

VILNIAUS UNIVERSITETAS MATEMATIKOS IR INFORMATIKOS
FAKULTETAS
DUOMENŲ MOKSLAS

KONKURUOJANČIŲ ĮVYKIŲ MODELIS

Tiriamasis darbas

Atliko: Simona Gelžinytė,
Ugnė Kniukšaitė,
Laineda Morkytė,
Austėja Valeikaitė DM 4k. 2gr.

Vilnius

2023

TURINYS

1. ĮVADAS	3
1.1 TYRIMO TIKSLAS	3
1.2 TYRIMO UŽDAVINIAI.....	3
1.3 DUOMENYS IR PROGRAMINĖ ĮRANGA.....	3
2. KONKURUOJANČIŲ ĮVYKIŲ MODELIS	4
2.1 PRADINĖ DUOMENŲ ANALIZĖ	4
2.2 KAPLAN - MEIER KREIVĖS	6
2.3 CIF KREIVĖS.....	10
2.4 FINE IR GRAY MODELIS	13
2.7 NOMOGRAMA IŠ KOKSO MODELIŲ.....	26
3. IŠVADOS.....	28

1. ĮVADAS

1.1 Tyrimo tikslas

Pritaikyti konkuruojančių įvykių modelį pasirinktiems duomenims.

1.2 Tyrimo uždaviniai

- Atlikti pirminę duomenų analizę;
- Nubrėžti KM ir CIF kreives;
- Sukonstruoti Fine ir Gray modelį, patikrinti jo prielaidas;
- Įvertinti modelio gerumą;
- Pateikti gauto modelio interpretacijas;
- Pateikti nomogramą;
- Apibendrinti gautus rezultatus, pateikti išvadas.

1.3 Duomenys ir programinė įranga

Pasirinktas duomenų rinkinys apie krūties vėžį¹. Pateikti įvairūs požymiai apie pacientes. Priklausomas kintamasis – dienų skaičius iki pacientės mirties ir 10 kovariančių:

- Statusas – cenzūravimo statusas: 0 – cenzūruota, 1 – mirtis nuo vėžio, 2 – mirtis ne nuo vėžio;
- Amžius – 5 amžiaus grupės ((29, 40], (40, 50], (50, 60], (60, 70], (70, 80], (80, 90]);
- I proteino lygis;
- II proteino lygis;
- III proteino lygis;
- IV proteino lygis;
- Stadija – 3 stadijos grupės (I, II, III);
- Histologija (vėžio tipas) – 3 vėžio tipai (infiltruojanti duktalinė karcinoma, infiltruojanti lobulinė karcinoma, mucinozinė karcinoma);
- HER2 statusas (teigiamas, neigiamas);
- Operacijos tipas – 4 operacijų tipai (kita, lumpektomija, modifikuota radikali mastektomija, paprasta mastektomija).

¹ <https://www.kaggle.com/datasets/kreeshrajani/breast-cancer-survival-dataset>

Iš viso yra 330 stebėjimų, pašalinus praleistas reikšmes stebėjimų liko 317. Tyrimo metu naudota „R“ programinė įranga.

2. KONKURUOJANČIŲ ĮVYKIŲ MODELIS

Parinkus konkuruojančių įvykių modelį pereiname visus modelio parinkimo etapus:

- Pradinė duomenų analizė – susipažįstame su kintamaisiais;
- Prielaidų tikrinimas:
 - Fine ir Gray modeliui – proporcingosios rizikos prielaida ir multikolinearumas;
 - Kokso modeliui – išskirtys, proporcingosios rizikos prielaida, multikolinearumas ir tiesiškumas;
- Reikšmingų kovariančių atranka;
- Parametrų ir kovariančių koeficientų įvertinimas, interpretacija;
- Modelio gerumo įvertinimas.

Duomenys buvo padalinti į mokymo ir testavimo aibes 80 : 20 santykiu.

2.1 Pradinė duomenų analizė

Prieš pradėdant taikyti modelį, susipažinome su duomenimis – kiekybiniais kintamiesiems nusibraižėme stačiakampes diagramas, o kategoriniams – pasižiūrėjome dažnių lenteles priklausomai nuo pacientės statuso, t. y. cenzūruota, mirtis nuo vėžio, mirtis ne nuo vėžio.

1 lentelė. Cenzūruotų stebėjimų ir konkuruojančių įvykių dažnių lentelė

Cenzūruota	Mirtis nuo vėžio	Mirtis ne nuo vėžio
191	62	64

Iš pateiktos lentelės (1 lentelė) matome, jog iš viso mirė 126 pacientės, o likusioms – įvykis dar neįvyko.

2 lentelė. Kategorinių kintamųjų dažnių lentelė (I)

Statusas	Amžius					
	(29, 40]	(40, 50]	(50, 60]	(60, 70]	(70, 80]	(80, 90]
0	8	45	58	39	28	12
1	3	16	16	15	8	3
2	7	13	16	15	10	3

Pagal amžiaus grupę matome, jog pacientės, mirusios nuo krūties vėžio ir nuo kitos priežasties, dažniausiai buvo 40 – 70 metų. Pacientės, kurioms įvykis dar neįvyko buvo (50, 60] amžiaus grupėje. Iš vis o tik 6 pacientės mirė sulaukusios 80 – 90 metų.

3 lentelė. Kategorinių kintamųjų dažnių lentelė (II)

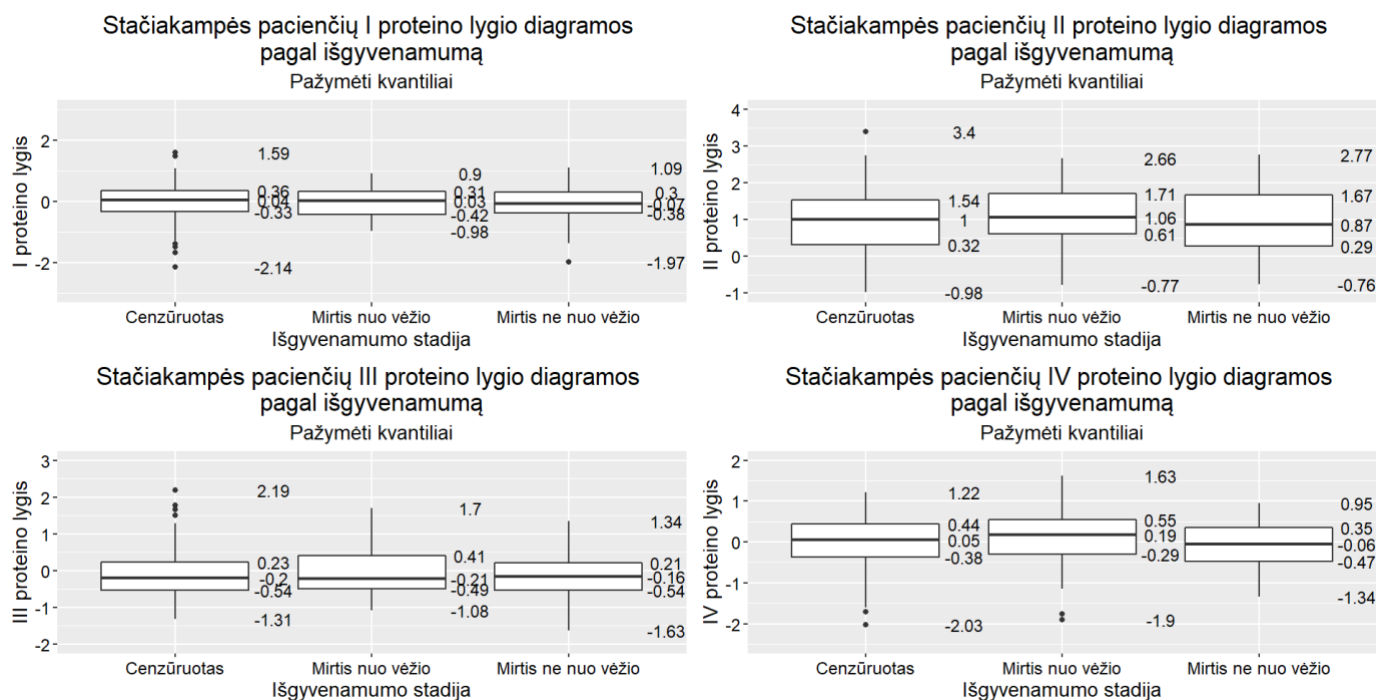
Statusas	Stadija			Histologija (vėžio tipas)		
	I	II	III	Infiltruojanti duktalinė karcinoma	Infiltruojanti lobulinė karcinoma	Mucinozinė karcinoma
0	36	112	43	129	54	8
1	9	36	17	45	14	3
2	15	32	17	50	13	1

Daugiausiai pacienčių (112), kurioms bet koks įvykis dar neįvyko sirgo II stadijos vėžiu, 36 mirė nuo krūties vėžio kai joms buvo diagnozuotas II stadijos vėžys ir 17 – kai buvo diagnozuotas III stadijos vėžys. Nuo kitos priežasties mirusių pacienčių daugiausiai buvo kai joms buvo diagnozuotas II stadijos vėžys. Dažniausias vėžio tipas buvo infiltruojanti duktalinė karcinoma, o rečiausias – mucinozinė karcinoma. Nuo krūties vėžio mirusios pacientės dažniausiai turėjo infiltruojančios duktalinės karcinomos vėžį, 14 iš jų mirė nuo infiltruojančios lobulinės karcinomos ir 3 – nuo mucinozinės karcinomos. Daugiausiai pacienčių, mirusių nuo kitos priežasties turėjo infiltruojančią duktalinę karcinomą ir tik viena mucinozinę karcinomą.

4 lentelė. Kategorinių kintamųjų dažnių lentelė (III)

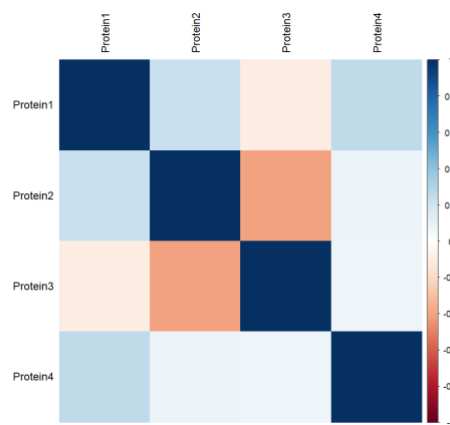
Statusas	HER2 statusas		Operacijos tipas			
	Neigiamas	Teigiamas	Kita	Lumpektomija	Modifikuota radikali mastektomija	Paprasta mastektomija
0	173	18	61	42	50	38
1	58	4	24	9	17	12
2	57	7	12	15	22	15

Dažniausias HER2 statusas buvo neigiamas. 58 nuo vėžio mirusios pacientės turėjo neigiamą HER2 statusą, o tik 4 – teigiamą. Panašūs rezultatai ir kai pacientės mirė nuo kitos priežasties. Pagal operacijos tipą dar neįvykus įvykiui dažniausiai buvo atlikta kito tipo operacija. Pacientėms, mirusioms nuo vėžio dažniausiai buvo atlikta modifikuota radikali mastektomija arba kito tipo operacija. Nuo kitos priežasties mirusioms pacientėms dažniausiai buvo atlikta modifikuota radikali mastektomija.



1 pav. Kiekybinių kintamųjų stačiakampės diagramos

Pagal stačiakampes diagramas matome, jog I proteino lygis visais atvejais buvo panašus. II ir III proteino lygis buvo šiek tiek aukštesnis mirusioms pacientėms. IV proteino lygis taip pat visais atvejais panašus.

