



Perearstiabi patsientide pöördumiskordade arvu ja haigusjuhu kestuse prognoosid Eesti Haigekassa ning tervise infosüsteemi andmete põhjal

Tervise Arengu Instituut

**Perearstiabi patsientide
pöördumiskordade arvu ja
haigusjuhu kestuse prognoosid
Eesti Haigekassa ning tervise
infosüsteemi andmete põhjal**

Kirill Smirnov, Eva Anderson, Diana Sokurova

Tallinn 2019

Tervise Arengu Instituudi **missioon** on luua ja jagada teadmisi, et tõenduspõhiselt mõjutada tervist toetavaid hoiakuid, käitumist, poliitikat ja keskkonda eesmärgiga suurendada inimeste heaolu Eestis.

Väljaande andmete kasutamisel viidata allikale.

Soovitav viide käesolevale väljaandele: Smirnov K., Anderson E., Sokurova D. Perearstiabi patsientide pöördumiskordade arvu ja haigusjuhu kestuse prognoosid Eesti Haigekassa ning tervise infosüsteemi andmete põhjal. Tallinn: Tervise Arengu Instituut; 2019.

Sisukord

Lühendid.....	4
Mõisted.....	4
Lühikokkuvõte	6
Summary.....	8
Sissejuhatus.....	10
1 Metoodika	11
1.1 Loendusandmete mudelid	11
1.2 Logistilised mudelid	11
1.3 Elukestusmudelid.....	12
1.4 Mudelite diagnostika.....	12
2 Mudelite tegemine Eesti Haigekassa andmestiku põhjal.....	13
2.1 Eesti Haigekassa andmestiku kirjeldus	13
2.2 Saadud loendusandmete mudeli kirjeldus	14
2.3 Saadud logistilise mudeli kirjeldus.....	19
3 Mudelite testimine tervise infosüsteemi andmestikuga	22
3.1 Tervise infosüsteemi andmestiku kirjeldus.....	22
3.2 Loendusandmete mudeli testimine.....	23
3.3 Logistilise mudeli testimine.....	25
4 Haigusjuhu kestuse prognoosimine tervise infosüsteemi andmestikus.....	26
Kasutatud kirjandus.....	32
LISA.....	33

Lühendid

EHK

Eesti Haigekassa

RHK-10

rahvusvaheline haiguste ja tervisega seotud probleemide statistiline klassifikatsioon, 10. versioon

TIS

tervise infosüsteem

Mõisted

Ambulatoorne haigusjuht

on patsiendi uuringute ja raviga seotud toimingud tervishoiuasutuses ühe eriala piires ambulatoorses arstiabis. Ambulatoorne haigusjuht koosneb esmasest (sh koduviisi) ja korduvatest visiitidest ning raviga seotud muudest toimingutest (uuringud, protseduurid). Ambulatoorne haigusjuht algab patsiendi suunamise (kiirabi, arsti saatekiri) või pöördumisega (patsient tuleb ise) visiidile ja lõppeb arsti otsusega haigusjuhu lõpetamise kohta. Üldjuhul kestab ambulatoorne haigusjuht kuni 3 kuud (WHO) (1).

Epikriis

kokkuvõtte patsiendi haigusjuhtumist, milles kajastatakse antud haigusjuhtumi dünaamika, lähtudes arsti käsutuses olevast sellekohasest teabest. Ambulatoorne epikriis koostatakse iga ambulatoorse haigusjuhtumi kohta selle lõppemisel (1).

Haigusjuhu kestus

patsiendi esimese ja viimase pöördumise ajavahemik ühe ja sama põhidiagnoosiga ühe kalendriaasta jooksul.

Muu tervishoiualane kontakt (Z00–Z99)

hõlmab nii väikelaste kui täiskasvanute profülaktilisi läbivaatlusi, tervisetõendite väljastamist, sõel-uuringute tegemist, retseptide väljastamist ja jälgimist haiguse kahtluse või sotsiaalsete probleemide korral jne (2).

Põhidiagnoos

patsiendi peamine pöördumise põhjus RHK-10 järgi, mis määratletakse tervisehoolduse episoodi lõpul. Peamiseks märgitakse seisund, mis oli patisendi ravi- ja uurimisvajaduse peamiseks põhjuseks. Kui seisundeid on rohkem kui üks, valitakse nende hulgast see, mis põhjustas suurimat ressurside kasutust.

Pöördumiskordade arv

patsiendi pöördumiste ehk visiitide arv perearstiabisse.

Raviarve	ravikulude arve, mille tervishoiuteenuse osutaja esitab patsiendile osutatud teenuste eest EHK-le. Ühel raviarvel võib olla info mitme visiidi kohta.
Tõenäosus	arv lõigul $[0,1]$, mis iseloomustab juhusliku sündmuse võimalikkust. Kõigil juhtudel alluvad tõenäosused põhilistele reeglitele, näiteks kindla sündmuse tõenäosus on 1 ja võimatu sündmuse tõenäosus on 0. (3)
Visiit	ehk pöördumine perearstiabisse, mille hulka kuuluvad antud töö raames kõik perearsti või -õe vastuvõttud (esmane, korduv, profülaktiline) ning telefoni ja e-posti teel toimunud nõustamised.

Lühikokkuvõte

Analüüsis prognoositi statistiliste mudelite abil patsientide pöördumiskordade arvu perearstiabisse ja tõenäosust, et patsient pöördub ühe ja sama põhidiagnoosiga aasta jooksul vähemalt kaks korda ning samuti, kui pikk on patsiendi haigusjuhu kestus, mis on patsiendi esimese ja viimase pöördumise ajavahemik ühe ja sama põhidiagnoosiga ühe kalendriaasta jooksul.

Prognoosimudeli koostamiseks kasutatakse kahte mudeliklassi: loendusandmete ja logistilist mudelit. Mudelite koostamiseks kasutati 2017. aasta Eesti Haigekassa (EHK) ravikindlustuse andmekogu perearstiabi raviarvete andmeid ja tervise infosüsteemi (TIS) edastatud ambulatoorsete epikriiside andmeid.

Saadud tulemused selgitasid, millest sõltub pöördumiskordade sagedus ja milliste patsientidega tuleb perearstiabis kõige enam tegeleda. Antud informatsioon aitab hinnata ja paremini planeerida perearstiabi tööjaotust teades patsienti, tema põhidiagnoosi ja esimese pöördumise andmeid.

Nii EHK kui ka TIS andmestikus oli raviarveid ja epikriise, kus oli osa andmeid puudu (patsiendi sugu, vanus või põhidiagnoos). Tähele tuleb panna, et TIS andmestikus on puuduvate andmetega epikriiside osatähtsus 6 korda suurem kui EHK raviarvete andmestikus. 2017. aastal oli TIS andmestikus kirjeldatud ainult 720 362 unikaalset patsienti, EHK andmestikus oli patsientide arv 1 011 794. Seega annab EHK andmestik meile rohkem informatsiooni ja selle kvaliteet on parem.

Loendusandmete mudelit kasutatakse juhul, kui on vaja prognoosida sündmuse toimumise arvu teatud ajavahemikus. Antud analüüsis on see patsiendi pöördumiskordade arv perearstiabisse 2017. aastal. EHK andmete põhjal saadud mudelist selgus, et aasta jooksul tehtud pöördumiskordade arv sõltub patsiendi vanusest, põhidiagnoosi rühmast, põhidiagnoosi tüübist (esialgne, esmane, korduv), esimesest pöördumise viisist (perearsti/-õe vastuvõtt, nõustamine telefoni või e-posti teel) ning esimesest pöördumise kuupäevast. Pöördumiskordade arv ei sõltu patsiendi soost.

Mudeli järgi pöördub perearstiabisse kõige sagedamini patsient vanuses 0–4, kelle esmane põhidiagnoos on rühmast S00–T98 (vigastused, mürgistused ja teatavad muud välispõhjuste toime tagajärjed) ja tema esimene kontakt on pereõe nõustamine telefoni teel, mis toimub aasta alguses. Keskmine pöördumiskordade arv aastas antud tingimustel on kuus.

Kõige harvem pöördub perearstiabisse patsient vanuses 15–24, kelle esialgne ehk kinnitamata põhidiagnoos on rühmast H00–H59 (silma- ja silmamanuste haigused) ja tema esimene kontakt on perearsti vastuvõtt aasta lõpus. Keskmine pöördumiskordade arv aastas on 0,34.

Prognoositud pöördumiskordade arv erineb reaalsest pöördumiskordade arvust 62% juhtudest kuni ühe visiidi võrra EHK testandmestikus ja 73,5% juhtudest TIS andmestikus. Ülejäänud juhtudel on prognoosid veel suurema veaga, vastavalt 38% EHK ja 26,5% TIS andmestikust.

Mudeli prognoos on täpne juhul, kui patsiendi reaalne pöördumiskordade arv aastas on üks, kaks või kolm. Näiteks, kui patsient pöördus aasta jooksul kaks korda, siis mudeli prognoos on keskmiselt 1,9. Küll aga ei saa mudeliga hinnata neid patsiente, kelle pöördumiskordade arv on suurem. Näiteks, kui patsient pöördus aastas realselt kuus korda on tema keskmine prognoositud pöördumiskordade arv 2,3.

Logistilisi mudeleid kasutatakse juhul, kui prognoositud tunnus on binaarne ehk tunnusel on ainult kaks võimalikku väärtust (0 või 1). Antud töö raames prognoositi logistilise mudeli abil tõenäosust, kas patsient pöördus perearstiabisse aastas rohkem kui korra. EHK andmete põhjal saadud mudelist selgus, et tõenäosus sõltub soost, vanuserühmast, põhidiagnoosi rühmast ja selle tüübist (esialgne, esmane, korduv), esimesest pöördumise viisist (perearsti/-õe vastuvõtt, nõustamine telefoni või e-posti teel) ja esimesest pöördumise kuupäevast.

Saadud mudeli järgi on suurem tõenäosus pöörduda perearstiasse rohkem kui ühe korra aastas poisil vanuses 0–4, kelle esmane põhidiagnoos on rühmast S00–T98 (vigastused, mürgistused ja teatavad muud välispõhjuste toime tagajärjed) ja tema esimene kontakt on pereõe nõustamine telefoni teel aasta alguses. Tõenäosus, et ta tuleb ka teist korda, on 0,93.

Väikseim tõenäosus pöörduda perearstiasse rohkem kui ühe korra aastas on naisel vanuses 15–24, kellel on korduv põhidiagnoos rühmast H00–H59 (silma- ja silmamanuste haigused) ja tema esimene kontakt on perearsti vastuvõtt aasta lõpul. Tõenäosus, et ta pöördub perearstiasse vähemalt kaks korda, on sel juhul 0,13.

Saadud mudel prognoosib õigesti, kas patsient pöördub perearstiasse aasta jooksul üks kord või pöördub vähemalt kaks korda, EHK andmestikul 64% juhtudest ja TIS andmestikul 52% juhtudest.

TIS andmestikust tehtud mudeli järgi sõltub haigusjuhu kestus soost, vanuserühmast, põhidiagnoosi rühmast ja selle tüübist (esialgne, esmane, korduv) ning esimese visiidi tüübist (perearsti/-õe vastuvõtt, nõustamine telefoni teel).

