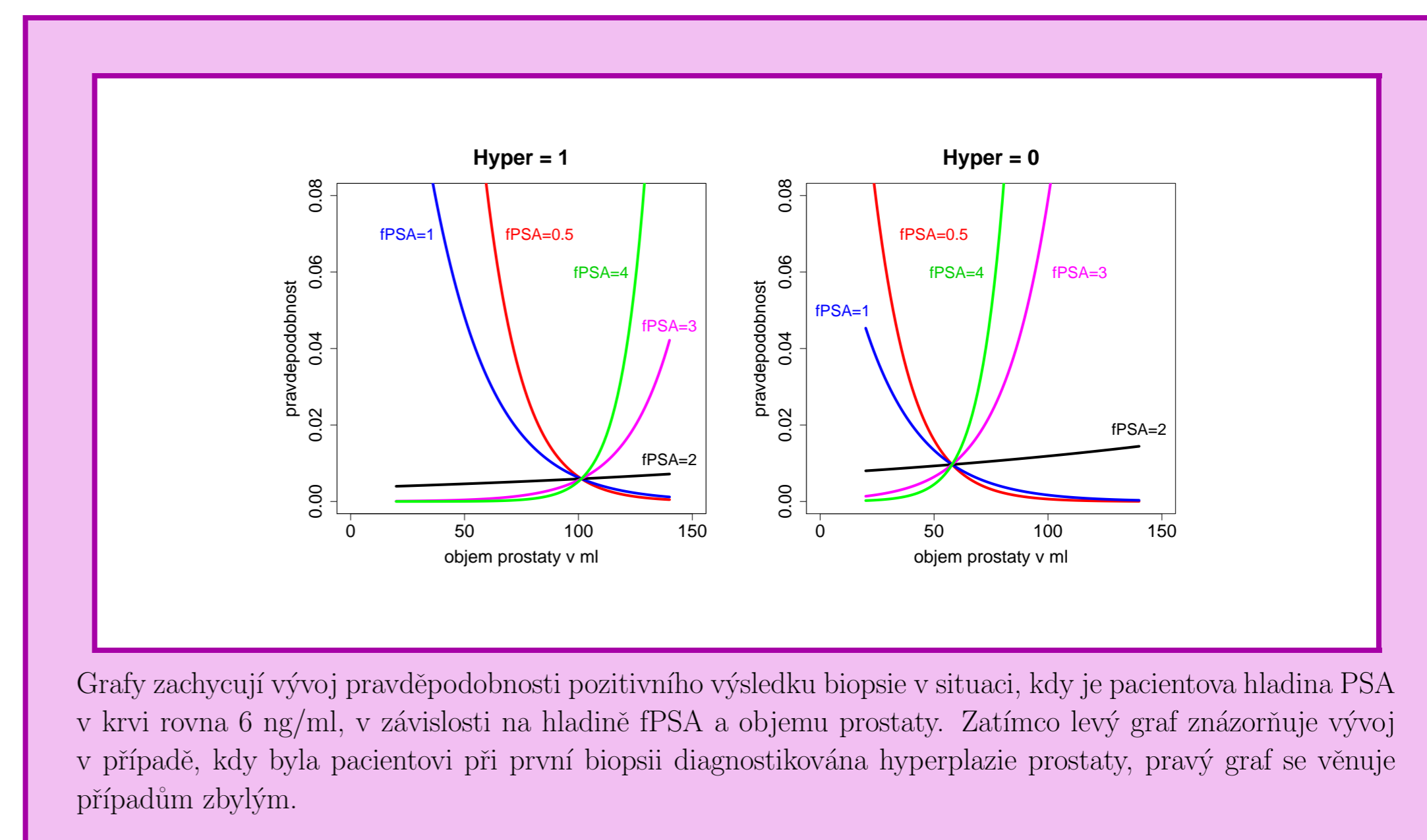
INTERPRETACE VÝSLEDKŮ

Pro správnou interpretaci modelu je potřeba rozlišit, zda pacient při první biopsii trpěl hyperplazií, nebo ne. Zabýváme se nejprve situací, kdy pacientovi při první biopsii byla diagnostikována hyperplazie, v tomto případě platí:

- Pokud je při rebiopsii naměřeno fPSA menší než asi 1,9 ng/ml, pak s rostoucím objemem klesá pravděpodobnost diagnostiky rakoviny, pokud je ale fPSA naopak větší než 1,9 ng/ml, roste s objemem i daná pravděpodobnost a vliv objemu je tak opačný než v předchozím případě.
- Obdobně platí, že při objemu menším než asi 100 ml, znamená větší fPSA menší pravděpodobnost pozitivní biopsie a při překročení hranice 100 ml je tomu opět naopak a s rostoucím fPSA roste i pravděpodobnost.