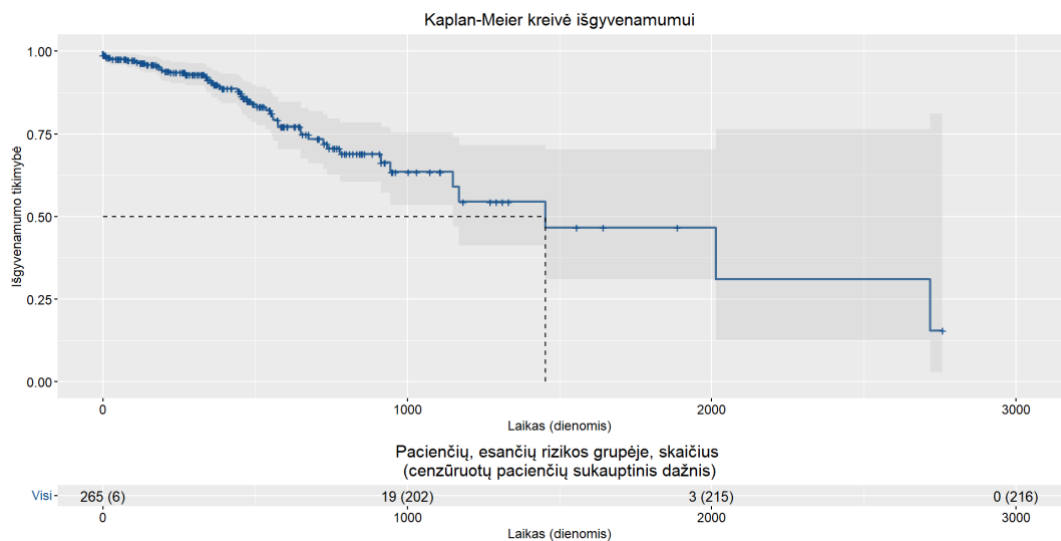


2 pav. Koreliacijų matrica

Iš koreliacijų matricos matome, jog tarp visų kiekybinių kintamųjų koreliacija yra nestipri. Šiek tiek stipriau neigiamai koreliuoja II ir III proteino lygiai.

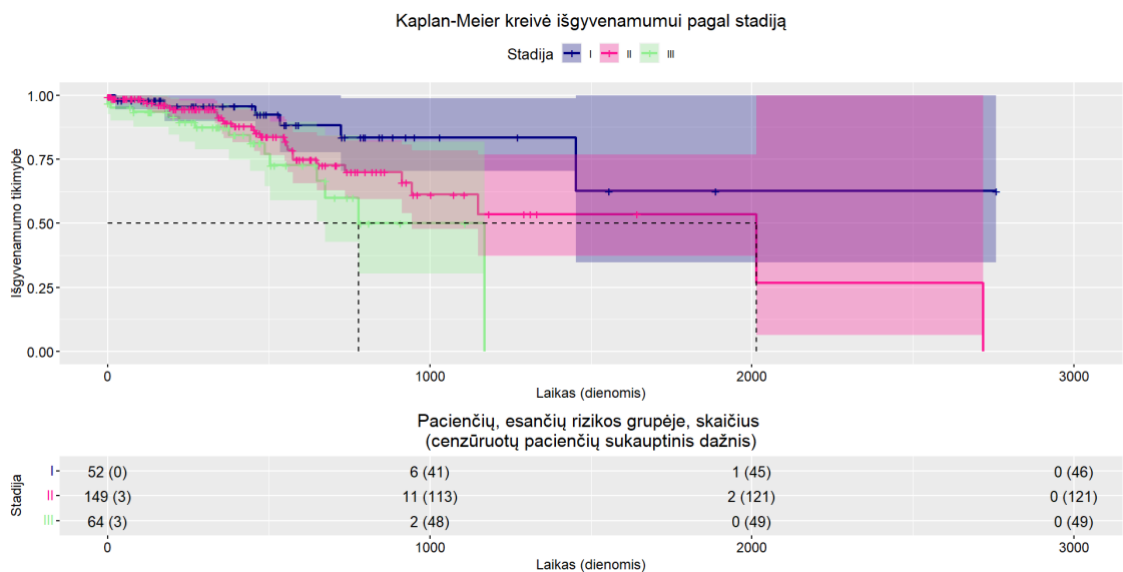
2.2 Kaplan - Meier kreivės

Taip pat buvo nubraižytos Kaplan – Meier kreivės, kad būtų galima pažiūrėti, kaip nuo kategorinių kintamųjų priklauso išgyvenamumo tikimybė bei apskritai kaip ji kinta. Buvo laikoma, jog įvykis – mirtis nuo krūties vėžio, o dar neįvykęs įvykis ir mirtis nuo kitos priežasties – cenzūruotas stebėjimas.



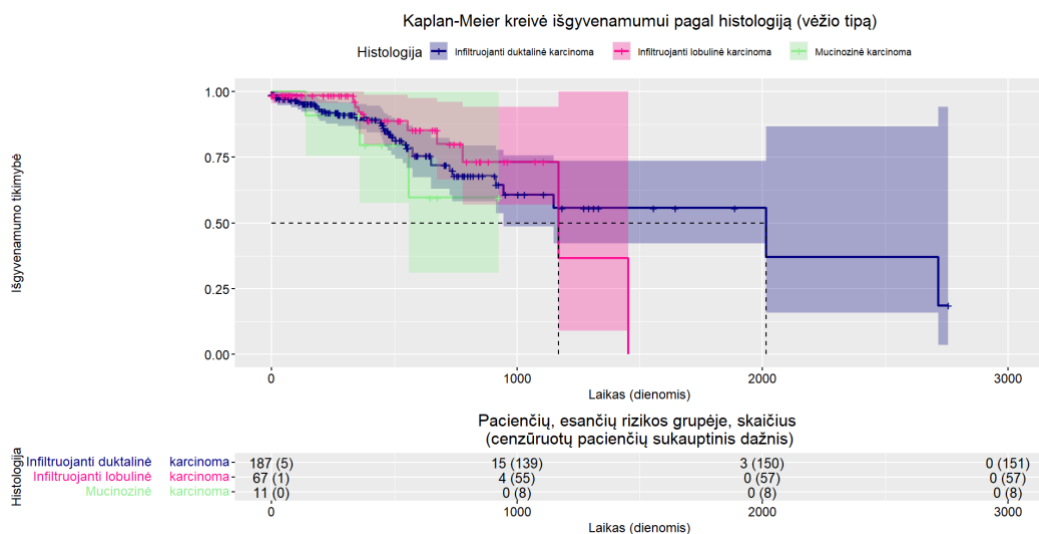
3 pav. Kaplan – Meier kreivė išgyvenamumui

Iš (3 pav.) matome, jog tikimybė išgyventi daugiau nei 1454 dienas (apie 4 m.) nukrenta iki 50 %. Paskutinė mirtis buvo 2717 dieną.



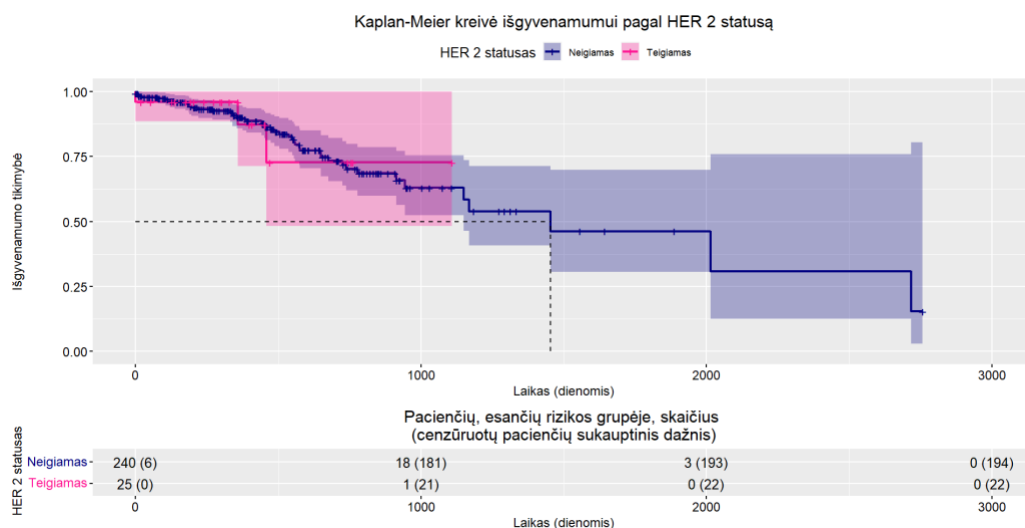
4 pav. Kaplan – Meier kreivė išgyvenamumui pagal stadiją

Iš (4 pav.) matome, jog 6 iš 52 pacientų mirė nuo vėžio, kai buvo diagnozuota I vėžio stadija, 28 iš 149 pacientėms buvo įvykis, kai jos sirgo II stadijos vėžiu, o 15 iš 64, kai buvo diagnozuota III vėžio stadija. Tikimybė, jog pacientė išgyvens ilgiau nei 2014 dienų (apie 5,5 m.), kai ji serga II vėžio stadija, nukrenta iki 50 %, o sergant III stadijos vėžiu tikimybė išgyventi ilgiau nei 778 dienas (apie 2 m.) nukrenta iki 50 %.



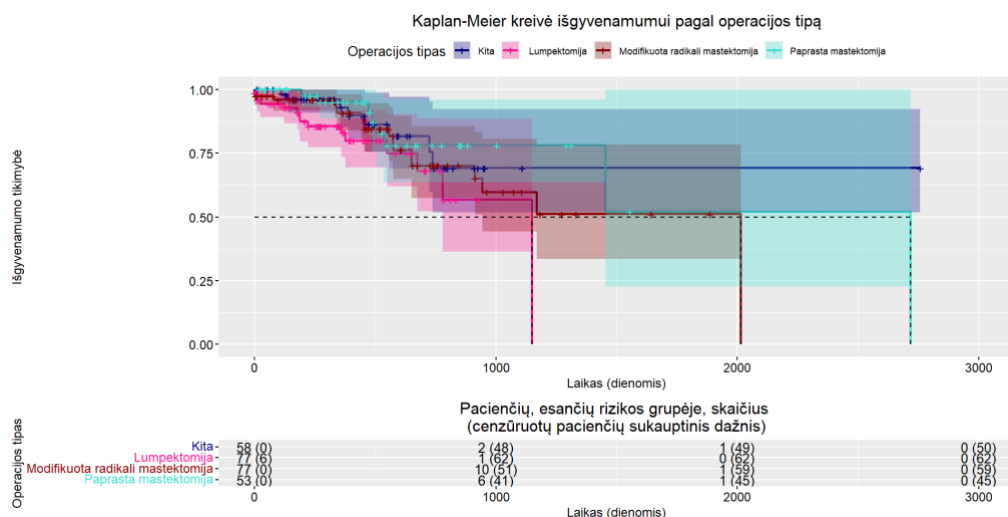
5 pav. Kaplan – Meier kreivė išgyvenamumui pagal histologiją

Iš (5 pav.) matome, jog 36 iš 187 pacienčių mirė nuo vėžio, kai sirgo infiltruojančia duktaline karcinoma, 10 iš 67 pacientėms buvo įvykis, kai jos infiltruojančia lobuline karcinoma, o 3 iš 11, kai buvo nustatyta mucinozinė karcinoma. Tikimybė, jog pacientė išgyvens ilgiau nei 2014 dienų (apie 5,5 m.), kai ji infiltruojančia duktaline karcinoma, nukrenta iki 50 %, o sergant infiltruojančia lobuline karcinoma, tikimybė išgyventi ilgiau nei 1169 dienas (apie 3 m.) nukrenta iki 50 %.



6 pav. Kaplan – Meier kreivė išgyvenamumui pagal HER2 statusą

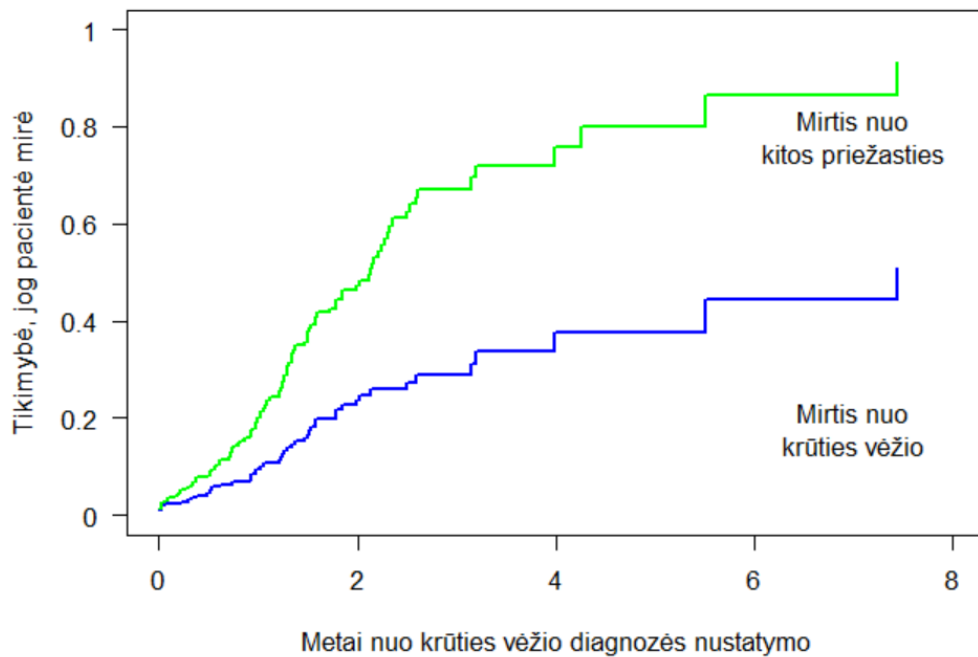
Iš 240 atvejų, kai pacientės HER2 statusas neigiamas, mirė 46, iš 25 atvejų kai HER2 statusas teigiamas – mirė 3 pacientės. Tikimybė, jog pacientė išgyvens ilgiau nei 1454 dienas (apie 4 m.), kai HER2 statusas yra neigiamas, nukrenta iki 50 %.