Saadud mudeli järgi on haigusjuhu kestus kõige lühem mehel vanuses 45–64, kellel on korduv põhidiagnoos rühmast Z00–Z99 (tervise seisundit mõjustavad tegurid ja kontaktid terviseteenistusega) ja tema esimene kontakt oli perearsti nõustamine telefoni teel. Haigusjuhu prognoositud pikkus on 1,84 päeva.

Kõige pikema haigusjuhu kestusega on naine vanuses 65–84, kellel on esmane põhidiagnoos rühmast S00–T98 ja tema esimene kontakt perearstiasse on pereõe nõustamine telefoni teel. Antud juhul on haigusjuhu kestus 24,96 päeva.

TIS andmestikus on haigusjuhu kestuse mudeli prognoosi vea varieeruvus ± 5 päeva 54% juhtudest. Ülejäänud 46% juhtudest on prognoosi vea varieeruvus veel suurem.

Summary

Performed analysis uses statistical models for predicting the number of contacts to family doctors' office made by one patient during one year with same basic diagnosis and the probability that patient visits family doctors' office at least twice during the year with the same basic diagnosis. Also, there is statistical model predicting the length of patient's disease, which defined as the interval between the patient's first and last contact over the year with the same basic diagnosis in this paper.

For forecasting were considered two classes of statistical models: count model and logistic model. Family medical care billing data from the Estonian Health Insurance Fund (EHIF) 2017 was used for modeling.

Research showed which factors are statistically significant in case of the frequency of contacts to family doctors' office made by one patient for one year and which patients contact family doctor most often. These results will help family doctors' office to plan their amount of work knowing the information about patient, his or her basic diagnosis and information about first contact. In this research were used datasets from Estonian Health Insurance Fund (EHIF) and Health Information System (e-Health or TIS).

There were found missing data (patient sex, age, basic diagnosis) in both EHIF and TIS databases. Note that the proportion of missing data in TIS dataset was 6 times higher than in case of EHIF dataset. In 2017 there were described 720 362 unique patients in TIS dataset and 1 011 794 patients in EHIF dataset. So EHIF data gives more information, and its quality is higher.

Count data model is used when it is necessary to predict the number of events that will occur over a period. In this analysis, this is the number of patient contacts to family doctors' office in 2017. The model based on EHIF data showed that the number of contacts per year depends on the patient's age group, basic diagnosis group, type of basic diagnosis (initial, primary, recurrent), type of first contact (family doctor or nurse visit, telephone or email consultation) and date of the first contact, but does not depend on the patient's sex.

According to the model, a patient in the age group 0-4 with the primary diagnosis from the block S00-T98 (injury, poisoning and certain other consequences of external causes) whose first visit(recurrence) was nurse's counselling by telephone at the beginning of the year make the most contacts to family doctors' office in year. The average number of contacts per year under these conditions is six.

A patient in the age group 15-24 with an initial (unconfirmed) diagnosis from the block H00-H59 (diseases of the eye and adnexa) whose first contact was family doctor visit at the end of the year, tend to contact the doctor's office least often. The average number of contacts per year is 0.34.

Predicted number of contacts differs from the actual number of contacts in 62% of cases up by one visit in the EHIF test dataset and in 73.5% of cases in the TIS dataset. In the remaining cases, respectively 38% EHIF and 26.5% TIS, the prediction errors are higher.

The model prediction is accurate if the patients' actual number of contacts per year is one, two or three. For example, if a patient made two contacts per year, the model's prediction is 1.9 contacts on average. However, the model cannot fit well patients who have made many contacts. For example, if a patient made six contacts per year, his or her average estimated number of contacts is 2.3.

Logistic models are used when the predicted variable is binary, i.e. the variable has only two possible values (0 or 1). In this paper, logistic model was used to predict the probability, that the patient consults in a family doctors' office more than one time during the year. The model based on EHIF dataset showed that the probability of a patient to make at least two contacts per year depends on his or her sex, age group, basic diagnosis group, type of diagnosis (initial, primary, recurrent), type

of the first contact (family doctor or nurse visit, telephone or email counselling) and the date of first contact.

According to the model, the most likely visits family doctors' office more than once per year a boy aged 0-4 with primary diagnosis from block S00-T98 (injury, poisoning and certain other consequences of external causes) whose first contact was nurse counselling by telephone at the beginning of the year. The probability that he will come a second time is 0.93.

The smallest chance of contacting family doctors' office more than once a year is for woman in the age group 15-24 with a recurrent diagnosis from H00-H59 (diseases of the eye and adnexa), whose first contact type was family doctor visit at the end of the year. The probability that she will go to family doctors' office at least twice is 0.13.

The model correctly predicts whether the patient will consult in the family doctors' office only once or at least two times per year in 64% of cases in EHIF data and in 52% of cases in TIS data.

According to the model based on the TIS data, the duration of the disease depends on sex, age group, basic diagnosis group, type of basic diagnosis (initial, primary, recurrent) and type of first contact (family doctor or nurse visit, telephone consultation).

According to this model, the duration of the disease is the shortest for man aged 45-64, who has a recurrent diagnosis from the block Z00-Z99 (factors influencing health status and contacts with the health services) and his first contact was telephone consultation with a family doctor. The estimated duration of the case is 1.84 days.

The longest duration of the disease is for 65-84-years-old woman, who has a primary diagnosis from the block S00-T98, and her first contact is telephone consultation with nurse. In this case, the duration of the disease is 24.96 days.

According to the TIS data, the prediction error is ± 5 days in 54% of cases. In the remaining 46% of cases, the variability of the forecast error is even greater.

Sissejuhatus

Esmatasandi arstiabi ehk üldarstiabi ja laialdaselt kasutatud perearstiabi on tasand, kus inimesed saavad abi enamiku terviseprobleemidega. Perearstiabi on Eestis inimesele üldjuhul tasuta. Üldarstiabi teenuse osutamise eest kindlustatud isikule tasub Eesti Haigekassa (EHK), kellele tervishoiuteenuse osutajad (TTO) esitavad teostatud teenuste eest raviarveid. TTO esitab EHK-le lõpetatud raviarved jooksvalt (4). 2008. aastal loodud tervise infosüsteem (TIS) on riiklik andmekogu, milles töödeldakse tervishoiuvaldkonnaga seotud andmeid, kuhu edastavad andmeid kõik Terviseametis tegevusloa saanud TTO-d (5). TIS-i edastatakse andmeid lõpetatud epikriiside kohta. Kokkuleppeliselt edastatakse ambulatoorsed epikriisid haigusjuhu alguskuupäevast kolme kuu möödudes (1).

Riikliku ravikindlustusega on kaetud 1.24 miljonit elanikku, mis on 94% Eesti rahvastikust ja kelle andmed kajastuvad ka EHK andmekogus. (6). TIS-is on seevastu erinevate ravitüüpide andmete kaetus erinev. Statsionaarse ravi andmed on EHK andmetega võrreldavad, kuid ambulatoorse ravi andmeid on vähe. Üheks näitajaks on visiidid perearstiasse, mille andmekaaetus on TIS-is vaid 31% (7, 8).

TIS-i edastatavate andmete vähesuse tingib osalt riiklikult kokku leppimata definitsioonid kui teadmatus, mida edastada ja ka soov seda tegelikult teha.

Tervishoiutöötajate töö hindamiseks on võimalik kasutada erinevad lähenemisi. Eelmisel aastal ilmunud analüüsis „Perearstiasse pöördumise põhjused“ hinnatakse perearstide ja -õdede tööd peamiselt läbi kirjeldavate tunnuste (9). Esmatasandi arstiabi tööd on võimalik hinnata ka läbi matemaatiliste mudelite. Prognooside koostamine sobivate mudelite abil annab meile võimaluse tulevikus oma tegevusi paremini ja efektiivsemalt planeerida.

Käesoleva töö esimeseks eesmärgiks on mudelite koostamine prognoosimaks, mitu korda aastas pöördub patsient ühe ja sama terviseprobleemiga või muul eesmärgil (terviseetendi, retsepti pikendamine jne) perearstiasse ja milline on tõenäosus, et patsient pöördub sama probleemiga aasta jooksul korduvalt EHK andmete põhjal.

Töö teiseks eesmärgiks on haigusjuhu kestuse prognoosimine, mida saab teha TIS andmete põhjal elukestus mudeliga.

Analüüs on jätk Tervise Arengu Instituudi 2018. aasta ilmunud analüüsile „Perearstiasse pöördumise põhjused, 2017“ (9).

1 Metoodika

1.1 Loendusandmete mudelid

Loendusandmete mudelit kasutatakse juhul, kui on vaja prognoosida sündmuse toimumise arvu teatud ajavahemikus (10). Käesoleva töö raames on see patsiendi pöördumiskordade arv perearstiabisse 2017. aastal. Olemas on mitu erinevat loendusandmete mudelit. Kõige enam kasutatakse neist Poissoni ning Negative Binomial (NB-mudel) mudelit (10).

Kuna Poissoni jaotuse korral peab kehtima vastavus $D(Y) = E(Y)$, on Poissoni mudelit mõistlik kasutada tunnuse Y prognoosimiseks ainult siis, kui andmestikus puudub üle- või alahajuvus ($D(Y) > E(Y)$ ülehajuvuse korral ning $D(Y) < E(Y)$ alahajuvuse korral). Õeldakse, et alahajuvus ei tekita probleeme mudeli parameetrite hindamisel ja seda on vähem uuritud, aga suur ülehajuvus võib tekitada vigu. (11) Kui ülehajuvus on suur ($D(Y) - E(Y) > 5$), siis Poissoni mudeli asemel on NB-mudel eelistatum. (10)

Peale tavaliste Poissoni ning NB-mudelite eksisteerivad ka ZTP (*zero truncated Poisson*) ning ZTNB (*zero truncated NB*). Neid mudeleid kasutatakse juhul, kui andmestikus puuduvad nullid ehk vaadeldud ajavahemikus prognoositud sündmus toimub vähemalt üks kord.

Poissoni ning NB-mudeli korral arvutatakse prognoositud tunnus y järgmise valemi järgi:

$$y = \exp(x^T B),$$

kus x – seletavate tunnuste väärtuste vektor, B – mudeli koefitsientide vektor. (10)

1.2 Logistilised mudelid

Logistilisi mudeleid kasutatakse juhul, kui prognoositud tunnus on binaarne ehk tunnus on ainult kaks võimalikku väärtust (0 või 1). Näiteks, kas patsient pöördus perearstiabisse rohkem kui ühe korra või mitte, on binaarne.

Logistilise mudeli prognoosimiseks on võimalik kasutada *Logit*, *Probit* ning *Clog-Log* seosefunktsioone.

Olgu π tõenäosus, et prognoositud tunnus on võrdne ühega, siis

$$\text{Logit mudeli korral } \pi = \frac{\exp(x^T B)}{1 + \exp(x^T B)},$$

$$\text{Probit mudeli korral } \pi = \Phi(x^T B);$$

$$\text{Clog-log mudeli korral } \pi = 1 - \exp(-\exp(x^T B)),$$

kus x – seletavate tunnuste väärtuste vektor, B – mudeli koefitsientide vektor ning Φ – standardse normaaljaotuse tihedusfunktsioon.