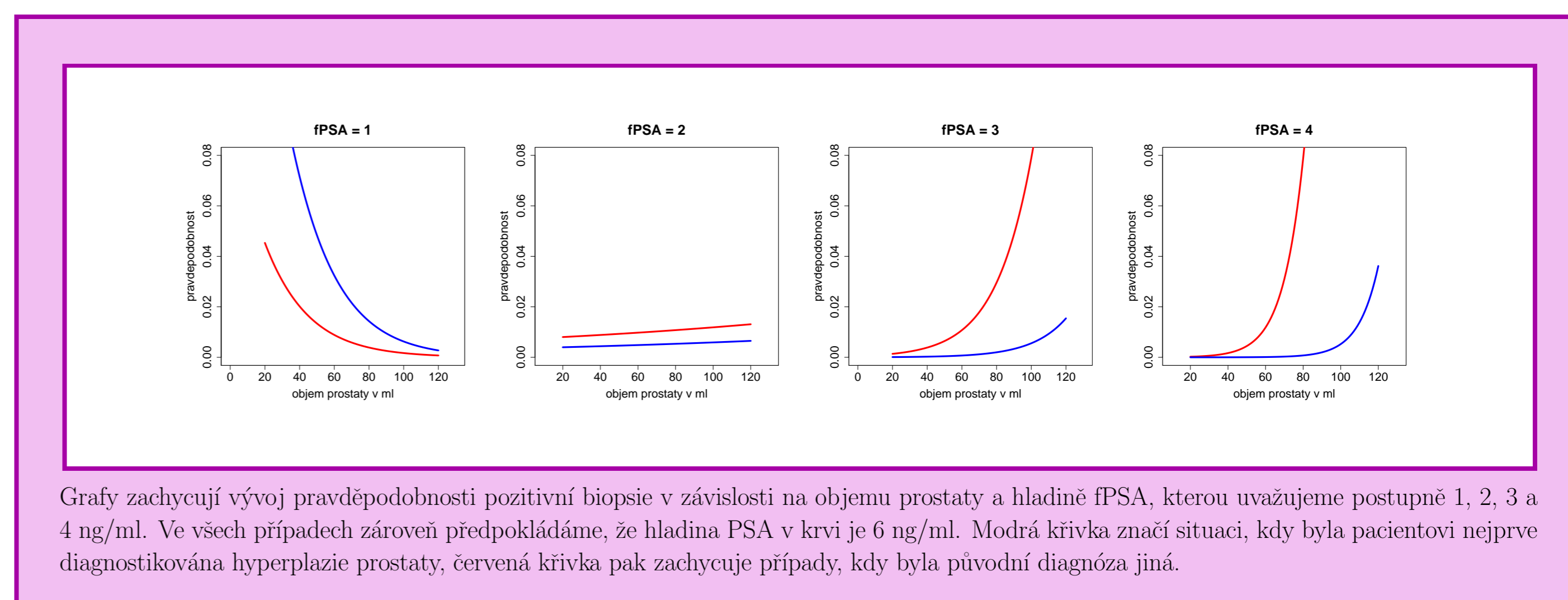
Druhou situací je, že výsledek první biopsie byl jiný než hyperplazie. I v tomto případě jsou ale závěry 1. a 2. platné s tím rozdílem, že hraniční hodnota objemu je nyní jen 57,5 ml.

Výše uvedené závěry lze shrnout tak, že nezávisle na výsledku první biopsie je vždy riziková kombinace vysokého fPSA a vysokého objemu prostaty nebo nízkého fPSA a malého objemu prostaty. Pokud je ale jeden z těchto faktorů vysoký a druhý nízký, je predikce výsledku rebiopsie příznivější. Tuto vlastnost lze snadno pozorovat i v následujících grafech, kde je zachycen vývoj pravděpodobnosti pozitivní biopsie, v závislosti na hladině fPSA a objemu prostaty. Hodnoty jsou počítány pro modelovou situaci, kdy je PSA rovno 6 ng/ml.