7 pav. Kaplan – Meier kreivė išgyvenamumui pagal operacijos tipą

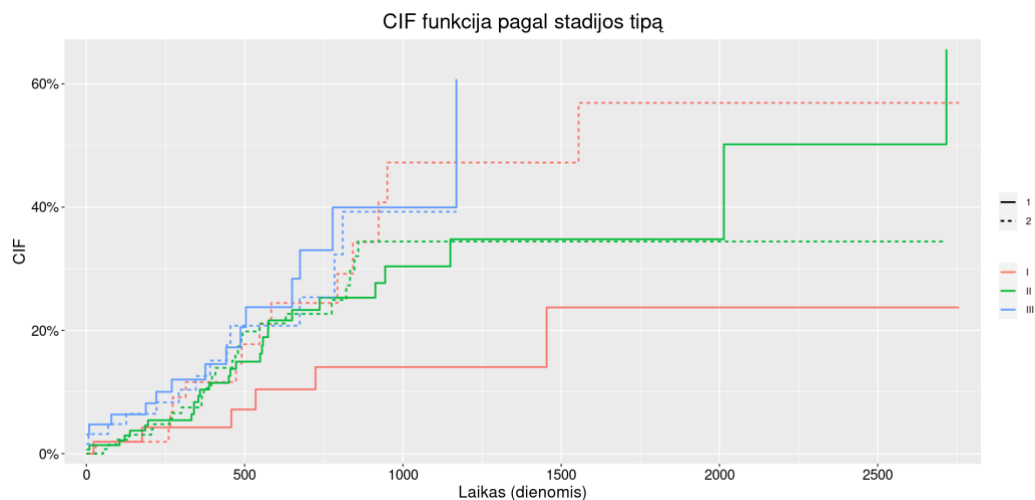
Iš (7 pav.) matome, jog 8 iš 58 pacienčių mirė nuo vėžio, kai joms buvo atlikta lumpektomija, 15 iš 77 pacienčių buvo įvykis, kai buvo atlikta modifikuota radikali mastektomija, 8 iš 53 pacienčių mirė, kai operacijos tipas buvo paprasta mastektomija, o 18 iš 77, kai buvo atlikta kita operacija. Tikimybė, jog pacientė išgyvens ilgiau nei 1150 dienų (apie 5,5 m.), kai jai buvo atlikta modifikuota radikali mastektomija, nukrenta iki 50 %, kai buvo atlikta paprasta mastektomija tikimybė išgyventi ilgiau nei 2717 dienų (apie 7,5 m.) nukrenta iki 50 %, o jei buvo atlikta kita operacija, tikimybė išgyventi ilgiau nei 2014 dienų (apie 5,5 m.) nukrenta iki 50 %.

2.3 CIF kreivės



8 8 pav. CIF funkcija abiem įvykių tipams (mirė nuo vėžio/nuo kitos priežasties)

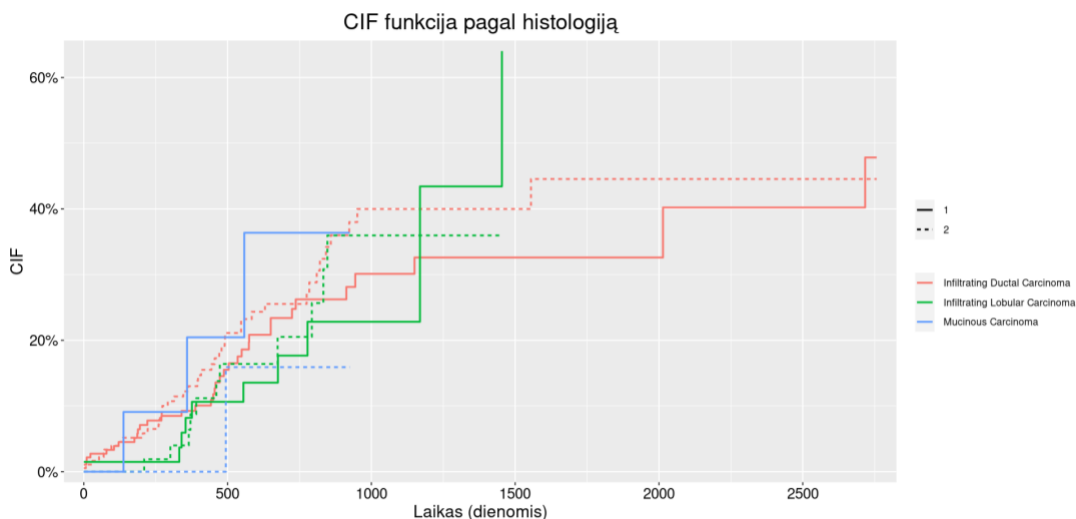
Iš (8 pav.) matome, jog tikimybė mirti nuo krūties vėžio yra 0,5, kai praėjo maždaug 8 metai nuo diagnozės pradžios, o sukauptinė tikimybė mirti dėl bet kurios iš dviejų priežasčių artėja į 1. Atstumas tarp mėlynos ir žalios kreivės – tikimybė mirti ne nuo vėžio (apie 0,4).



99 pav. CIF funkcija pagal stadijos tipą

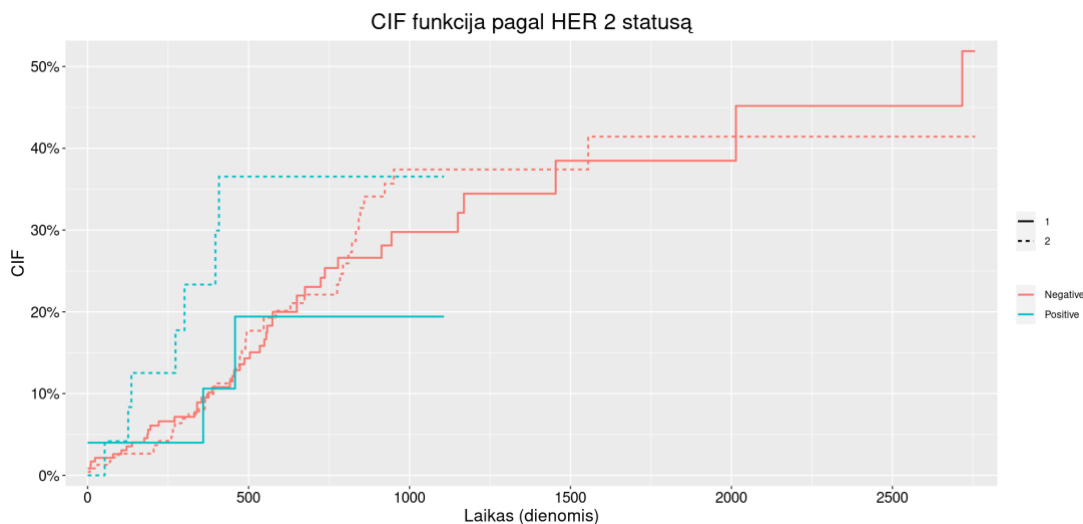
Jei pacientė serga I stadijos vėžiu, tikimybė mirti nuo krūties vėžio išauga iki 22 %, kai praėjo daugiau nei 2500 dienų (apie 7 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties išauga beveik iki 60 %. Kai diagnozuotas II stadijos vėžys, tikimybė mirti nuo krūties vėžio yra virš 60 %, kai praėjo daugiau nei 2500 dienų (apie 7 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties – beveik 40 %. Jei

pacientei buvo diagnozuotas III stadijos vėžys, tikimybė mirti nuo krūties vėžio išauga iki 60 %, kai praėjo daugiau nei 1000 dienų (apie 3 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties – apie 40 %. Patikrinus su Gray testu, ar rizika mirti po 3 m. yra tarp visų grupių vienoda gavome p reikšmę 0,029 – yra reikšmingi skirtumai tarp stadijų.



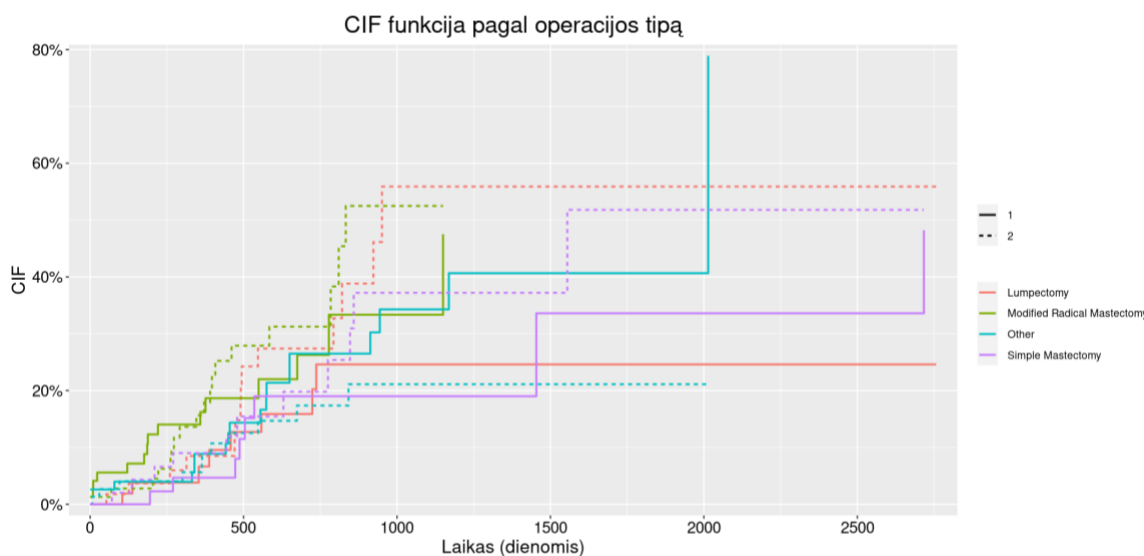
10 10 pav. CIF funkcija pagal histologiją

Jei pacientei buvo diagnozuota infiltruojanti duktalinė karcinoma, tikimybė mirti nuo krūties vėžio išauga iki 50 %, kai praėjo daugiau nei 2500 dienų (apie 7 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties išauga apie 45 %. Kai diagnozuota infiltruojanti lobulinė karcinoma, tikimybė mirti nuo krūties vėžio yra apie 40 %, kai praėjo daugiau nei 800 dienų (apie 2 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties – apie 20 %. Jei pacientei buvo diagnozuota mucinozinė karcinoma, tikimybė mirti nuo krūties vėžio išauga virš 60 %, kai praėjo daugiau nei 1400 dienų (apie 4 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties – apie 40 %. Patikrinus su Gray testu, ar rizika mirti po 3 m. yra tarp visų grupių vienoda gavome p reikšmę 0,6 – nėra statistiškai reikšmingų skirtumų tarp histologijos.



11 11 pav. CIF funkcija pagal HER2 statusą

Jei pacientės HER2 statusas yra neigiamas, tikimybė mirti nuo krūties vėžio išauga virš 50 %, kai praėjo daugiau nei 2500 dienų (apie 7 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties išauga apie 40 %. Kai HER2 statusas yra teigiamas, tikimybė mirti nuo krūties vėžio yra apie 20 %, kai praėjo daugiau nei 1000 dienų (apie 3 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties – apie 40 %. Patikrinus su Gray testu, ar rizika mirti po 3 m. yra tarp visų grupių vienoda gavome p reikšmę 0,7 – nėra statistiškai reikšmingų skirtumų tarp HER2 statusų.