Logistilise mudeli interpreteerimiseks on tihti kasutusel šanss. Šansiks nimetatakse sündmuse esinemise tõenäosuse ja sündmuse mitteesinemise tõenäosusesuhet $\Pi = \frac{\pi}{1-\pi}$. Logit mudeli korral $\Pi = \exp(x^T B)$.

Mudeli sobivuse hindamiseks on võimalik kasutada *Hosmer-Lemeshow*i testi. Kui saadud testi p -väärtus on $< 0,05$, siis saadud mudel ei sobi hästi andmete kirjeldamiseks. (10, 11)

1.3 Elukestusmudelid

Elukestusmudeleid kasutatakse juhul, kui on vaja prognoosida sündmuse toimumise aega. Näiteks, käesolevas töös prognoositakse aega päevades haigusjuhu algusest kuni hetkeni, mil patsient saab terveks või haigusjuht lõpetati.

Oluline mõiste elukestusmodelite kasutamisel on *hazard* funktsioon. *Hazard* funktsiooni saab interpreteerida kui tõenäosust, et haigusjuht lõpeb antud ajavahemikus $(t, t + \Delta t)$, kui on teada, et patsient pöördus ajal t ning $\Delta t \rightarrow 0$. (12)

Eksisteerib mitu erinevat elukestusmodelit. Kõige enam kasutatavad on Weibull, Log-normal ning Log-logistic mudelid. Nende mudelite kasutamisel eeldatakse, et uuritud tunnus on pärit vastavast jaotusest.

Alternatiivne variant on mitteparameetiline Cox mudel. Selle mudeli kasutamine ei eelda mingit konkreetset jaotust, kuid vajab erinevate vaatluste jaoks *hazard* funktsioonide proportsionaalsust. (12) Selleks, et seda kontrollida, kasutatakse *Grambsch-Therneau* testi. Kui testi p-väärtus on $< 0,05$, siis saadud Cox mudel ei sobi andmete kirjeldamiseks. (12)

1.4 Mudelite diagnostika

Mudelite võrdlemiseks kasutatakse *Akaike Information Coefficient* (AIC), mis arvutatakse järgmise valemi järgi:

$$AIC = 2k - 2\ln(L),$$

kus k – parameetrite arv mudelis, L – tõepärafunktsiooni maksimaalne väärtus. Tõepärafunktsiooniks nimetatakse funktsiooni, mis esitab valimi saamise tõenäosuse sõltuvalt jaotusparameetrite väärtusest.

Väiksema AIC-i väärtusega mudel kirjeldab andmeid paremini (12). Seega, kui võrrelda omavahel kaht mudelit AIC väärtustega vastavalt AIC_1 ja AIC_2 nii, et $AIC_1 > AIC_2$ siis teine mudel (väiksema AIC-i väärtusega) on eelistatum kasutamiseks ja prognoosimiseks.

Prognoosi täpsuse uurimiseks kasutakse töös RMSE (*Root mean square error*), MAPE (*Mean Absolute Percentage Error*) ning MAD (*Mean Absolute Difference*) kriteeriumeid.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}},$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|,$$

$$MAD = \frac{1}{n} \left| \sum_{i=1}^n \hat{y}_i - y_i \right|,$$

kus $\hat{y}_i$ – prognoositud väärtus, y_i – reaalne väärtus. (10)

2 Mudelite tegemine Eesti Haigekassa andmestiku põhjal

1.1 Eesti Haigekassa andmestiku kirjeldus

Eesti Haigekassa ravikindlustuse andmekogus on kõik nii ravikindlustatute raviarved kui ka ravikindlustamata patsientide üldarstiabi arved. Käesolev töö hõlmab kõiki raviarveid, mis algasid ning lõppesid 2017. aastal. Kokku on andmestikus 3 940 681 sobivat arvet. Igast raviarvest on teada:

- **Raviarve number**
Unikaalne raviarve kood.
- **Patsiendi ID**
Kaheksanumbiline kood, mis eristab patsienti. Andmestikus on olemas 1 011 794 unikaalset patsienti.
- **Sugu**
40% raviarvetest on kirjutatud meespatsientidele ja 60% naispatsientidele. 10-nel raviarvel oli patsiendi sugu teadmata.
- **Vanus**
Isikute vanus varieerub andmestikus 0-st kuni 107-aastani. Neljal raviarvel oli vanus teadmata.
- **Vanuserühm**
Käesoleva töö raames kasutatakse seitset vanuserühma (0–4; 5–14; 15–24; 25–44; 45–64; 65–84; 85+). Kõige rohkem raviarveid on välja antud vanuserühma 65–84 patsientidele (29,01%). Väiksemad rühmad on vanuses 15–24 (5,54%) ning 85+ (5,52%).
- **Põhidiagnoos**
Raviarvele märgitud patsiendi pöördumise põhjuse põhidiagnoos või seisund, mis on kodeeritud rahvusvahelise haiguste klassifikatsiooni RHK-10 järgi. Kokku on andmestikus 7962 unikaalset RHK-10 koodi. Kõige sagedamini kasutatav kood on Z76.0 ehk korduva retsepti väljakirjutamine.
- **Põhidiagnoosi rühm**
Patsiendi pöördumise põhjuse põhidiagnoosi või seisundi rühm. Kõik põhidiagnoosid on jagatud 20-ks rühmaks (LISA). Andmestikust puudus V00–Y98 (Haigestumise ja surma välispõhjused) rühm. Kõige suurem koodide rühm on Z00–Z99 (22,21%), mis on terviseseisundit mõjustavad tegurid ja kontaktid terviseteenistusega. Väikseim põhidiagnoosi rühm on P00–P96 ehk sünniperioodis tekkivad teatavad seisundid, mille osatähtsus on alla 0,02%.
- **Põhidiagnoosi tüüp**
Kõik põhidiagnoosid ja seisundid on jagatud kolmeks tüübiks. „Esialgne“ tüüp tähendab, et patsiendi diagnoos ei ole veel kliiniliselt kinnitatud. „Esmane“ tüüp tähendab, et patsiendil on antud diagnoos kinnitatud esmakordselt. „Korduv“ tüüp tähendab, et patsiendil on antud diagnoos kinnitatud juba varem. Esialgsed diagnoosid on väga harvad ja esinevad ainult 2,65%-l raviarvetel. Kõige sagedasemad on korduvad diagnoosid, nende osatähtsus on 69,60%.
- **Pöördumiskordade arv (visiitide arv)**
Patsiendi kõikide visiitide (nii perearsti kui -õe vastuvõtt, telefoni või e-posti nõustamine) arv ühe raviarve raames.

- **Pöördumise viis (visiidi tüüp)**

Perearstiabisse pöördumise viisi aluseks kasutati vastavalt EHK tervishoiuteenuste loetelu koodide (13):

- perearsti esmane vastuvõtt (9001);
- perearsti korduv vastuvõtt (9002);
- perearsti profülaktiline vastuvõtt (9003);
- perearsti kindlustamata isikute plaaniline vastuvõtt (9016);
- perearsti telefoni teel toimunud nõustamine (9018);
- perearsti elektronposti teel toimunud nõustamine (9019);
- pereõe vastuvõtt (9061);
- pereõe telefoni teel toimunud nõustamine (9064);
- pereõe elektronposti teel toimunud nõustamine (9065)

Analüüsis grupeeriti kõik perearsti vastuvõtukoodid (9001, 9002, 9003 ja 9016) kokku.

- **Alguskuupäev**

Raviarve alguse kuupäev.

- **Lõppkuupäev**

Raviarve lõpu kuupäev.

Analüüsi ei kaasatud 33 raviarvet, kuna neil puudus patsiendi sugu, vanus või põhidiagnoositüüp. Kokku on vaatluse all **3 940 648** raviarvet.

Ühel patsiendil võib aasta jooksul olla mitu raviarvet sama pöördumise põhjusega ehk põhidiagnoosiga. Kuna eesmärk on analüüsida patsiendi summaarset pöördumiskordade arvu aastas ühe ja sama põhidiagnoosiga, mitte raviarve jooksul, grupeeriti raviarved patsiendi ning põhidiagnoosi järgi (2 724 733 rida). Grupeeritud raviarvete andmestikus on järgmised tunnused:

- patsiendi ID;
- patsiendi sugu;
- patsiendi maksimaalne vanus sel aastal;
- vanuserühm;
- patsiendi põhidiagnoos;
- põhidiagnoosi rühm;
- päevade arv esimesest pöördumisest aasta lõpuni sama diagnoosiga;
- põhidiagnoosi tüüp esimese pöördumise ajal antud aastal;
- esimese pöördumise viis selle diagnoosiga antud aastal;
- patsiendi pöördumiste arv selle diagnoosiga aasta jooksul.

1.2 Saadud loendusandmete mudeli kirjeldus

Järgnevalt kirjeldatakse mudelit prognoosimaks patsiendi pöördumiskordade arvu perearstiabi kasutamisel ühe ja sama põhidiagnoosiga aastas.

Andmestiku analüüsimisel leiti mitu patsienti, kelle summaarne pöördumiskordade arv ei ole loogiline. Näiteks üks patsient pöördus perearstiabisse 906 korda 87 päeva jooksul. Suure pöördumiskordade arvuga patsientide osatähtsus on üsna väike (tabel 1). Seega, eelpool nimetatud patsient ja teised patsiendid, kellel oli rohkem kui seitse pöördumist, lisati „7+“ rühma. See parandab andmete kvaliteeti ja lihtsustab mudeli koostamist.

Tabel 1 Patsientide pöördumiskordade arvu jaotus

Pöördumiskordade arv	1	2	3	4	5	6	7+
Patsientide arv	1 203 746	721 523	335 517	192 976	95 295	62 247	113 429
Osatähtsus	44,19%	26,48%	12,31%	7,08%	3,50%	2,28%	4,16%

Pöördumiskordade arvu prognoosimiseks kasutati järgmisi tunnuseid:

- patsiendi sugu;
- patsiendi maksimaalne vanus sel aastal või vanuserühm;
- patsiendi põhidiagnoos (pöördumise põhjus);
- patsiendi põhidiagnoosi rühm;
- päevade vahe esimesest pöördumisest aasta lõpuni sama põhidiagnoosiga;
- põhidiagnoosi tüüp (esialgne, esmane või korduv) esimese pöördumise ajal antud aastal;
- esimese pöördumise viis (perearsti või -õe vastuvõtt, telefoni- või e-posti nõustamine) selle diagnoosiga sel aastal.

Mudeli tegemiseks on EHK andmestik jagatud kaheks osaks. Esimene osa on treeningandmed ehk need, mida on kasutatud mudeli õppimiseks ja teine osa on testandmed, mis on vajalikud prognoosimiseks ning prognoosi täpsuse leidmiseks. Treeningvaatluste valimiseks on andmestik jagatud seitsmeks kihiks pöördumiskordade arvu järgi. Igast kihist valitakse juhuslikult 70% treeningandmete hulka, ülejäänud 30% on testandmed.