Grafy zachycující vývoj pravděpodobnosti pozitivního výsledku biopsie v situaci, kdy je pacientova hladina PSA v krvi rovna 6 ng/ml, v závislosti na hladině fPSA a objemu prostaty. Zatímco levý graf znázorňuje vývoj v případě, kdy byla pacientovi při první biopsii diagnostikována hyperplazie prostaty, pravý graf se věnuje případům zbylým.

Analýza dále ukázala rozdílnost pravděpodobností výskytu karcinomu mezi skupinou pacientů, jimž byla při první biopsii diagnostikována hyperplazie a skupinou pacientů s diagnózou odlišnou. Při fPSA menším než 1,65 ng/ml byla tato pravděpodobnost vyšší pro první skupinu, je-li ale fPSA vyšší, je tomu právě naopak. I tuto skutečnost lze pozorovat v grafu.

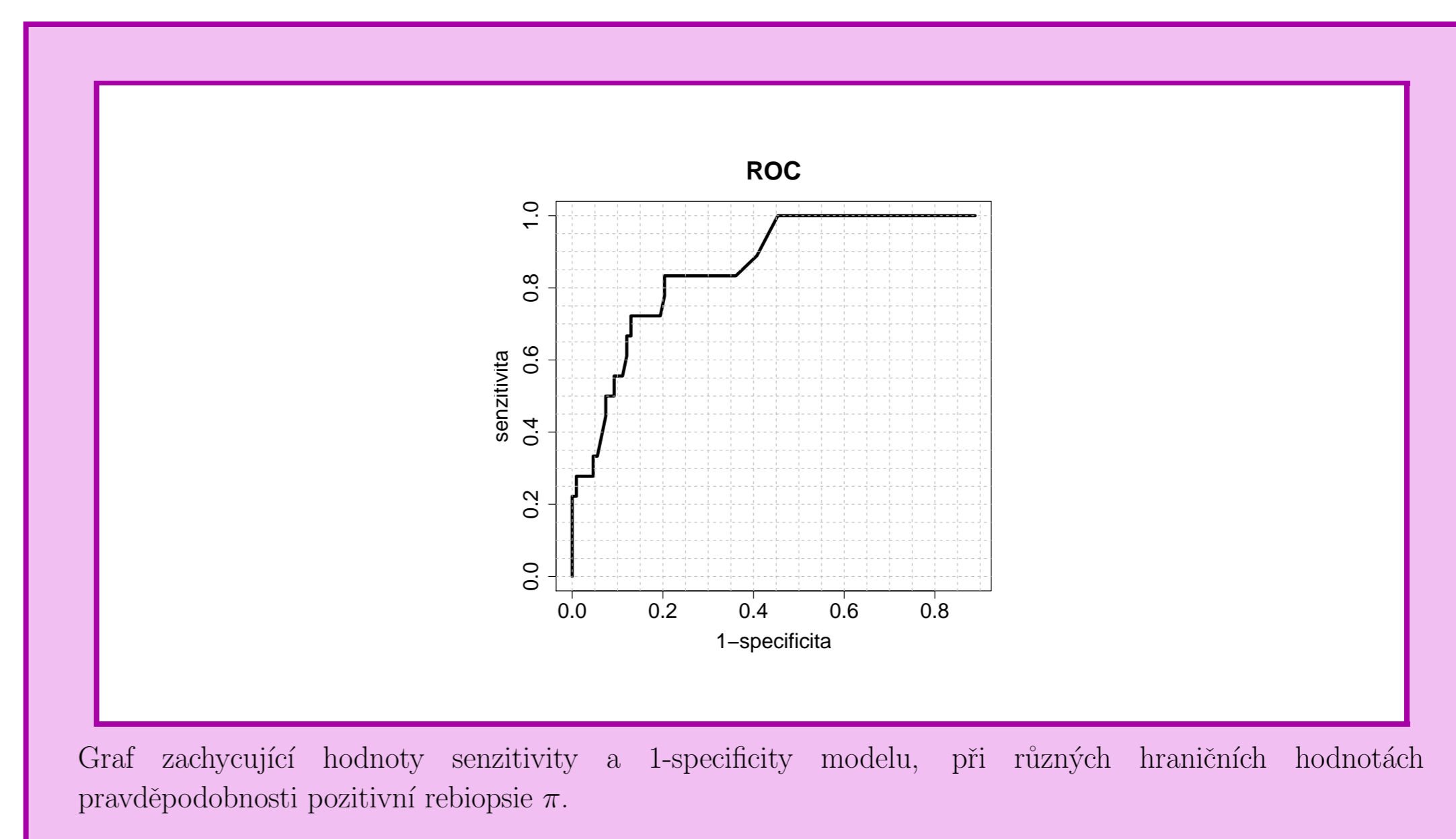


Grafy zachycující vývoj pravděpodobnosti pozitivní biopsie v závislosti na objemu prostaty a hladině fPSA, kterou uvažujeme postupně 1, 2, 3 a 4 ng/ml. Ve všech případech zároveň předpokládáme, že hladina PSA v krvi je 6 ng/ml. Modrá křivka značí situaci, kdy byla pacientovi nejprve diagnostikována hyperplazie prostaty, červená křivka pak zachycuje případy, kdy byla původní diagnóza jiná.

Vliv věku pacienta a PSA indexu na výsledek rebiopsie, se ukázal jako nevýznamný.

SENZITIVITA A SPECIFICITA

Dosažením vstupních dat do modelu a stanovením pevné hodnoty π , při jejímž překročení předpokládáme, že bude výsledek biopsie pozitivní, lze určit senzitivitu a specifitu tohoto modelu. Senzitivita je pravděpodobnost, že bude model predikovat pozitivní výsledek rebiopsie v případě, kdy se karcinom na prostatě skutečně nachází. Její odhad získáme ze vstupních dat jako podíl počtu pacientů s pozitivním výsledkem rebiopsie a odhadem pravděpodobnosti větším než stanovené π a počtu všech pacientů, u nichž byl při rebiopsii odhalen karcinom. Oproti tomu specifita značí