12 12 pav. CIF funkcija pagal operacijos tipą

Jei pacientei buvo atlikta lumpektomija, tikimybė mirti nuo krūties vėžio išauga iki 25 %, kai praėjo daugiau nei 2500 dienų (apie 7 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties išauga apie 60 %. Kai buvo atlikta modifikuota radikali mastektomija, tikimybė mirti nuo krūties vėžio išauga iki 50 %, kai praėjo daugiau nei 1000 dienų (apie 3 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties

– virš 50 %. Jei pacientei buvo atlikta paprasta mastektomija, tikimybė mirti nuo krūties vėžio išauga iki 50 %, kai praėjo daugiau nei 2500 dienų (apie 7 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties – virš 60 %. Jei pacientei buvo atlikta kito tipo operacija, tikimybė mirti nuo krūties vėžio išauga iki 80 %, kai praėjo daugiau nei 2000 dienų (apie 5,5 m.) nuo diagnozės pradžios, o tikimybė mirti nuo kitos priežasties – virš 20 %. Patikrinus su Gray testu, ar rizika mirti po 3 m. yra tarp visų grupių vienoda gavome p reikšmę 0,2 – nėra statistiškai reikšmingų skirtumų tarp operacijos rūšių.

2.4 Fine ir Gray modelis

Turimiems duomenims taikysime Fine ir Gray modelį, konkuruojantiems įvykiams. Šis modelis bet kokio tipo įvykiui, modeliavo rizikos funkciją (*sub – distributional hazard*). Norint pritaikyti modelį patikrinsime multikolinearumo ir proporcingųjų rizikų prielaidas.

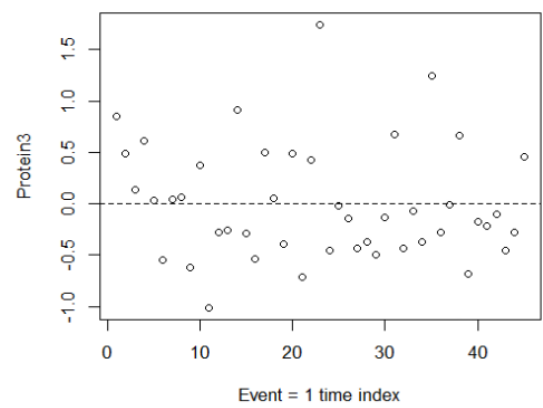
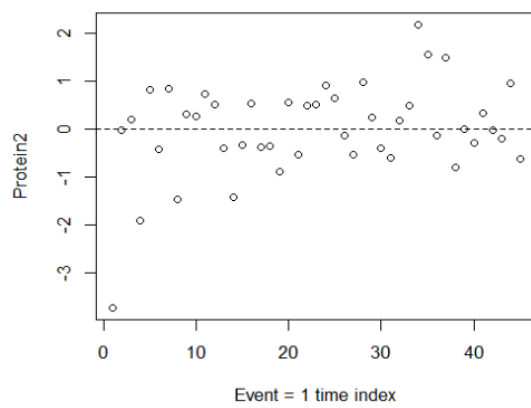
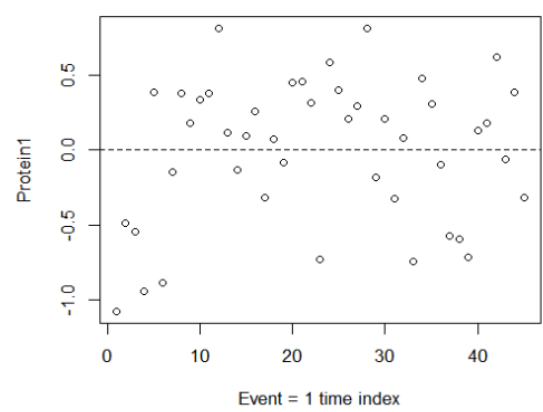
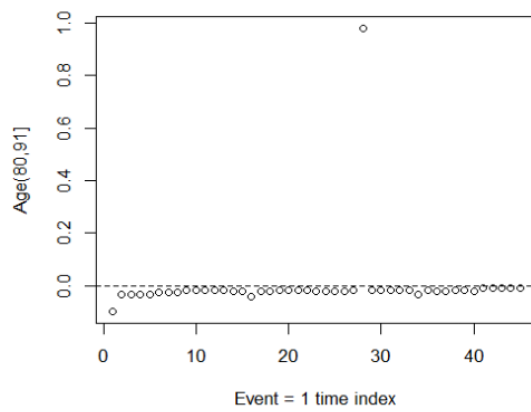
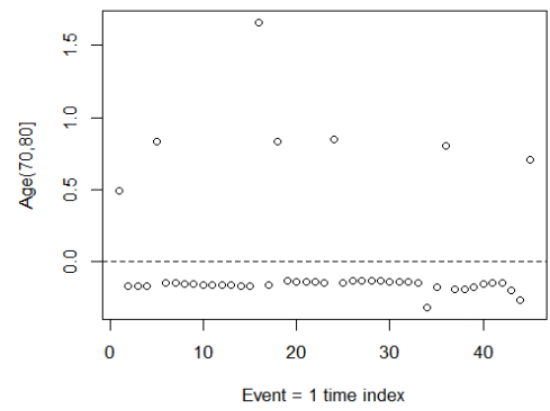
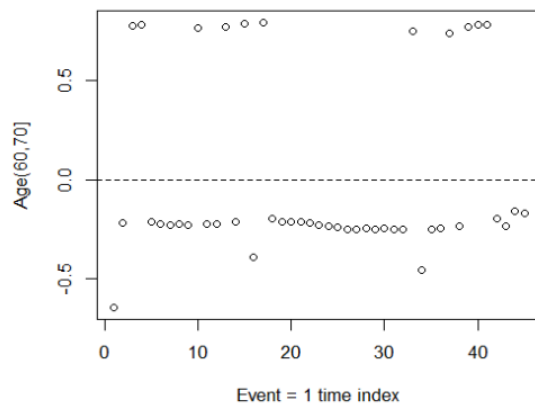
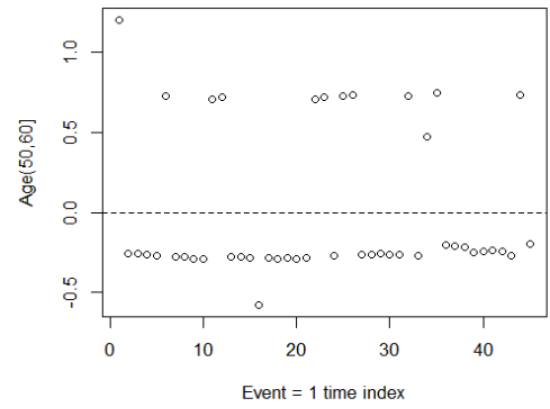
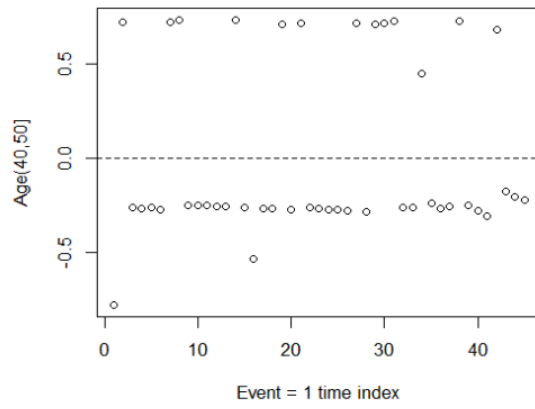
Multikolinearumą ištyrėme naudojantis pilnu modeliu ir modeliu su konkrečia kiekybine kovariante, žiūrėdami, ar koeficientų ženklai sutampa.

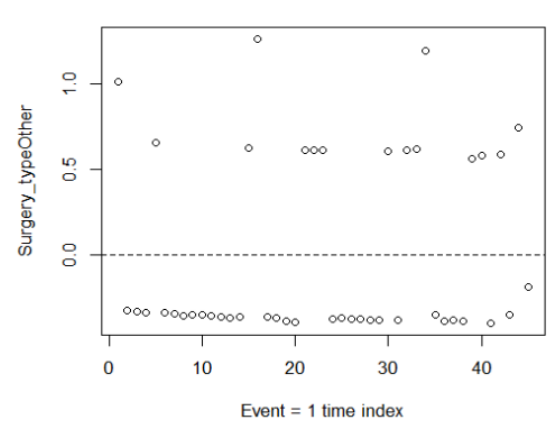
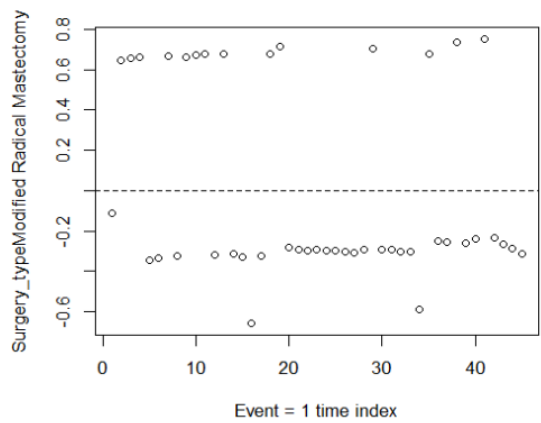
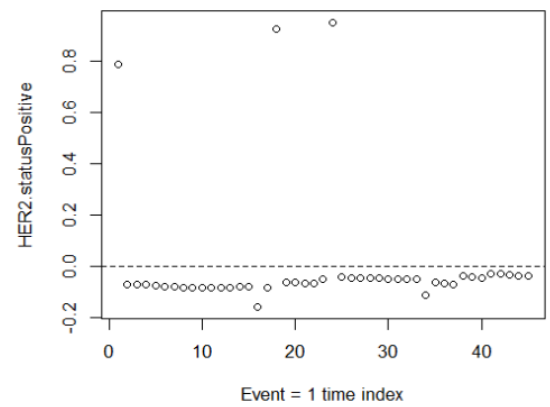
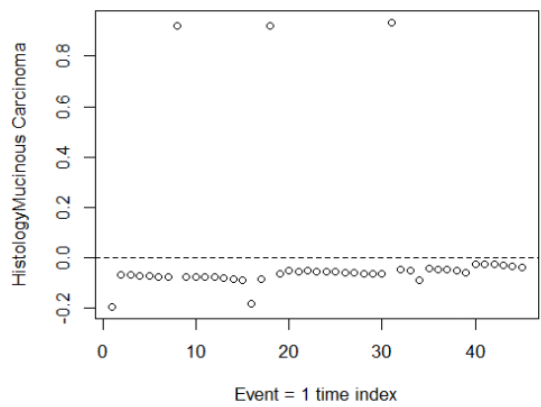
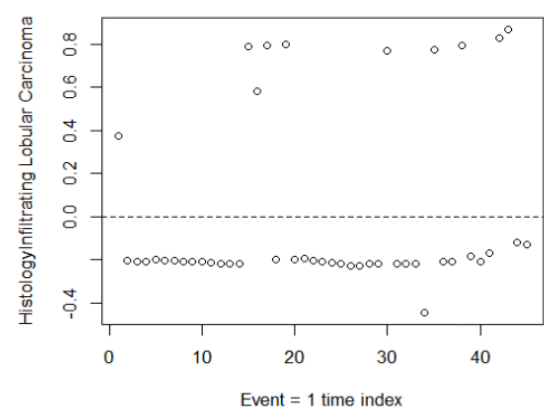
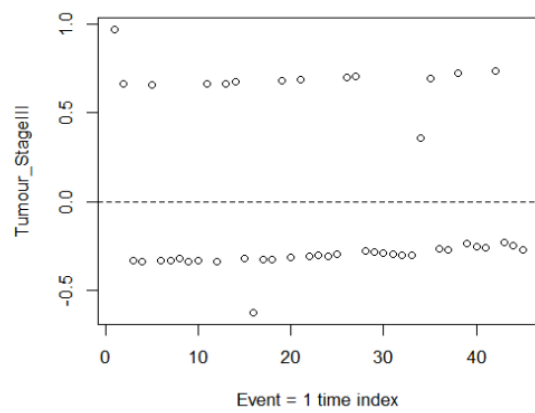
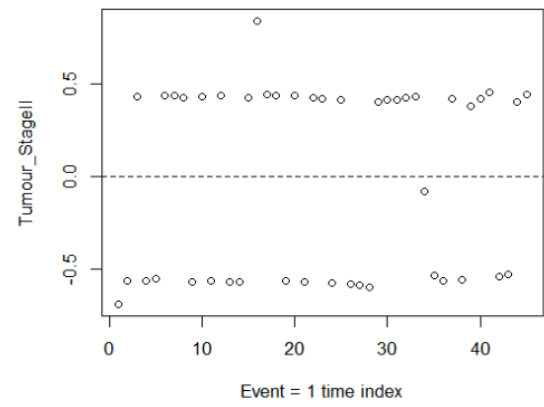
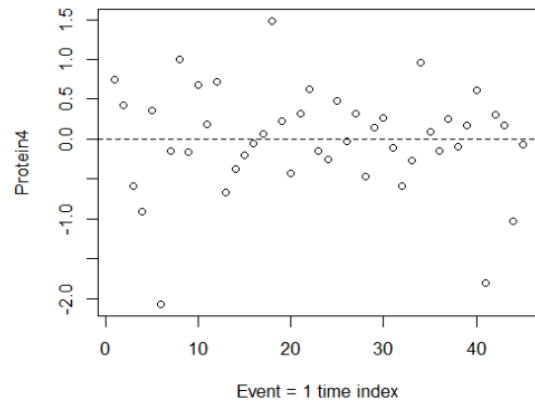
5 lentelė. Multikolinearumo problemos tikrinimas