Analüüsi protsessis on vaatluse all erinevad Poissoni mudelid, sealhulgas ZTP mudelid. NB-mudelit pole vaadeldud, kuna andmestikus puudub ülehajuvus. Soo tunnus ei ole oluline iga mudeli korral. Seega töötati edasi ainult järgmiste tunnustega: vanus või vanuserühm, põhidiagnoosi rühm, päevade arv aasta lõpuni, põhidiagnoosi tüüp ning esimese pöördumise viis.

Väikseima AIC-i väärtusega mudel kirjeldab andmeid paremini (12). Tabelis 2 on näha, et AIC-i väärtus on väiksem, kui mudelis kasutatakse tunnuse vanus asemel vanuserühm. Kõige väiksem AIC on ZTP mudelil parameetritega vanuserühm, põhidiagnoosi rühm, põhidiagnoosi tüüp, pöördumise viis ning päevade arv esimesest pöördumisest aasta lõpuni. ZTP mudel (joonis 1) on eelistatum ka nullide puudumise tõttu (andmestikus on patsiente, kes on vähemalt korra pöördunud perearstiabisse).

Tabel 2. Mudelite võrdlus

	AIC
Poisson vanuserühmaga	7 136 603
Poisson vanusega	7 158 217
ZTP vanuserühmaga	6 407 448
ZTP vanusega	6 441 551

Pöördumiskordade arv

$$\begin{aligned}
 &= \exp(-0,07 - 0,35 \times I(\text{vanus} = 5 - 14) - 0,43 \times I(\text{vanus} = 15 - 24) \\
 &- 0,26 \times I(\text{vanus} = 25 - 44) - 0,14 \times I(\text{vanus} = 45 - 64) \\
 &- 0,14 \times I(\text{vanus} = 65 - 84) - 0,15 \times I(\text{vanus} = 85 +)) \\
 &+ 0,05 \times I(\text{diagngr} = C00 - D48) + 0,06 \times I(\text{diagngr} = D50 - D89) \\
 &- 0,06 \times I(\text{diagngr} = E00 - E90) + 0,04 \times I(\text{diagngr} = F00 - F99) \\
 &+ 0,02 \times I(\text{diagngr} = G00 - G99) - 0,56 \times I(\text{diagngr} = H00 - H59) \\
 &- 0,27 \times I(\text{diagngr} = H60 - H95) + 0,14 \times I(\text{diagngr} = I00 - I99) \\
 &+ 0,03 \times I(\text{diagngr} = J00 - J99) - 0,18 \times I(\text{diagngr} = K00 - K93) \\
 &- 0,38 \times I(\text{diagngr} = L00 - L99) - 0,02 \times I(\text{diagngr} = M00 - M99) \\
 &- 0,1 \times I(\text{diagngr} = N00 - N99) + 0,007 \times I(\text{diagngr} = O00 - O99) \\
 &- 0,52 \times I(\text{diagngr} = P00 - P96) - 0,46 \times I(\text{diagngr} = Q00 - Q99) \\
 &- 0,25 \times I(\text{diagngr} = R00 - R99) + 0,23 \times I(\text{diagngr} = S00 - T98) \\
 &- 0,39 \times I(\text{diagngr} = Z00 - Z99) + 0,27 \times (\text{tüüp} = \text{esmane}) \\
 &+ 0,17 \times (\text{tüüp} = \text{korduv}) + 0,0025 \times \text{päevade arv} \\
 &+ 0,02 \times (\text{vis tüüp} = 9018) + 0,11 \times (\text{vis tüüp} = 9019) \\
 &+ 0,2 \times (\text{vis tüüp} = 9061) + 0,29 \times (\text{vis tüüp} = 9064) \\
 &+ 0,36 \times (\text{vis tüüp} = 9065))
 \end{aligned}$$

Märkus: Mudelis tunnus „vis tüüp“ tähistab pöördumise viisi (visiidi tüüp).

Joonis 1. Zero truncated Poisson mudel pöördumiskordade arvu prognoosimiseks.

Vaatamata sellele, et lõppmudel (vt. joonis 1) on parameetreid palju, on tulemuste interpreteerimine üsna lihtne. Parameetrid positiivse koefitsiendiga suurendavad prognoositud pöördumiste arvu ja negatiivsed vähendavad seda. Mudelist on näha, et koefitsiendid on kõikide vanuserühmade ees negatiivsed. See tähendab, et vanuserühmas 0–4 pöördutakse perearstiasse sagedamini kui vanemates vanuserühmades. Näiteks, kui on kaks isikut erineva põhidiagnoosiga, kuid samast diagnoosirühmast, sama põhidiagnoosi tüübi ja esimese pöördumise viisiga ning pöördumiskuupäev on sama, kuid üks neist on vanuserühmas 0–4 ja teine 5–14, siis vanem patsient pöördub perearstiasse aasta jooksul keskmiselt $\frac{1}{\exp(-0,35)} = 1,43$ korda harvemini, kui patsient vanuserühmas 0–4.

Erinevaid põhidiagnoosi rühmi võrreldi põhidiagnoosi rühmaga A00–B99 (teatavad nakkus- ja parasiithaigused). See tähendab, et kui on kaks isikut samast vanuserühmast, sama pöördumise põhjuse (põhidiagnoosiga) ja esimese pöördumise viisiga ning pöördumiskuupäev on sama, kuid ühel neist on pöördumise põhjuseks märgitud põhidiagnoos rühmast A00–B99 ja teisel H00–H59 (silma- ja silmamanuste haigused), siis teine patsient pöördub samal aastal perearstiasse keskmiselt $\frac{1}{\exp(-0,56)} = 1,75$ korda harvemini, kui esimene. Aga, kui teisel patsiendil on pöördumise põhjuseks märgitud põhidiagnoos rühmast I00–I99 (vereringeelundite haigused), siis pöördub ta $\exp(0,14) = 1,15$ korda sagedamini kui esimene patsient.

Võrreldes kaht samade parameetritega patsienti, kellest ühel on esimesel pöördumisel esialgne põhidiagnoos ehk diagnoos ei ole veel kinnitatud, siis ta pöördub aasta jooksul perearstiasse keskmiselt $\exp(0,27) = 1,31$ korda vähem kui patsient esimese põhidiagnoosiga ja $\exp(0,17) = 1,19$ korda vähem kui korduva põhidiagnoosi patsient.

Pöördumiskordade arv sõltub kuupäevast, millal patsient esimest korda perearstiasse pöördus. Kui kaks sarnast patsienti pöörduvad perearstiasse, siis patsient, kes pöördus päeva võrra varem, pöördub sel aastal keskmiselt $\exp(0,0025) = 1,0025$ korda sagedamini kui hilisem patsient.

Samuti sõltub pöördumiskordade arv esimesest pöördumise viisist (vis tüüp) ehk kellega ja mil viisil toimub patisendi esimene kontakt perearstiasis. Näiteks patsient, kelle esimene kontakt on pereõe

vastuvõtt (kood 9061), pöördub $\exp(0,2) = 1,22$ korda sagedamini kui sarnane patsient, kelle esmaseks kontaktiks on perearsti vastuvõtt (koodid: 9001, 9002, 9003, 9016).

Saadud mudeli järgi pöördub kõige enam perearstiabisse patsient vanuses 0–4, kelle pöördumise põhjuse esmane põhidiagnoos on rühmast S00–T98 (vigastused, mürgistused ja teatavad muud välispõhjuste toime tagajärjed) ja tema esimeseks kontaktiks on pereõe nõustamine telefoni teel ning see toimus aasta alguses. Nende patsientide keskmine pöördumiste arv aastas on 5,9.

Kõige harvem pöördub perearstiabisse patsient vanuses 15–24, kelle esialgne pöördumise põhjus on põhidiagnoos rühmast H00–H59 (silma- ja silmamanuste haigused) ja tema esimene kontakt on perearsti vastuvõtt, mis toimus aasta lõpul. Sel juhul on nende patsientide keskmine pöördumiskordade arv aastas 0,34.

Saadud mudeli (ZTP vanuserühmaga) RMSE on 1,56; MAPE on 51,11%; MAD on 1,08 (tabel 3). Seega ei saa öelda, et mudel on väga täpne, kuid see on parim mudel vaadeldud mudelite hulgas MAPE ja MAD järgi (RMSE põhjal on Poisson vanuserühmaga parem, aga erinevus pole suur).

Tabel 3. Mudelite võrdlus

	RMSE	MAPE	MAD
Poisson vanuserühmaga	1,52	63,76%	1,14
Poisson vanusega	1,52	64,24%	1,15
ZTP vanuserühmaga	1,56	51,11%	1,08
ZTP vanusega	1,57	51,27%	1,09

Tabelis 4 on näha, et 36,57% andmetest on prognoositud täpsusega 0,5 ning 62,42% juhtudest prognoosi viga on ühest väiksem.

Vaatleme eraldi prognoosi täpsust patsientidele erinevate pöördumiskordade arvuga.

Tabel 4. Prognoosi täpsus testandmetes

Prognoosi erinevus reaalsest väärtusest (pöördumiskordade arv)	<0,5	0,5–1	1–2	2–3	>3
Osatähtsus	36,57%	25,85%	23,00%	6,71%	7,87%

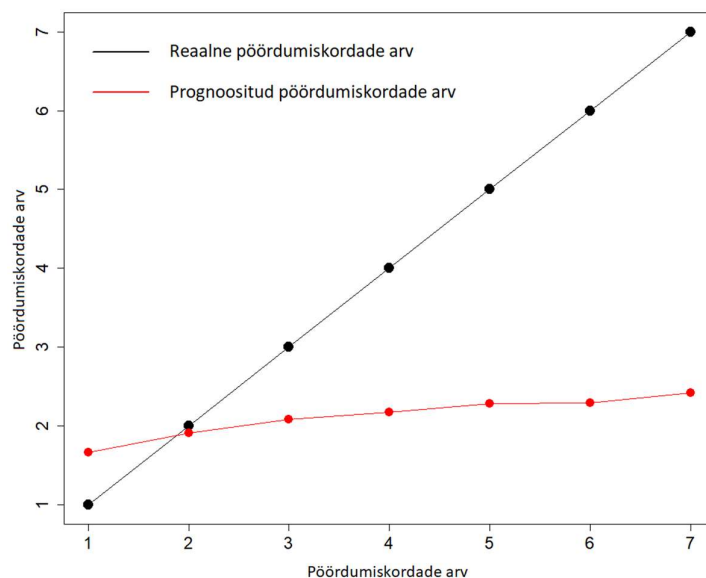
Tabelis 5 on näha, et nii ühe, kahe kui kolme reaalse pöördumise korral töötab mudel üsna täpselt ja 96–99% juhtudest on prognoosi viga väiksem kui kaks. Kuid suurema reaalsete pöördumiskordade arvu korral prognoosib mudel ebatäpselt. Näiteks kuue pöördumise puhul on prognoosi viga kahest suurem 99,41% juhtudel.