pravděpodobnost, že bude model u zdravých pacientů predikovat negativní výsledek rebiopsie. Tentokrát získáme její odhad jako podíl počtu všech pacientů s negativní rebiopsií a odhadem pravděpodobnosti menším než π a počtu všech pacientů s negativním výsledkem rebiopsie. Vztah mezi senzitivitou a specifitou, při různých hraničních hodnotách π , lze znázornit pomocí ROC křivky.



Graf zachycující hodnoty senzitivity a 1-specifity modelu, při různých hraničních hodnotách pravděpodobnosti pozitivní rebiopsie π .

SROVNÁNÍ S MODELEM PRO 1. BIOPSIÍ

Pro první biopsii byl v [2] nalezen odlišný model:

$$\ln \frac{\pi}{1-\pi} = -3,917 + 0,053 \cdot \text{věk} - 0,027 \cdot \text{objem} + 0,144 \cdot \text{PSA} - 0,305 \cdot \text{fPSA}.$$

Při odhadu pravděpodobnosti byl tedy nově významný vliv věku pacienta. Ze srovnání intervalových odhadů parametrů v tomto modelu a v modelu pro rebiopsii navíc vyplývá, že vliv hladiny PSA v krvi je v případě rebiopsie významně vyšší. Zatímco při první biopsii je intervalový odhad příslušného parametru $\langle 0,0888; 0,1992 \rangle$, v modelu pro rebiopsii je to $\langle 0,2252; 0,6529 \rangle$. Rozdíly mezi odhady zbylých porovnatelných parametrů nebyly statisticky významné.

Poděkování

Na tomto místě bych ráda poděkovala dr. Tomáši Füstovi a dr. Ondřeji Vencálkovi, za jejich pomoc při tvorbě tohoto příspěvku.

Literatura

- [1] AGRESTI, ALAN, *Categorical Data Analysis*, second edition, John Wiley & Sons, Inc., 2002.
- [2] FAČEVICOVÁ, KAMILA, *Použití logistické regrese pro diagnostiku výskytu rakoviny prostaty*, diplomová práce, Univerzita Palackého v Olomouci, 2012.
- [3] GREPL, M., ŠTUDENT, V., FÜRST, T., FÜRSTOVÁ, J., *Prostate cancer detection yield in repeated biopsy is independent of the diagnosis of earlier biopsies*, Biomedical papers 4/2009, 297-305.