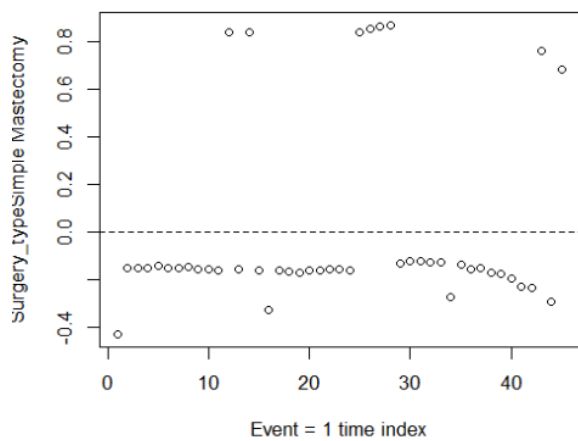
Kiekybinė kovariantė	Kai įvykis = 1 (mirtis nuo krūties vėžio)		Kai įvykis = 2 (mirtis ne nuo vėžio)	
	Ženklas modelyje su visomis kovariantėmis	Ženklas modelyje su viena kovariante	Ženklas modelyje su visomis kovariantėmis	Ženklas modelyje su viena kovariante
I proteino lygis	-	-	-	-
II proteino lygis	+	+	-	-
III proteino lygis	+	+	-	-
IV proteino lygis	+	+	+	-

Iš (5 lentelės) matome, jog IV proteino lygio ženklai pilname ir modelyje iš vieno kintamojo nesutampa. Dėl to, šios kovariantės nenaudosime konstruojant modelyje, kai įvykis traktuojamas mirties ne nuo vėžio.

Proporcingųjų rizikų prielaidą tikrinsime, naudojantis grafikais su *schoenfeld* liekanomis. Iš pradžių tikrinsime proporcingųjų rizikų prielaidą, kai įvykis – mirtis nuo vėžio.

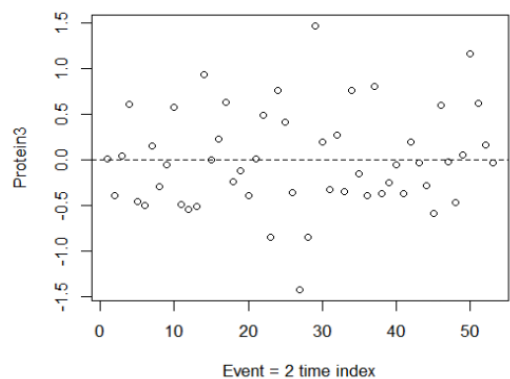
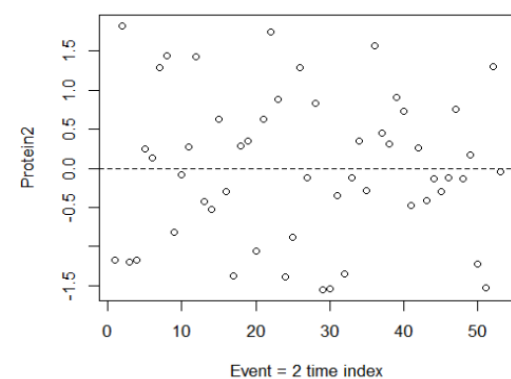
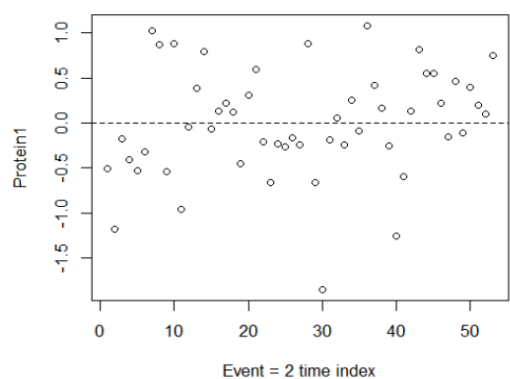
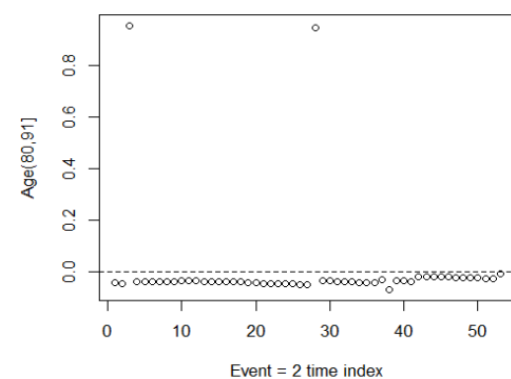
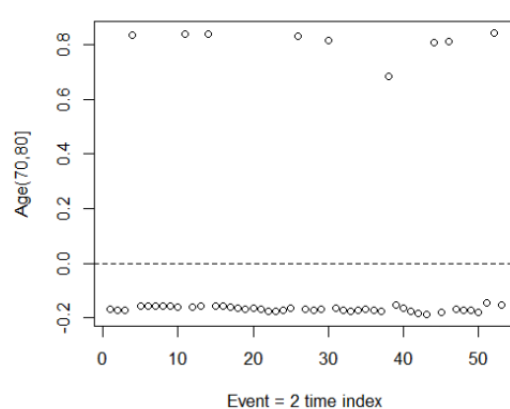
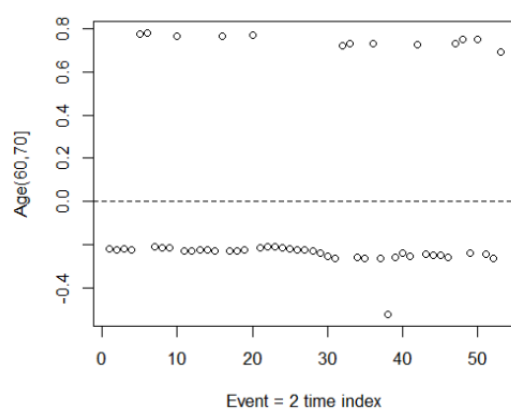
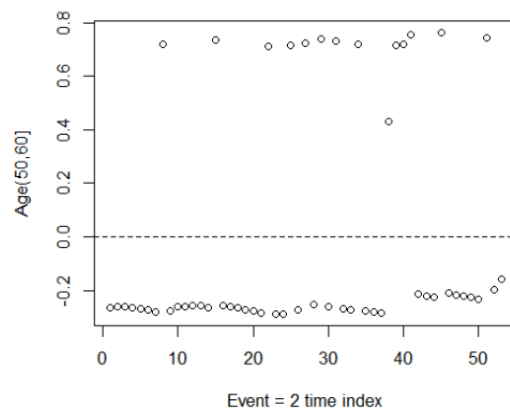
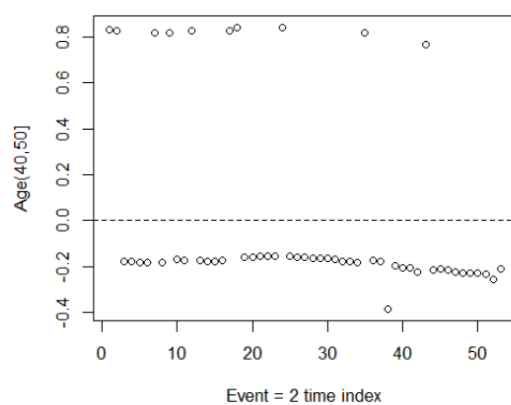


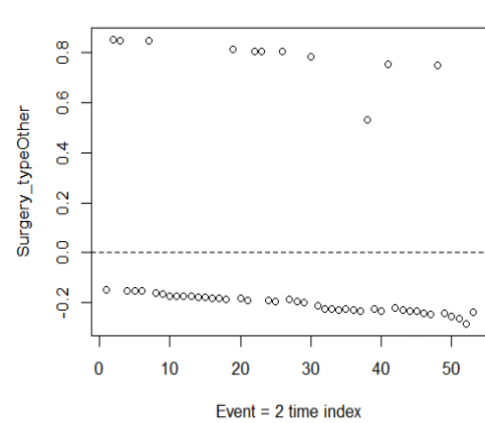
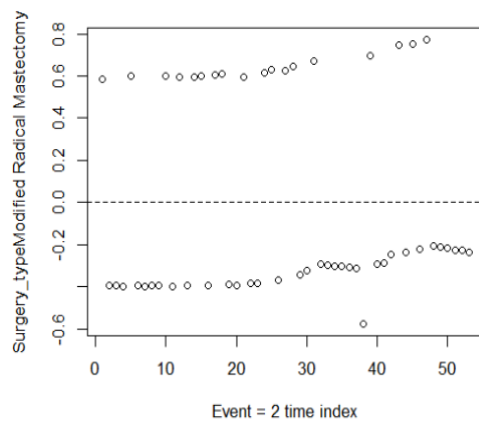
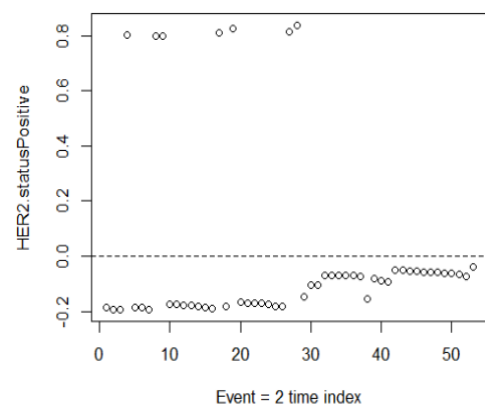
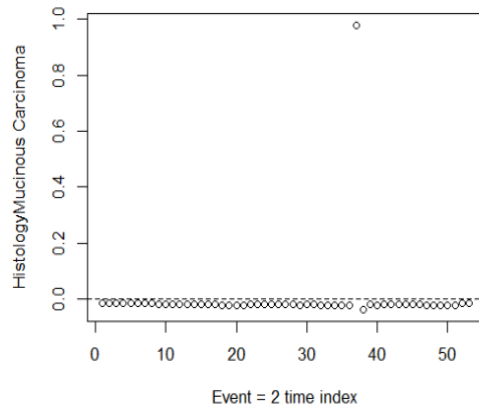
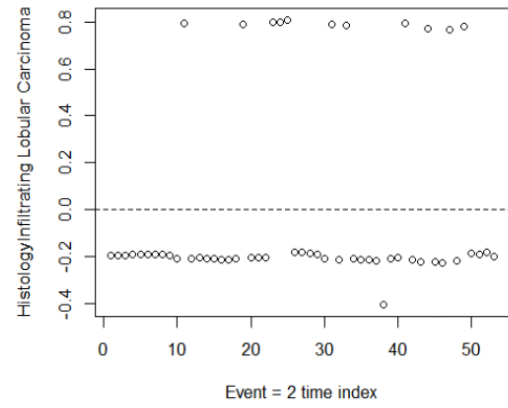
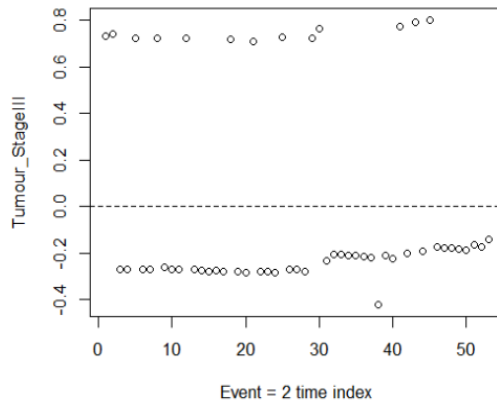
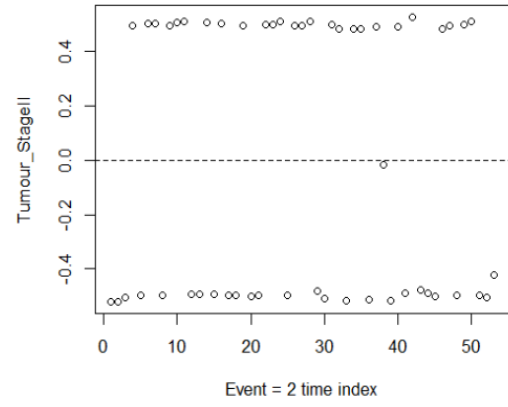
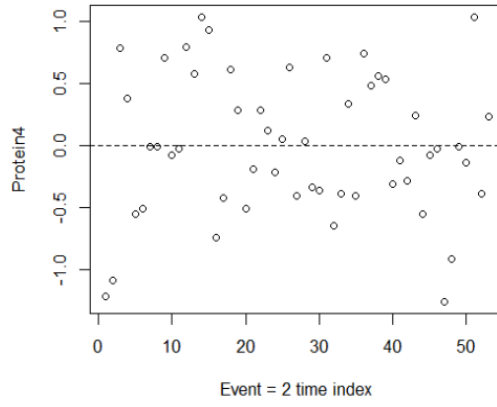