Tabel 5. Prognoosi täpsus testandmetes reaalsete pöördumiskordade järgi

Prognoosi erinevus reaalsest väärtusest	<0,5	0,5–1	1–2	2–3	>3
Ühe pöördumisega patsientide hulk	46,46%	26,51%	24,37%	2,57%	0,09%
Kahe pöördumisega patsientide hulk	49,55%	39,47%	10,76%	0,22%	0,00%
Kolme pöördumisega patsientide hulk	25,70%	26,44%	45,39%	2,46%	0,00%
Nelja pöördumisega patsientide hulk	2,83%	8,56%	46,60%	40,11%	1,90%
Viie pöördumisega patsientide hulk	0,02%	0,78%	13,40%	50,91%	34,90%
Kuue pöördumisega patsientide hulk	0,00%	0,00%	0,59%	13,68%	85,73%
Seitse ja enam pöördumisega patsientide hulk	0,00%	0,00%	0,00%	1,30%	98,70%

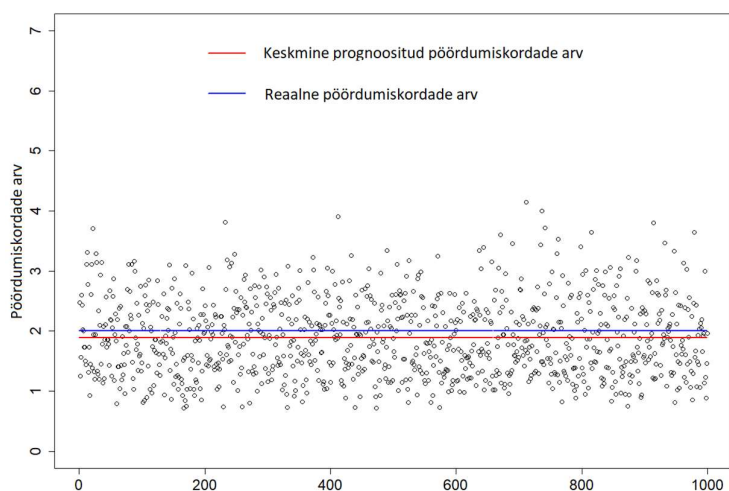
Sama on kujutatud ka joonisel 2, kus punane joon näitab, kuidas muutub keskmine prognoositud pöördumiskordade arv reaalsete pöördumiskordade arvu (must sirge) suurenemisel. Samuti on

näha, et keskmine prognoositud pöördumiskordade arv kasvab väga aeglaselt, mistõttu suure pöördumiskordade arvu korral on prognoos ebatäpne.



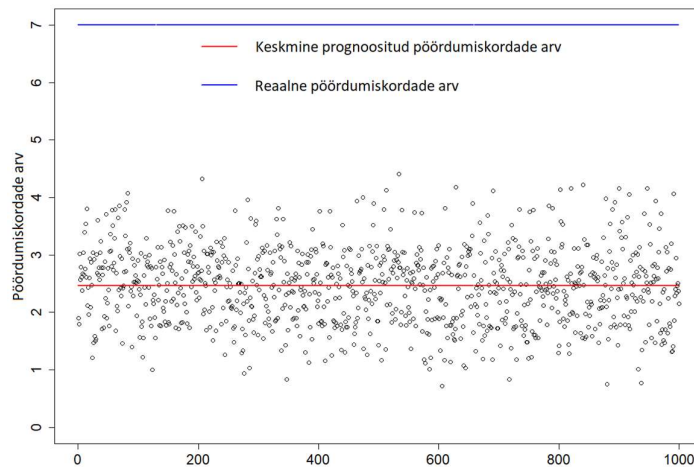
Joonis 2. Keskmine prognoositud ning reaalne pöördumiste arv perearstiasse EHK andmete põhjal

See ebatäpsus on ülevaatlilikult visualiseeritud joonisel 3 ja joonisel 4. Kui reaalne pöördumiskordade arv testandmestikus on võrdne kahega (joonis 3), siis keskmine prognoositud pöördumiskordade arv on umbes 1,9 ja suur osa prognoosidest (mustad punktid) on reaalse väärtuse lähedal.



Joonis 3. Prognoositud pöördumiskordade arv, kui reaalne pöördumiste arv on võrdne kahega

Kui reaalne pöördumiskordade arv on seitse või rohkem (joonis 4), siis keskmine prognoositud pöördumiskordade arv on ainult 2,6 ning minimaalne prognoosi viga on kahest suurem.



Joonis 4. Prognoositud pöördumiskordade arv, kui reaalne pöördumiste arv on võrdne seitsmega

1.3 Saadud logistilise mudeli kirjeldus

Antud peatükis kirjeldatakse mudelit tõenäosuse prognoosimiseks, et patsient pöörduv perearstiabisse sama põhidiagnoosiga vähemalt kaks korda aastas. Selleks loodi uus binaarne tunnus, mis on võrdne nulliga, kui pöörduti vaid korra ja ühega, kui pöördumiskordi oli rohkem.

Mudeli õppimiseks ning prognoosi täpsuse kontrollimiseks kasutatakse sama treening- ja testandmestikku nagu eelmises punktis. Treeningandmestikus on 44,45% patsientidest, kes pöördusid perearstiabisse ühe ja sama diagnoosiga ainult korra aastas ning 55,55% pöördusid vähemalt kaks korda (tabel 6).

Tabel 6. Pöördumiskordade arv treeningandmestikus

Pöördumiskordade arv	Üks	Kaks või rohkem
Patsientide arv	946 167	1 182 335
Osatähtsus	44,45%	55,55%

Binaarse tunnuse prognoosimiseks on võimalik kasutada *Logit*, *Probit* ning *Clog-Log* seosefunktsiooni. Need kolm mudelit testiti parameetritega sugu, vanuserühm, põhidiagnoosi rühm ja -tüüp, esimese pöördumise viis ning päevade vahe aasta lõpuni. *p*-väärtused näitavad, et kõik parameetrid on olulised kõigi mudelite korral. AIC väärtuse põhjal (tabel 7) parimaks mudeliks on mudel *Logit* seosefunktsiooniga (joonis 5). Kuid *Hosmer-Lemeshow*i testi põhjal võib öelda, et mitte ükski mudel ei kirjelda andmeid hästi (*p*-väärtus <0,05).

Tabel 7. Seosefunktsioonide võrdlemine

Seosefunktsioon	AIC	<i>Hosmer-Lemeshow</i> i testi <i>p</i> -väärtus
Logit	2 711 414	$< 2 \times 10^{-16}$
Probit	2 711 513	$< 2 \times 10^{-16}$
Clog-log	2 715 391	$< 2 \times 10^{-16}$

$$\begin{aligned}
\pi = \exp & (-0,42 - 0,03 \times I(\text{sugu} = \text{naine}) - 0,48 \times I(\text{vanus} = 5 - 14) \\
& - 0,52 \times I(\text{vanus} = 15 - 24) - 0,3 \times I(\text{vanus} = 25 - 44) \\
& - 0,13 \times I(\text{vanus} = 45 - 64) - 0,16 \times I(\text{vanus} = 65 - 84) \\
& - 0,24 \times I(\text{vanus} = 85 +) - 0,23 \times I(\text{diagngr} = C00 - D48) \\
& + 0,25 \times I(\text{diagngr} = D50 - D89) - 0,04 \times I(\text{diagngr} = E00 - E90) \\
& - 0,2 \times I(\text{diagngr} = F00 - F99) - 0,18 \times I(\text{diagngr} = G00 - G99) \\
& - 0,96 \times I(\text{diagngr} = H00 - H59) - 0,51 \times I(\text{diagngr} = H60 - H95) \\
& + 0,22 \times I(\text{diagngr} = I00 - I99) + 0,07 \times I(\text{diagngr} = J00 - J99) \\
& - 0,31 \times I(\text{diagngr} = K00 - K93) - 0,77 \times I(\text{diagngr} = L00 - L99) \\
& - 0,14 \times I(\text{diagngr} = M00 - M99) - 0,2 \times I(\text{diagngr} = N00 - N99) \\
& + 0,007 \times I(\text{diagngr} = O00 - O99) - 0,75 \times I(\text{diagngr} = P00 - P96) \\
& - 0,78 \times I(\text{diagngr} = Q00 - Q99) - 0,37 \times I(\text{diagngr} = R00 - R99) \\
& + 0,26 \times I(\text{diagngr} = S00 - T98) - 0,92 \times I(\text{diagngr} = Z00 - Z99) \\
& + 0,47 \times (\text{tüüp} = \text{esmane}) - 0,11 \times (\text{tüüp} = \text{korduv}) \\
& + 0,0037 \times \text{päevade arv} + 0,05 \times (\text{vis tüüp} = 9018) \\
& + 0,26 \times (\text{vis tüüp} = 9019) + 0,47 \times (\text{vis tüüp} = 9061) \\
& + 0,79 \times (\text{vis tüüp} = 9064) + 0,97 \times (\text{vis tüüp} = 9065))
\end{aligned}$$

Märkus: Mudelis tunnus „vis tüüp“ tähistab pöördumise viisi (visiidi tüüp).

Joonis 5. Mudel šanssi prognoosimiseks *Logit* seosefunktsiooniga

Positiivsed koefitsiendid mudelis (joonis 5) suurendavad šanssi, et patsient pöördub perearstiabisse vähemalt kaks korda aastas sama põhidiagnoosiga ja negatiivsed koefitsiendid vähendavad seda.

Näiteks, kui perearstiabisse pöördus samal kuupäeval mees ja naine ühest vanuserühmast, kellel on sama pöördumise põhjus ehk põhidiagnoosi rühm ja tüüp ning samasugune esimene pöördumise viis, siis šanss, et mees pöördub rohkem kui ühe korra on $\frac{1}{\exp(-0,03)} = 1,03$ korda suurem kui naise puhul.

Patsientidel vanuserühmas 0–4 on suurem šanss pöörduda perearstiabisse enam kui ühe korra, kui teistes vanuserühmades. Näiteks, kui perearstiabisse pöörduvad kaks sarnast patsienti, kellest üks on vanuserühmas 0–4 ja teine 15–24, siis šanss, et väikelapsega pöördutakse veel üks kord on $\frac{1}{\exp(-0,52)} = 1,68$ suurem, kui patsiendil vanuserühmas 15–24.

Patsient, kelle pöördumise põhjuseks olev haigus või muu tervishoiualane kontakt on esmakordne (põhidiagnoosi tüüp on esmane), tulevad perearsti juurde $\exp(0,47) = 1,6$ korda suurema šansiga, kui sama patsient, kelle põhidiagnoos või haiguskasutus jääb kinnitamata (põhidiagnoosi tüüp on esialgne).

Teist korda perearstiabisse pöördumise šanss sõltub ka esimesest pöördumise kuupäevast. Kahe sarnase patsiendi korral ühe päeva varem pöördunud patsiendi šanss pöörduda sel aastal veel korra on $\exp(0,0037) = 1,0037$ korda suurem.

Samuti sõltub pöördumise šanss põhidiagnoosi rühmast ja esimesest pöördumise viisist. Näiteks, kui ühel patsiendil on pöördumise põhjuseks põhidiagnoos rühmast H00–H59 (silma- ja silmamanuste haigused) ja teisel on A00–B99 (teatavad nakkus- ja parasiithaigused) ning ülejäänud näitajad on sarnased, siis teise patsiendi šanss veel korra pöörduda on $\frac{1}{\exp(-0,96)} = 2,61$ korda suurem kui esimesel.

Teades šanssi, on võimalik leida tõenäosust, et patsient pöördub aastas perearstiabisse sama põhidiagnoosiga ka teist korda.

Saadud mudeli järgi on suurim tõenäosus, et patsient pöördub perearstiabisse rohkem kui üks kord aastas, poistel vanuserühmas 0–4, kelle esmane põhidiagnoos on rühmast S00–T98 (vigastused, mürgistused ja teatavad muud välispõhjuste toime tagajärjed) ja tema esimene kontakt on pereõe nõustamine telefoni teel aasta alguses ($\pi = 0,93$). Väikseim tõenäosus, et patsient pöördub rohkem kui üks kord aastas, on naistel vanuses 15–24, kelle pöördumise põhjuseks on korduv diagnoos rühmast H00–H59 (silma- ja silmamanuste haigused) ja tema esmane pöördumine perearsti vastuvõtule toimus aasta lõpul ($\pi = 0,13$).

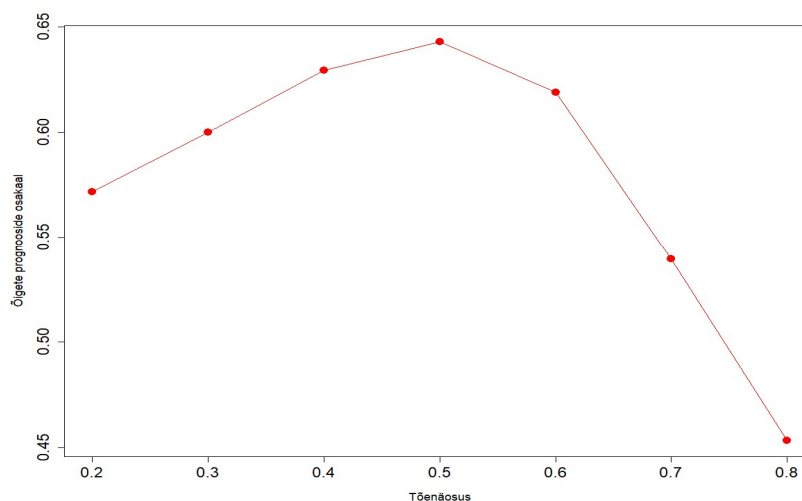
Kasutades testandmestikku, kontrollime, kui hästi töötab valitud mudel. Eeldame, et kui teise pöördumise tõenäosus on $\geq 0,5$, siis järeldame, et patsient pöördub perearstiabisse sama põhidiagnoosiga rohkem kui üks kord aastas.

Tabelist 8 on näha, et mudel annab õige tulemuse 64,3% juhtudest ehk kaks kolmandikku patsientidest on õigesti klassifitseeritud.

Tabel 8. Prognoosi täpsus

Prognoos	Reaalne pöördumiskordade arv	
	Üks	Vähemalt kaks
Üks	20,37%	12,96%
Vähemalt kaks	22,74%	43,93%

Proovides teisi valitud tõenäosusi, muutub mudel ebatäpsemaks. Seda on näha joonisel 6. Näiteks, kui valida tõenäosuseks 0,8, klassifitseerib mudel patsiente õigesti ainult 45% juhtudest.



Joonis 6. Mudeli täpsus erinevate valitud tõenäosuste korral

3 Mudelite testimine tervise infosüsteemi andmestikuga

Eelmises peatükis kirjeldatud mudelid koostati Eesti Haigekassa andmestiku põhjal. Järgnevas peatükis vaatleme, kui hästi töötavad saadud mudelid (ZTP ning Logit-mudel) tervise infosüsteemi andmetega.

1.4 Tervise infosüsteemi andmestiku kirjeldus

Tervise infosüsteemi (TIS) andmestik on oma struktuurilt sarnane EHK andmestikuga. TIS-i koondatakse sarnane informatsioon perearstiabis tehtud pöördumistest üle Eesti. Töö hõlmab kõiki perearstiabide poolt edastatud epikriise, mis algasid ning lõppesid 2017. aastal. Kokku on andmestikus 1 669 820 sobivat epikriisi. Igast epikriisist on teada:

- **Dokumendi number**
Unikaalne kood epikriisil.
- **Patsiendi ID**
Numbriline kood, mis identifitseerib patsiendi. Kodeerimine erineb EHK andmestikust. TIS andmestikus on olemas 720 362 unikaalset patsienti, mida on 1,5 korda vähem, kui EHK andmestikus.
- **Sugu**
42% epikriisidest on kirjutatud meespatsientidele ning 58% naispatsientidele. 12-nel raviarvel oli patsiendi sugu teadmata.
- **Vanus**
Patsiendi vanus epikriisi alustamise hetkel. Isikute vanus varieerub andmestikus 0-st kuni 107-aastani nagu EHK andmestikus. Kümnel epikriisil on vanus teadmata.
- **Vanuserühm**
Patsiendi vanuserühm epikriisi alustamise hetkel. Vanuserühmad moodustati sarnaselt EHK andmestikuga (0–4; 5–14; ..., 85+).
- **Eriala kood**
Epikriisi koostaja eriala kood; peremeditsiin (E300) ja pereõendus (N320)
- **Põhidiagnoos**
Ambulatoorsele epikriisile märgitud patsiendi pöördumise põhjuse põhidiagnoos või seisund, mis on kodeeritud rahvusvahelise haiguste klassifikatsiooni RHK-10 järgi nagu EHK andmestikus. Kokku on andmestikus 7467 unikaalset RHK-10 koodi. 68-l epikriisil on kood teadmata
- **Põhidiagnoosi tüüp**
Patsiendi põhidiagnoosi tüüp. Kõik põhidiagnoosid on jagatud sarnaselt EHK andmesikuga kolmeks rühmaks (esialgne, esmane, korduv)
- **Visiitide arv (pöördumiskordade arv)**
Patsiendi kõikide visiitide/pöördumiskordade arv ühe epikriisi raames.
- **Visiidi tüüp (pöördumise viis)**
TIS-is jagunevad visiidid järgmiselt:
 - 1 – vastuvõtt,
 - 2 – koduviit,
 - 3 – nõustamine telefoni teel,
 - 4 – nõustamine elektronposti teel.Koduvisiite ja nõustamine elektronposti teel antud analüüsis ei vaadelda.

- **Algus- ja lõpukuupäev**
Epikriisi alguse ja lõpu kuupäev.

TIS andmestikus puudub tunnus „põhidiagnoosi rühm“. Kuna see on vajalik saadud mudelite kasutamiseks, moodustati selleks eraldi tunnus.

Analüüsi ei kaasatud 90 epikriisi, kuna nendes puudus patsiendi sugu, vanus või põhidiagnoos. Seega vaadeldi **1 669 730** epikriisi. Vigaste epikriiside osakaal TIS andmestikus on 6,42 korda suurem kui EHK andmestikus.

Eelmises peatükis kirjeldatud mudelite kasutamiseks moodustakse uus andmestik grupeeritud epikriisidega patsiendi ning põhidiagnoosi (pöördumise põhjuse) järgi (1 384 424 rida), nagu tehti EHK andmestikuga. Uues andmestikus on järgmised tunnused:

- patsiendi ID;
- patsiendi sugu;
- patsiendi maksimaalne vanus antud aastal;
- vanuserühm;
- patsiendi põhidiagnoos
- põhidiagnoosi rühm;
- päevade vahe esimesest visiidist sama põhidiagnoosiga aasta lõpuni;
- põhidiagnoosi tüüp esimese visiidi ajal antud aastal;
- esimese visiidi tüüp sama põhidiagnoosiga antud aastal;
- patsiendi visiitide arv sama põhidiagnoosiga aasta jooksul.

1.5 Loendusandmete mudeli testimine

Tervise infosüsteemi andmestikus erineb patsientide pöördumiskordade (visiitide) arvu jaotus EHK andmete põhjal saadud jaotusest. TIS-is on ühekordsete pöördumiste osatähtsus kaks korda suurem (EHK – 44,19%, TIS – 84,78%). Seetõttu suurema pöördumiskordade arvuga patsientide osatähtsus on palju väiksem kui EHK andmetes (tabel 1 ja tabel 9).

Tabel 9. Patsientide pöördumiskordade arvu jaotus

Pöördumiskordade arv	1	2	3	4	5	6	7+
Patsientide arv	1 153 340	137 459	37 333	14 847	7 287	3 928	6 242
Osatähtsus	84,78%	10,10%	2,74%	1,09%	0,54%	0,29%	0,46%

Peatükis 2.2 kirjeldatud ZTP mudeli RMSE ja MAD väärtused TIS andmetel on järgmised 0,97 ja 0,71. See on 1,5 korda väiksem kui EHK testandmestikus. Seega võib öelda, et mudel annab täpsema prognoosi TIS andmestikus.

Tabelis 10 on näha, et 47,55% andmetest on prognoositud täpsusega 0,5 ning 73,48% juhtudest on prognoosi viga ühest väiksem. See on 10% võrra täpsem, kui oli EHK testandmestikus. Nagu peatükis 2.2 sai kirjeldatud, on saadud ZTP mudeli prognoosid väga ebatäpsed, kui pöördumiskordade arv on suurem kui neli. Kuna TIS andmestikus on vähem neid patsiente, kelle pöördumiskordade arv ületab nelja, näitab mudel paremat tulemust.

Tabel 10. Prognoosi täpsus TIS andmetes

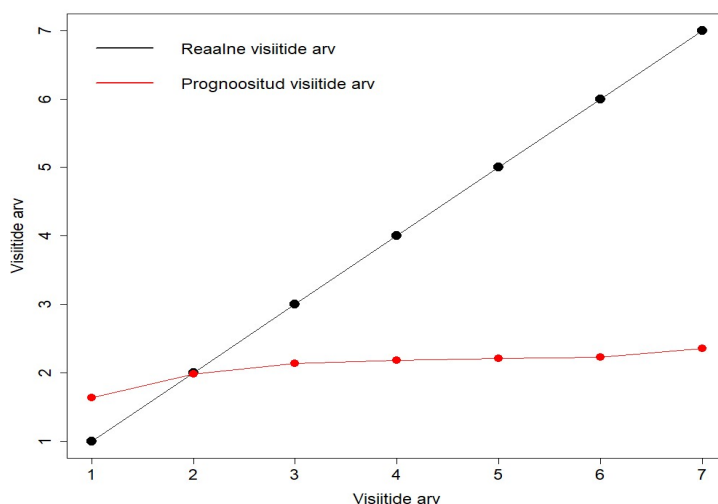
Prognoosi erinevus reaalsest väärtusest	<0,5	0,5–1	1–2	2–3	>3
Osatähtsus	47,55%	25,93%	22,75%	2,82%	0,94%

Vaatleme eraldi prognoosi täpsust patsientide erinevate pöördumiskordade arvuga. Kui võrrelda omavahel tabeli 5 ja tabeli 11 andmeid, siis on näha, et prognoosi täpsuses ei ole suurt erinevust. Mudel töötab hästi, kui reaalsete pöördumiste arv on üks, kaks või kolm. Sel juhul on prognoosi viga väiksem kui ± 2 visiiti (~98% juhtudest). Samaselt EHK testandmestiku prognoosiga on suure pöördumiskordade arvu korral prognoosi täpsus madal.