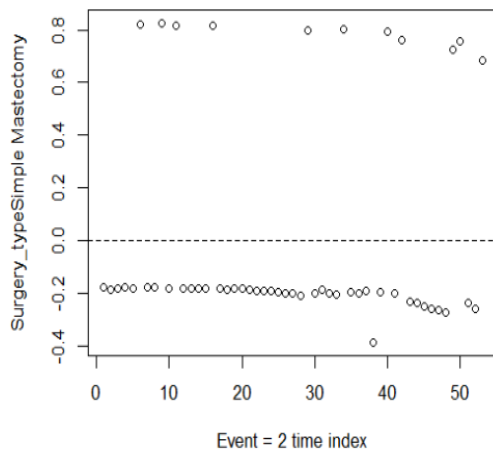


13 pav. Proporcingųjų rizikų prielaidos tikrinimas (I)

Iš pateiktų grafikų matosi, jog II proteino lygis gali netenkinti PH prielaidos, atsižvelgus į y ašies ribas, todėl šios kovariantės nenaudosime (modelyje ji buvo nereikšminga).







14 pav. Proporcinių rizikų prielaidos tikrinimas (II)

Toliau patikrinome proporcingųjų rizikų prielaidą, kai įvykis traktuojamas, kaip mirtis ne nuo vėžio. Iš gautų grafikų galime teigti, jog visos kovariantės tenkina prielaidą.

Kiekvienam galimam įvykiui atrinkome reikšmingas kovariantes – šalinome po vieną kovariantę, kurios p reikšmė buvo didžiausia.

Gautas modelis, kai įvykis yra mirtis nuo vėžio, su reikšmingomis kovariantėmis, kai reikšmingumo lygmuo 0,05:

```
> summary(FGR(Hist(days_of_last_follow_up, Patient_Status)~ Protein4+
+ Tumour_Stage,data=train, cause=1))#visi reikšmingi
Competing Risks Regression
```

```
Call:
FGR(formula = Hist(days_of_last_follow_up, Patient_Status) ~
      Protein4 + Tumour_Stage, data = train, cause = 1)
```

	coef	exp(coef)	se(coef)	z	p-value
Protein4	0.609	1.84	0.303	2.01	0.044
Tumour_StageII	0.832	2.30	0.461	1.81	0.071
Tumour_StageIII	1.177	3.24	0.499	2.36	0.018

Iš pateiktų koeficientų ženklų matome, jog padidėjus IV proteino lygiui rizika numirti nuo krūties vėžio. Taip pat didėja rizika numirti dėl krūties vėžio, sergant 2 stadija ir 3 stadija, lyginant su 1 stadija. Tikslėni įvertinimą galime pateikti pažvelgę į $\exp(coef)$ – rizikos funkcijų santykį. Matome, jog padidėjus 1 vnt. IV proteino lygio koncentracijos rizika numirti padidėja 84 %, sergant 1 stadija (lyginamoji stadija). Taip pat galime pastebėti, jog sergant 2 stadija rizika numirti yra 2,3 karto didesnė nei sergant 1 stadija, o sergant 3 stadija rizika numirti yra 3,24 karto didesnė nei lyginant su 1 stadija.

Modelio lygtis:

$$h_{1,CIC} = h_{01,CIC}(t) \times e^{0,609 \times IV \text{ proteinas} + 0,832 \times Stadija(2) + 1,177 \times Stadija(3)}.$$

Gautas modelis, kai įvykis yra mirtis ne nuo vėžio, su reikšmingomis kovariantėmis, kai reikšmingumo lygmuo 0,15:

```
> summary( with(train, crr(days_of_last_follow_up, Patient_Status, cov1 = model.matrix(m2)[,c(14,15,16,17)],
+                               failcode = 2, cencode = 0)), Exp=T)#HER2 ir surgery kai al
fa0,15
Competing Risks Regression

Call:
crr(ftime = days_of_last_follow_up, fstatus = Patient_Status,
    cov1 = model.matrix(m2)[, c(14, 15, 16, 17)], failcode = 2,
    cencode = 0)
```

	coef	exp(coef)	se(coef)	z	p-value
HER2.statusPositive	0.646	1.908	0.443	1.458	0.140
Surgery_typeModified Radical Mastectomy	0.196	1.217	0.344	0.571	0.570
Surgery_typeOther	-0.719	0.487	0.403	-1.783	0.075
Surgery_typeSimple Mastectomy	-0.162	0.850	0.389	-0.417	0.680

Iš pateiktų koeficientų matome, jog turint teigiamą HER2 statusą, lyginant su neigiamu, rizika numirti ne nuo vėžio yra didesnė. Jei operacija buvo modifikuota radikali mastektomija lyginant su lumpektomija, rizika numirti didėja, o, jei operacija buvo paprasta mastektomija ar kitos rūšies rizika numirti mažėja lyginant su operacija – lumpektomija (lyginamoji kategorija). Pažvelgę į stulpelį $\exp(coef)$ galime daryti išvadas, jog turint teigiamą HER2 statusą rizika numirti padidėja 91 %. Jei operacija buvo modifikuota radikali mastektomija 22 % padidėja rizika numirti, lyginant su lumpektomijos operacija. Jei operacija buvo kitos rūšies lyginant su lumpektomija, rizika numirti ne nuo vėžio sumažėja apie 51 %. Taip pat, jei buvo daryta paprasta mastektomija lyginant su lumpektomija, rizika numirti ne nuo vėžio mažėja 15 %.

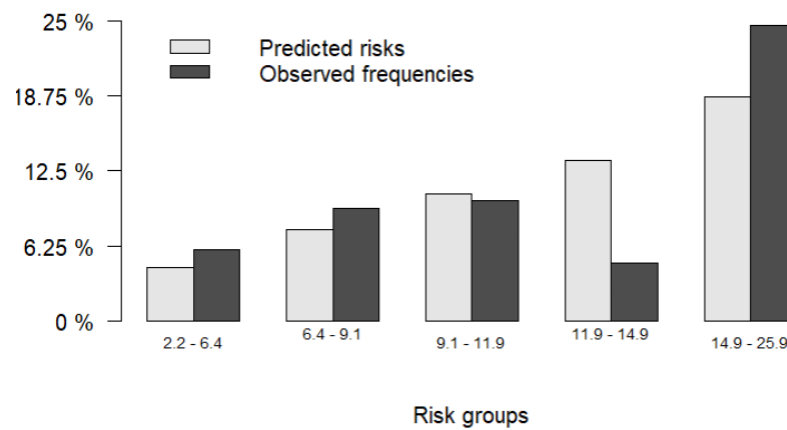
Modelio lygtis:

$$h_{2,CIC} = h_{02,CIC}(t) \times e^{0,646 \times HER2Status(+) + 0,196 \times Operacija(Modifikuota radikali mastektomija) - 0,719 \times Operacija(kita) - 0,162 \times Operacija(paprasta mastektomija)}.$$

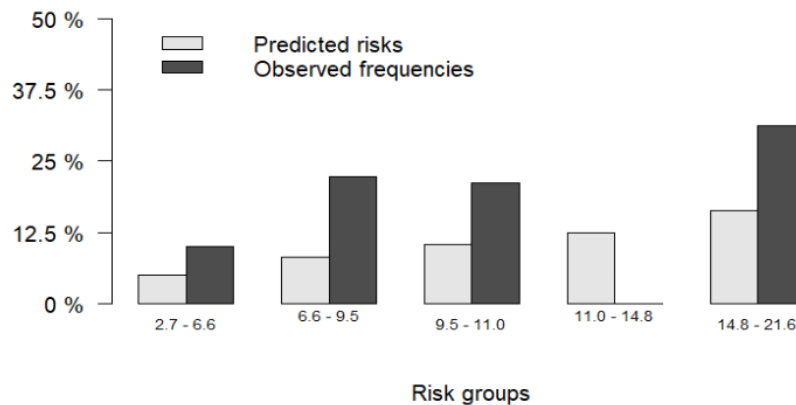
Kiekvienam modeliui paskaičiavome konkordacijos indeksą. Iš lentelės matome, jog gautas geresnis modelis, kai įvykis – mirtis ne nuo vėžio, nors abu modeliai nėra pakankamai geri.

6 lentelė. Konkordacijos indeksas

	1 įvykiui mokymosi aibėje	1 įvykiui testavimo aibėje	2 įvykiui mokymosi aibėje	2 įvykiui testavimo aibėje
Konkordacijos reikšmė	0,65	0,52	0,63	0,68

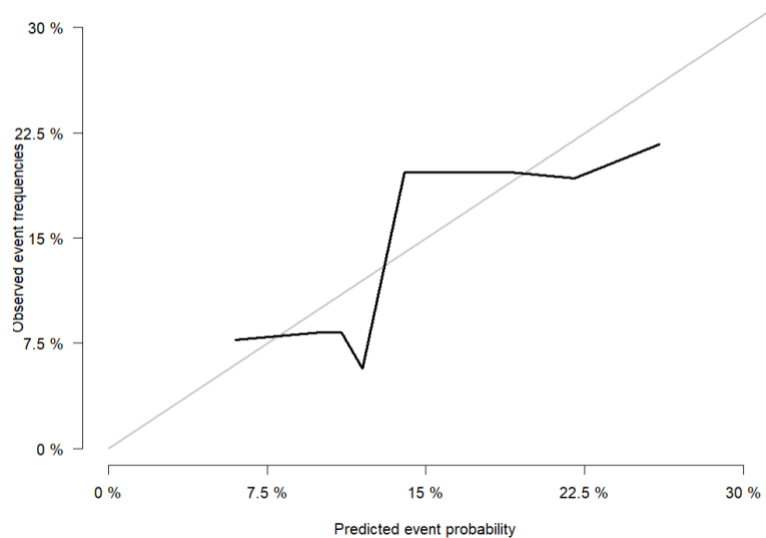


15 15 pav. Prognozės tikslumas, kai įvykis - mirtis nuo vėžio mokymosi imtyje

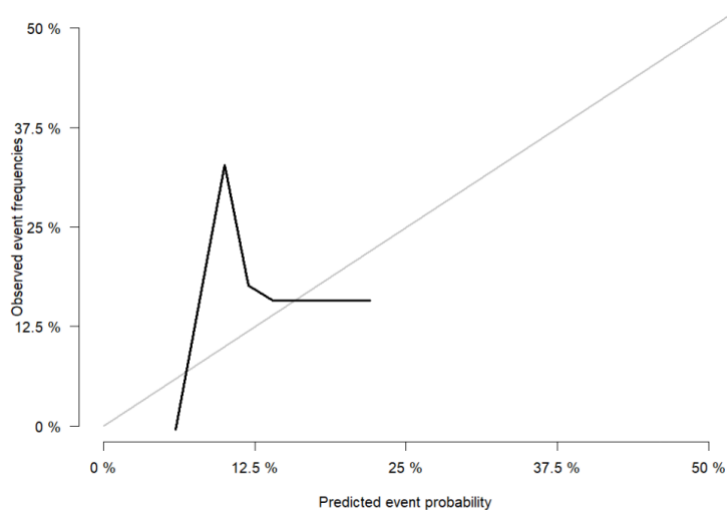


16 16 pav. Prognozės tikslumas, kai įvykis - mirtis nuo vėžio testavimo imtyje

Iš (15 ir 16 pav.) matome prognozavimo tikslumus, kai įvykis – mirtis nuo vėžio, mokymosi ir testavimo imtyje. Šiuose grafikuose buvo paimti 5 intervalai, kuriuose suskaičiavo vidurkius prognozuotos rizikos. X ašyje pateikiama rizika numirti nuo vėžio procentais. Modelis mokymosi imtyje prognozuoja neblogai, tačiau testavimo imtyje pasirodo prastai.