Tabel 11. Prognoosi täpsus TIS andmetes reaalsete pöördumiskordade arvu järgi

Prognoosi erinevus reaalsest väärtusest	<0,5	0,5–1	1–2	2–3	>3
Ühe pöördumisega patsientide hulk	48,75%	25,10%	23,72%	2,42%	0,01%
Kahe pöördumisega patsientide hulk	53,64%	37,27%	9,04%	0,05%	0,00%
Kolme pöördumisega patsientide hulk	28,58%	28,59%	41,37%	1,46%	0,00%
Nelja pöördumisega patsientide hulk	1,25%	9,52%	49,01%	39,29%	0,94%
Viie pöördumisega patsientide hulk	0,05%	0,38%	11,35%	48,40%	39,81%
Kuue pöördumisega patsientide hulk	0,00%	0,00%	0,46%	12,81%	86,74%
Seitse ja enam pöördumisega patsientide hulk	0,00%	0,00%	0,00%	0,70%	99,30%

Joonisel 7 on näha, kuidas muutub keskmine prognoositud pöördumiskordade arv reaalsete pöördumise arvu suurenemisel. Samuti nähtub, et keskmine prognoositud pöördumiskordade arv kasvab väga aeglaselt. Seetõttu on suure pöördumiskordade arvu korral prognoos ebatäpne. Olukord on täpselt sama, mis oli EHK testandmestikus (joonis 2).

**Joonis 7.** Keskmine prognoositud ning reaalne pöördumiskordade (visiitide) arv TIS andmete põhjal

1.6 Logistilise mudeli testimine

Selles peatükis analüüsime, kui hästi töötab saadud *logit*-mudel, mis prognoosib tõenäosust, et patsient pöörduv perearstiasse sama põhidiagnoosiga vähemalt kaks korda aastas. Selleks on loodud TIS andmestikus uus binaarne tunnus, mis on võrdne nulliga, kui pöörduti vaid korra ja ühega, kui pöördumiskordade arv oli suurem.

TIS andmestiku järgi pöördus 84,78% patsientidest perearstiasse ühe ja sama põhidiagnoosiga ainult üks kord ja 15,22% pöördusid vähemalt kaks korda aastas (tabel 12).

Tabel 12. Pöördumiskordade arv TIS andmestikus

Pöördumiskordade arv	Üks	Kaks või enam
Patsientide arv	1 153 340	207 096
Osatähtsus	84,78%	15,22%

Eeldame, et kui prognoositud tõenäosus on $\geq 0,5$, siis patsient pöörduv perearstiasse sama põhidiagnoosiga rohkem kui üks kord aastas (nagu peatükis 2.3).

Tabelis 13 on näha, et mudel prognoosib õigesti 40,8% juhtudest, kui patsient pöörduv perearstiasse ühe korra aastas ja vaid 11,31% juhtudest õigesti, kui patsient pöörduv vähemalt kaks korda. Ehk ainult pooled patsiendid on õigesti klassifitseeritud. Saadud tulemus on 12% võrra madalam, kui oli EHK testandmestikus. Mudeli järgi pöörduv patsient perearstiasse rohkem kui üks kord aastas, kuid tegelikult pöördus patsient vaid ühe korra. Selline viga esineb 43,97% juhtudest. Teist tüüpi viga, kus patsient pöörduv mudeli järgi ühe korra aastas, kuid tegelikult teeb mitu pöördumist esineb palju harvem (3,91%).

Tabel 13. Prognoosi täpsus

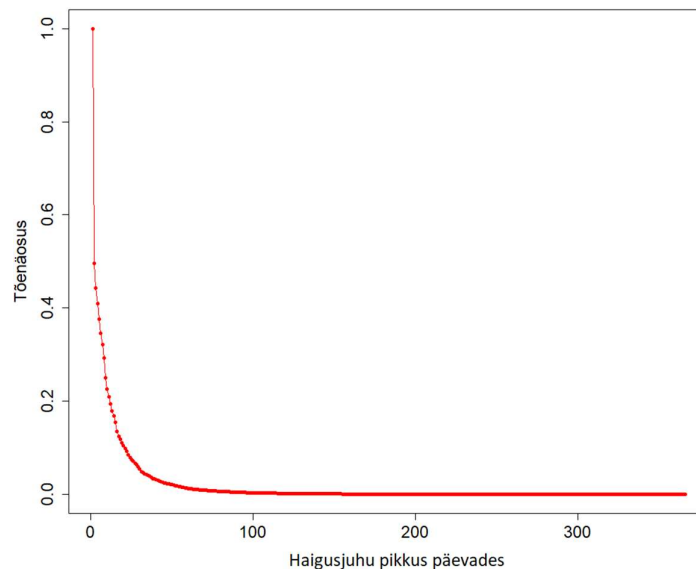
Prognoos	Reaalne pöördumiskordade arv	
	Üks	Vähemalt kaks
Üks	40,80%	3,91%
Vähemalt kaks	43,97%	11,31%

4 Haigusjuhu kestuse prognoosimine tervise infosüsteemi andmestikus

Haigusjuhu kestuse prognoosimiseks TIS andmete põhjal kasutatakse elukestusmodelit. Kuna EHK andmetes puudub informatsioon haigusjuhu lõpust, siis neid andmeid antud prognoosis ei kasutata. Järgnevate andmete puhul loeti haigusjuhuks ka patsiendile retseptiravimite väljakirjutamist ning arstlikku läbivaatust, mis tegelikult on pigem tervises seisundi või haiguskahtluse hindamine ning patsient võib olla terve.

Antud peatükis on huvipakkuvaks tunnuseks patsiendi ühe haigusjuhu pikkus 2017. aasta jooksul. Selle tunnuse modelleerimiseks valitakse elukestusmodel (*Weibull*, *log-logistic*, *log-normal* või *Cox* mudel), mis kõige paremini andmeid kirjeldab. Andmed on võetud TIS andmestikust. Lihtsa juhusliku valiku järgi on andmestik jagatud kaheks osaks – treeningandmestik (70,0%) ning testandmestik (30,0%).

Joonisel 8 on esitatud haigusjuhu kestuse funktsioon, mis näitab tõenäosust p , et patsient on haige kauem kui x päeva. Haiguse alguses (nullpunkt) on kõik patsiendid haiged ehk tõenäosus, et patsient on haige kauem kui null päeva on võrdne ühega. Treeningandmestiku põhjal saavad pooled patsiendid terveks ühe päeva jooksul. Seega tõenäosus, et patsient on haige kauem kui ühe päeva, on 0,496. Haiguse pikkus enam kui seitse päeva on tõenäosusega 0,29.



Joonis 8. Haigusjuhu kestuse funktsioon

Esmalt vaatleme kolme parameetrilist mudelit (*Weibull*, *Log-logistic*, *Log-normal*) kasutades järgmisi tunnuseid: sugu, vanuserühm, põhidiagnoosi rühm, esimese pöördumise viis ja põhidiagnoosi tüüp. Iga mudeli korral on kõik parameetrid olulised (p -väärtus $< 0,05$). Võrreldes mudeleid omavahel AIC'i põhjal on parimaks elukestusmodeliks *Log-normal* (tabel 14).

Tabel 14. Parameetriliste modelite saadud AIC

Mudel	AIC
Weibull	6 577 055
Log-logistic	6 250 598
Log-normal	6 214 736

Mitteparameetrilise Cox mudeli korral on kõik tunnused olulised. Mudeli sobivuse kontrollimiseks kasutame *Grambsch-Therneau* testi, mis kontrollib, kas Cox mudeli eeldused on täidetud (2). Testi põhjal ei sobi saadud mudel andmete kirjeldamiseks (p-väärtus <0,05). Seega edasi prognoosimiseks kasutatakse *Log-normal* mudelit (joonis 9).

$$\begin{aligned}
E(\ln(T)) = & \exp(1,23 + 0,01 \times I(\text{sugu} = \text{naine}) - 0,14 \times I(\text{vanus} = 5 - 14) \\
& - 0,002 \times I(\text{vanus} = 15 - 24) + 0,11 \times I(\text{vanus} = 25 - 44) \\
& - 0,2 \times I(\text{vanus} = 45 - 64) + 0,13 \times I(\text{vanus} = 65 - 84) \\
& - 0,02 \times I(\text{vanus} = 85 +) - 0,03 \times I(\text{diagngr} = C00 - D48) \\
& + 0,06 \times I(\text{diagngr} = D50 - D89) + 0,02 \times I(\text{diagngr} = E00 - E90) \\
& - 0,31 \times I(\text{diagngr} = F00 - F99) - 0,34 \times I(\text{diagngr} = G00 - G99) \\
& - 0,55 \times I(\text{diagngr} = H00 - H59) - 0,28 \times I(\text{diagngr} = H60 - H95) \\
& - 0,03 \times I(\text{diagngr} = I00 - I99) + 0,03 \times I(\text{diagngr} = J00 - J99) \\
& + 0,015 \times I(\text{diagngr} = K00 - K93) - 0,46 \times I(\text{diagngr} = L00 - L99) \\
& - 0,007 \times I(\text{diagngr} = M00 - M99) + 0,06 \times I(\text{diagngr} = N00 - N99) \\
& + 0,16 \times I(\text{diagngr} = O00 - O99) + 0,16 \times I(\text{diagngr} = P00 - P96) \\
& - 0,37 \times I(\text{diagngr} = Q00 - Q99) - 0,08 \times I(\text{diagngr} = R00 - R99) \\
& + 0,54 \times I(\text{diagngr} = S00 - T98) - 0,8 \times I(\text{diagngr} = Z00 - Z99) \\
& + 0,22 \times (\text{tüüp} = \text{esmane}) - 0,12 \times (\text{tüüp} = \text{korduv}) \\
& - 0,24 \times (\text{vis tüüp} = 9018) - 0,03 \times (\text{vis tüüp} = 9061) \\
& + 0,35 \times (\text{vis tüüp} = 9064))
\end{aligned}$$

Märkus: Mudelis tunnus „vis tüüp“ tähistab pöördumise viisi (visiidi tüüp).