17 17 pav. Prognozės tikslumas, kai įvykis - mirtis ne nuo vėžio mokymosi imtyje

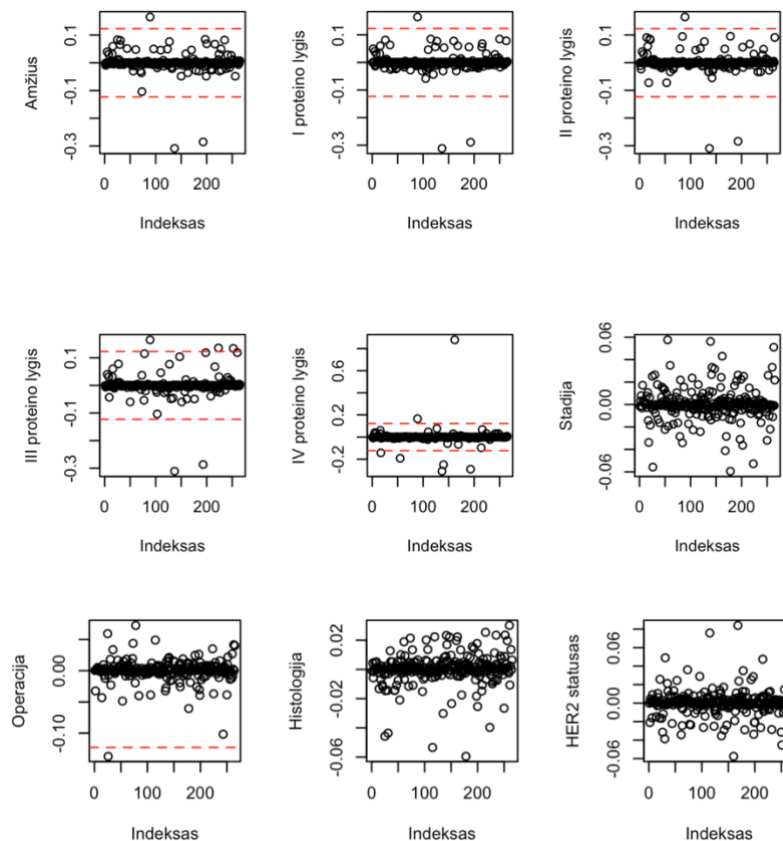


18 18 pav. Prognozės tikslumas, kai įvykis - mirtis ne nuo vėžio testavimo imtyje

Iš (17 ir 18 pav.) matome, jog rizikos prognozės nėra tikslios, prognozuoja žymiai blogiau testavimo imtyje.

2.6 Kokso modelis

Pabandykime sukonstruoti atskirus Kokso modelius kiekvienam įvykiui. Pradėkime, kai įvykis lygus 1. Kad tai padaryti pirmiausia patikrinime prielaidas: išskirtis, tiesiškumą bei proporcingosios rizikos prielaidą be to iš karto tarkime, jog nepriklausomumo prielaida patenkinta, nes mirti nuo vėžio rizika yra nesusijusi su rizika mirti dėl kitų priežasčių. Kitų prielaidų tikrinimą pradėkime nuo išskirčių - tam naudosime DFBet'ų statistiką. Gavome slenksčius $[-0,12; 0,12]$, kurios viršijant, stebėjimas laikomas išskirtimi.



19 19 pav. *DFBetų statistikos*

Iš 19 pav. matome, jog išskirčių turi tokios kovariantės kaip amžius, visi proteino lygiai bei operacija. Be to, daugumos kovariančių išskirtys turi vienodus indeksus, todėl visus juos pašalinkime (buvo pašalinta 10 išskirčių). Dabar patikrinime ar kiekybiniai kintamieji neturi multikolinerumo problemos.

7 lentelė. *VIF statistika*

Kovariantės	I proteino lygis	II proteino lygis	III proteino lygis	IV proteino lygis
VIF	1,17	1,30	1,26	1,18

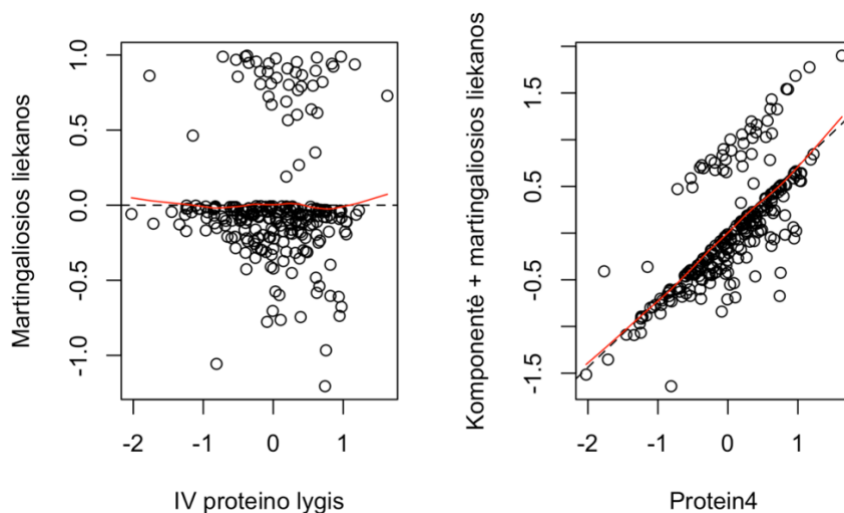
Kaip galime matyti iš lentelės, nei vienos kovariantės VIF koeficientas neviršija 4, o tai rodo, jog multikolinerumo problemos nėra.

Galiausiai atrinkime kovariantes, kurios yra statistiškai reikšmingos bei tenkina proporcingųjų rizikų prielaidą, su reikšmingumo lygmeniu 0,05.

8 lentelė. *Proporcingųjų rizikų prielaida*

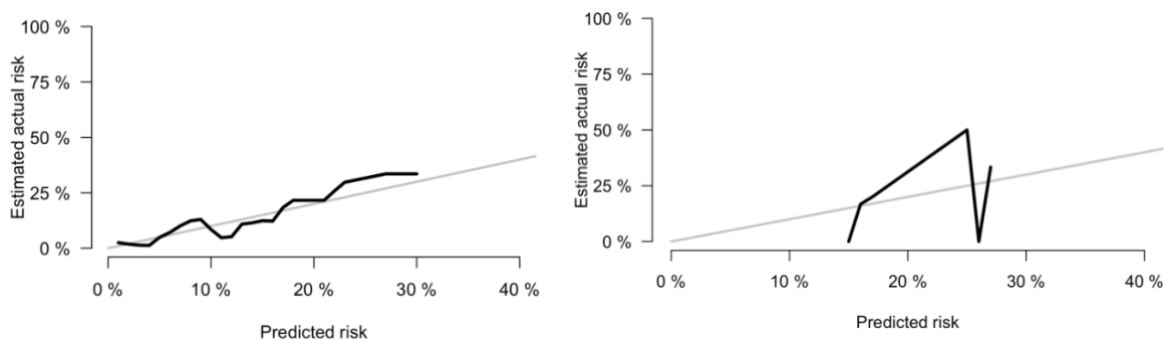
Kovariantė	P reikšmė	Proporcingųjų rizikų p reikšmė
IV proteino lygis	0,01	0,45
Stadija	0,01	0,72

Kadangi kovariantė IV proteino lygis yra kiekybinė - reikia patikrinti tiesiškumą. Ta darysime su marginaliosiomis liekanomis.



20 20 pav. Marginaliosios liekanos

Iš 20 pav. galime matyti, jog prielaida tenkinama. Dabar galime palyginti prognozės tikslumus testinėje ir mokymosi aibėse.



21 21 pav. Prognozės tikslumas mokymosi (kairėje) ir testinėje (dešinėje) imtyse

Kaip galime matyti iš 21 pav. modelis geriau prognozuoja mokymo imtyje nei testinėje.

9 lentelė. Eksponentės

Kovariantė	β	$\exp(\beta)$
IV proteino lygis	0,8226	2,2764
Stadija II	1,3158	3,7278
Stadija III	1,7399	5,6968

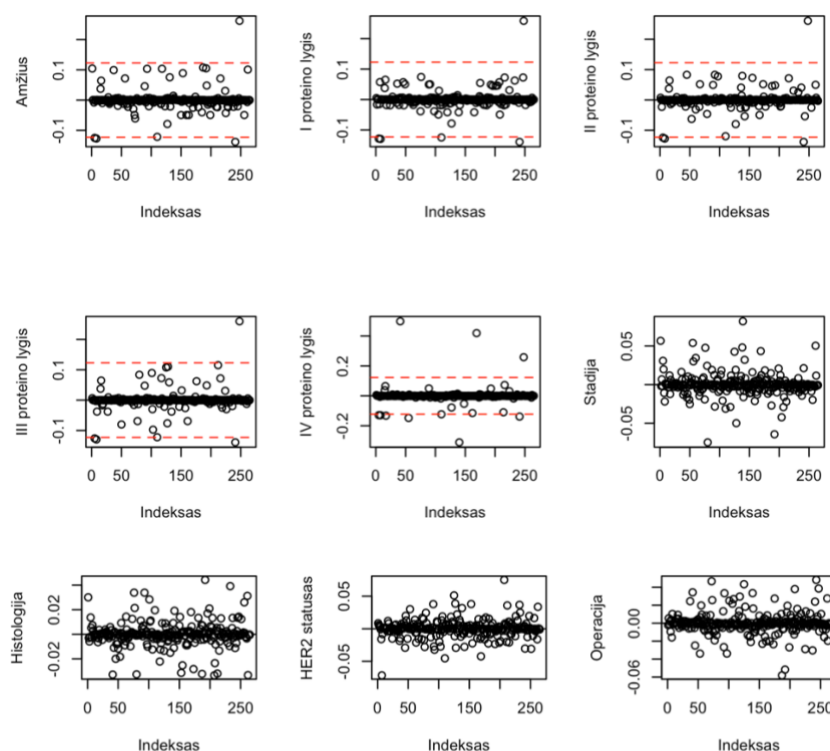
Iš lentelės matome, jog padidėjus IV proteino lygiui rizika mirti didėja, taip pat sergant II stadijos vėžiu rizika didėja lyginant su I stadijos vėžiu. Sergant III stadijos vėžiu taip pat rizika numirti nuo vėžio didėja nei sergant I stadijos vėžiu. Taip pat galime pastebėti, jog padidėjus proteino IV lygiui 1 vnt. rizika numirti padidėja 2,28 karto. Sergant II stadijos vėžiu rizika numirti yra 3,73 karto didesnė, o sergant III stadijos vėžiu rizika numirti dėl vėžio yra 5,7 k. didesnė nei sergant I stadijos vėžiu.

Galutinis modelis atrodo taip:

$$h(t|\mathbf{x}) = h_0(t) \times e^{(0,8226 \times IV \text{ proteino lygis} + 1,3158 \times Stadija(II) + 1,7399 \times Stadija(III))}$$

Gauto modelio Akaikės informacinis indeksas mokymo imtyje lygus 361,94, o Harrello konkordacijos indeksas – 0,679.

Dabar tuos pačius žingsnius pakartokime kai įvykis lygus 2. Pradėkime nuo išskirčių.



22 22 pav. *DFBetų statistikos*

Iš 22 pav. galime matyti, jog išskirčių turi amžius bei visi proteino lygiai, kadangi dauguma indeksu sutampa, juos pašalinsime (pašalinta 10 išskirčių). Dabar patikrinkime ar kiekybiniai kintamieji neturi multikolinerumo problemos.

10 10 lentelė. *VIF statistika*

Kovariantės	I proteino lygis	II proteino lygis	III proteino lygis	IV proteino lygis
-------------	------------------	-------------------	--------------------	-------------------

VIF	1,04	1,20	1,20	1,02
------------	------	------	------	------

Kadangi visi VIF koeficientai neviršija 4, tai galime teigti, jog multikolinerumo problemos – nėra. Galiausiai atrinkime reikšmingas bei proporcingųjų rizikų prielaidą tenkinančias kovariantes su reikšmingumo lygmeniu 0,10.

1111 lentelė. Proporcingųjų rizikų prielaida

Kovariantė	P reikšmė	Proporcingųjų rizikų p reikšmė
Operacija	0,06	0,59

Kaip matome iš 10 lentelės, kai įvykis 2, gauta tik viena reikšminga ir proporcingųjų rizikų prielaidą tenkinanti kovariantė.

12 12 lentelė. Eksponentės

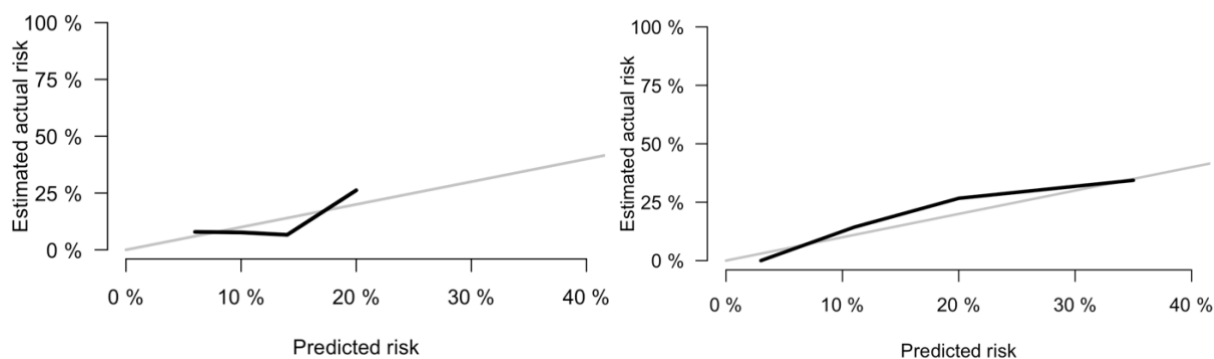
Kovariantė	β	$\exp(\beta)$
Operacija (kitos)	-0,9297	0,3947
Operacijos (modifikuota radikali mastektomija)	0,4278	1,5338
Operacijos (paprasčia mastektomija)	-0,3035	0,7383

Iš lentelės galime matyti, jog rizika numirti nuo ne vėžio didėja, jei operacija buvo modifikuota radikali mastektomija, o, jei buvo kitos rūšies ar paprasčia mastektomija mažėja, lyginant su lumpektomija. Apie 40 % yra mažesnė rizika numirti, jei operacija buvo kitos rūšies ir apie 26 % yra mažesnė rizika numirti ne nuo vėžio, jei operacija buvo paprasčia mastektomija, lyginant su lumpektomijos operacija. Apie 53 % yra didesnė rizika numirti, jei buvo atlikta modifikuota radikali mastektomija nei lumpektomija.

Sudarykime vieno kintamojo modelį:

$$h(t|x) = h_0(t) \times e^{\left(\begin{matrix} -0.9297 \times \text{Operacija(kitos)} + 0.4278 \times \text{Operacija(modifikuota radikali mastektomija)} \\ -0.3035 \times \text{Operacija(paprasčia mastektomija)} \end{matrix} \right)}.$$

Gauto modelio Akaikės informacinis indeksas mokymo imtyje lygus 414,03, o Harrello konkordacijos indeksas – 0,61. Taip pat pažiūrėkime modelio prognozavimo tikslumą.



23 23 pav. Prognozės tikslumas mokymo (kairėje) ir testinėje (dešinėje) imtyse

Kaip matome iš 23 pav. prognozavimas testinėje aibėje visai neblogas.

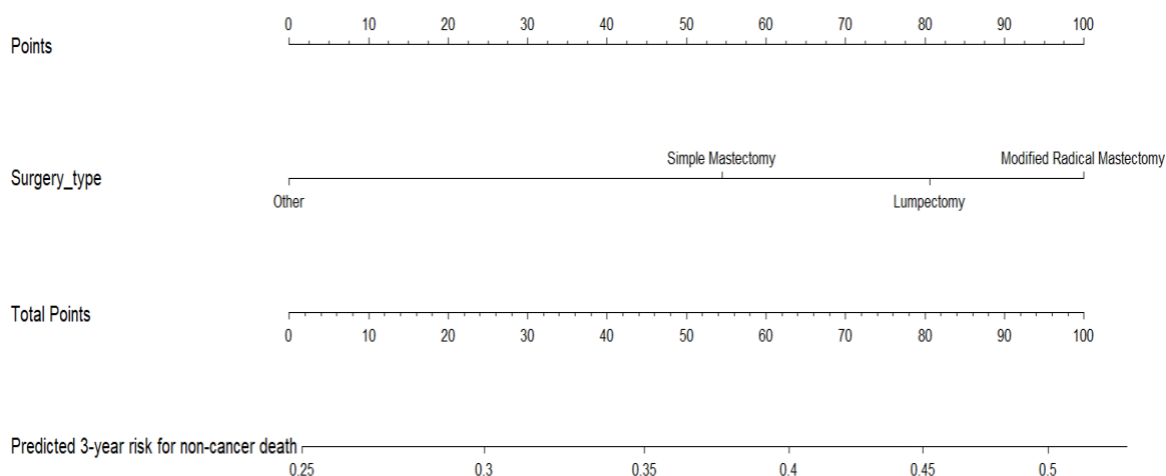
Dabar palyginkime gautus modelius.

13 13 lentelė. Modelių palyginimas mokymo imtyje

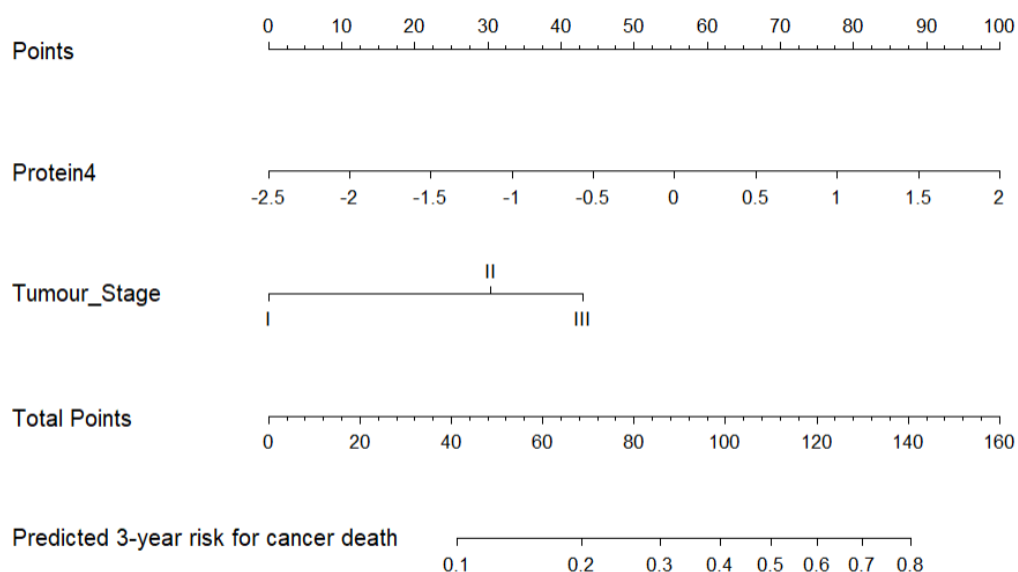
	AIC	Harrello konkordacijos indeksas
Kai įvykis = 1	361,94	0,68
Kai įvykis = 2	414,03	0,61

Kaip galime matyti iš AIC, modelis, kai įvykis – mirtis ne nuo vėžio yra geresnis nei modelis su įvykiu – mirtis nuo vėžio.

Taip pat gautiems modeliams sukūrėme nomogramas, kurios įvertina 3 m. laikotarpiu riziką numirti dėl tam tikros priežasties, kai galima ir kita priežastis.



24 pav. Nomograma mirčiai ne nuo vėžio



25 pav. Nomograma mirčiai nuo vėžio

3. IŠVADOS

Iš viso yra 330 stebėjimų, pašalinus praleistas reikšmes stebėjimų liko 317. Duomenys buvo padalinti į mokymo ir testavimo aibes santykiu 80 : 20.

Iš pradinės duomenų analizės pastebėta, jog pacientės, mirusios nuo krūties vėžio ir nuo kitos priežasties, dažniausiai buvo 40 – 70 metų. Daugiausiai pacienčių (112), kurioms įvykis dar neįvyko sirgo II stadijos vėžiu, 36 mirė nuo krūties vėžio kai joms buvo diagnozuotas II stadijos vėžys ir 17 – kai buvo diagnozuotas III stadijos vėžys. Nuo krūties vėžio mirusios pacientės dažniausiai turėjo infiltruojančios duktalinės karcinomos vėžį, 14 iš jų mirė nuo infiltruojančios lobulinės karcinomos ir 3 – nuo mucinozinės karcinomos. Pacientėms, mirusioms nuo vėžio dažniausiai buvo atlikta modifikuota radikali mastektomija arba kito tipo operacija.

Nusibraižius Kaplan – Meier kreives tikimybė išgyventi daugiau nei 1454 dienas (apie 4 m.) nukrenta iki 50 %, kai įvykis – mirtis nuo krūties vėžio, o dar neįvykęs įvykis ir mirtis nuo kitos priežasties – cenzūruotas stebėjimas.

Taikant Kokso modelius, kai įvykis – mirtis nuo vėžio bei kai įvykis – mirtis ne nuo vėžio, buvo gauta, jog modelis su antruoju įvykiu buvo geresnis, t. y. gautas didesnis Akaikės informacinis indeksas bei iš grafikų buvo galima matyti geresnį prognozavimą. Visos prielaidos buvo tenkinamos. Gauta modelio lygtis, kai įvykis – mirtis nuo vėžio:

$$h(t|\mathbf{x}) = h_0(t) \times e^{(2,275 \times IV \text{ proteino lygis} + 3,728 \times Stadija(II) + 5,697 \times Stadija(III))}.$$

Lygtis kai įvykis – mirtis ne nuo vėžio:

$$h(t|\mathbf{x}) = h_0(t) \times e^{\left(\begin{array}{c} -0.9297 \times \text{Operacijos tipas(kitos)} + 0.4278 \times \text{Operacijos(modifikuota radikali mastektomija)} \\ -0.3035 \times \text{Operacijos(paprasta mastektomija)} \end{array} \right)}.$$

Taikant Fine ir Gray modelį, II proteino lygio kovariantė netenkino PH prielaidos, kai įvykis – mirtis nuo vėžio, o IV proteino lygis buvo multikolinearus, kai įvykis – mirtis ne nuo vėžio. Gauta modelio lygtis, kai įvykis – mirtis nuo vėžio:

$h_{1,CIC} = h_{01,CIC}(t) \times e^{0,609 \times IV \text{ proteinas} + 0,832 \times Stadija(2) + 1,177 \times Stadija(3)}$. Šio modelio konkordacijos indeksas – 0,62 (mokymo imtyje) ir 0,52 (testavimo imtyje). Gauta modelio lygtis, kai įvykis – mirtis ne nuo vėžio:

$$h_{2,CIC} = h_{02,CIC}(t) \times e^{\left(\begin{array}{c} 0,646 \times HER2Status(+) + 0,196 \times Operacija(Modifikuota radikali mastektomija) \\ -0,719 \times Operacija(kita) - 0,162 \times Operacija(paprasta mastektomija) \end{array} \right)}.$$

Šio modelio konkordacijos indeksas – 0,63 (mokymo imtyje) ir 0,68 (testavimo imtyje).