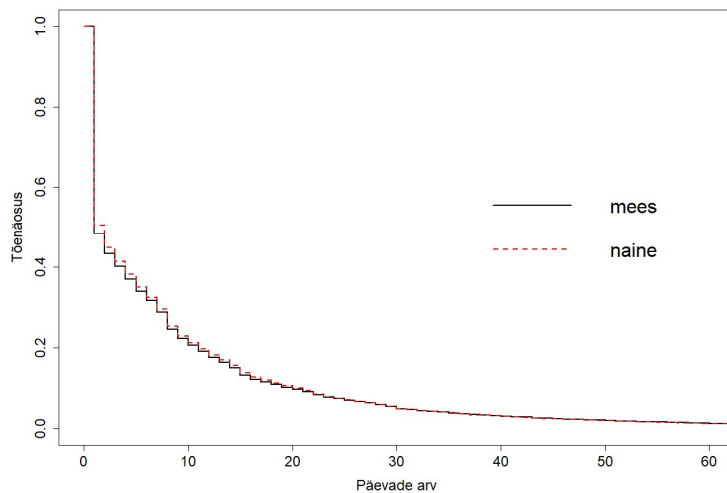
Joonis 9 *Log-normal* elukestusmudel haigusjuhu kestuse prognoosimiseks

Saadud mudel näitab keskmist logaritmitud haigusjuhu kestust. Oodatud haigusjuhu kestuse saamiseks on vaja kasutada saadud skaala parameetrit σ . Antud juhul $\sigma = 1,2$. Seega oodatud haigusjuhu kestuse saab järgmisest valemist:

$$\text{Haigusjuhu kestus} = E(\ln(T)) \cdot \exp\left(\frac{\sigma^2}{2}\right) = E(\ln(T)) \cdot 2,09,$$

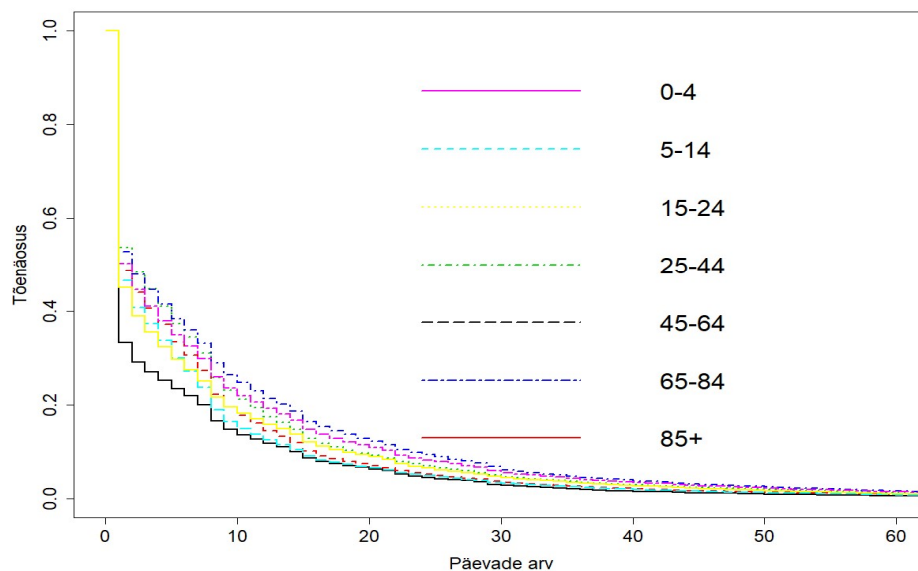
kus $E(\ln(T))$ on eelnevalt kirjeldatud mudel.

Positiivsed koefitsiendid suurendavad haigusjuhu kestuse aega ja negatiivsed vähendavad seda. Kui vaadeldakse kaht patsienti ühest vanuserühmast, sama põhidiagnoosi rühmast, sama diagnoosi tüübi ja esimese pöördumise viisiga, kuid üks on mees ja teine naine, siis naise haigusjuhu kestus on keskmiselt $\exp(0,01) = 1,01$ korda pikem. Joonisel 10 on näha, et naiste ja meeste haigusjuhu kestuse funktsioonid on väga sarnased.



Joonis 10. Haigusjuhu kestuse funktsioon meeste ja naiste vahel

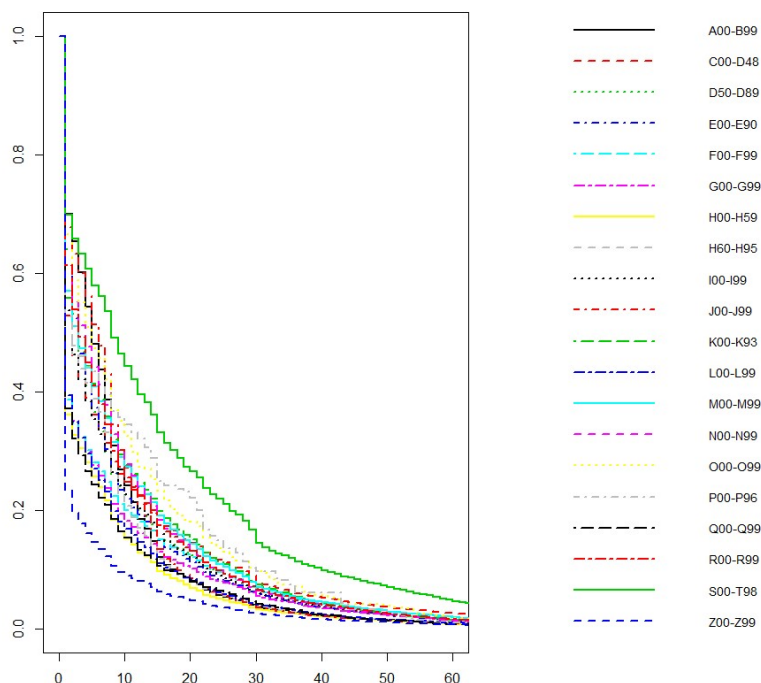
Haigusjuhu kestust mõjutab patsiendi vanuserühm (joonis 11). Kui üks patsient on vanuserühmast 0–4 ja teine 45–64 ning kõik teised tunnused on neil võrdsed, siis vanema patsiendi haigusjuhu kestus on $\frac{1}{\exp(-0,2)} = 1,22$ korda lühem.



Joonis 11. Haigusjuhu kestuse funktsioon erinevates vanuserühmades

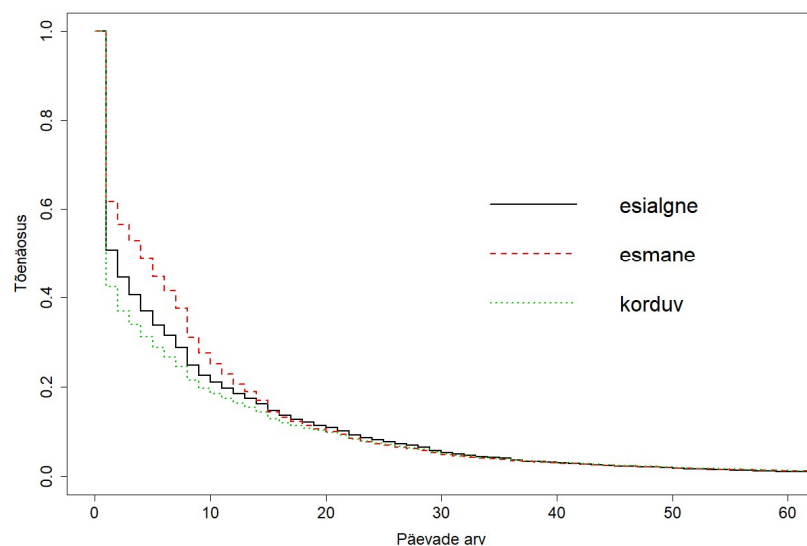
Haigusjuhu kestus erineb põhidiagnoosi rühmades (joonis 12). Kõige pikem haigusjuhu kestus on patsiendil, kelle põhidiagnoos kuulub rühma S00–T98 (vigastused, mürgistused ja teatavad muud välispõhjuste toime tagajärjed), kuid kõik teised tunnused on võrdsed. Kui võrrelda kaht sarnast patsienti, kellest ühel on põhidiagnoos S00–T98 ja teisel A00–B99 (teatavad nakkus- ja parasiithaigused) rühmast, siis esimese patsiendi haigusjuhu kestus on keskmiselt $\exp(0,54) = 1,72$ korda pikem kui teisel. Kui aga võrrelda sarnaseid patsiente põhidiagnoosidega rühmast Z00–Z99 (tervise seisundit mõjustavad tegurid ja kontaktid terviseteenistusega) ja A00–B99 rühmast,

siis esimese patsiendi haigusjuhu kestus on $\frac{1}{\exp(-0,8)} = 2,23$ korda lühem kui teisel. Kuna Z00–Z99 diagnoosirühm hõlmab enamasti pigem profülaktilisi läbivaatlusi ja erinevate tervisetõendite vormistamist, siis on loogiline, et antud põhjusega pöördumisel on juhu kestus teistest oluliselt lühem.



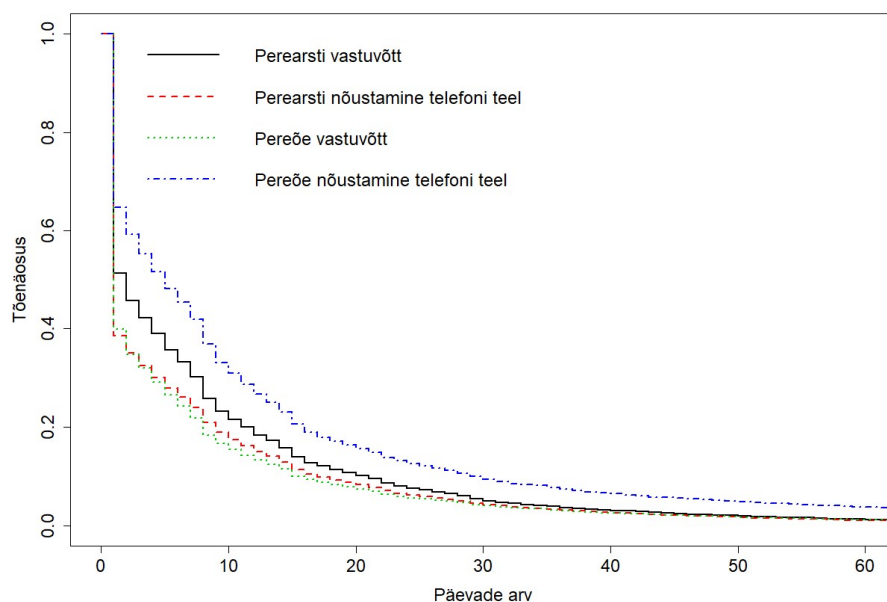
Joonis 12. Haigusjuhu kestuse funktsioon erinevates põhidiagnoosi rühmades

Põhidiagnoosi tüüp on veel üks oluline tunnus, mis mõjutab haigusjuhu kestust (joonis 13). Kui vaatluse all on kaks patsienti ühest vanuserühmast, põhidiagnoosidega ühest rühmast, sama soo ning esimese visiidi tüübiga, kuid ühel neist on esialgne põhidiagnoos (diagnoos ei ole kinnitatud) ja teisel esmane (esmakordselt elus), siis esialgse diagnoosiga patsiendi haigusjuhu kestus on keskmiselt $\exp(0,22) = 1,25$ korda lühem, kui patsiendil esmase põhidiagnoosi tüübiga. Kui aga teise patsiendi põhidiagnoosi tüüp on korduv, siis on tema haiguskestus $\frac{1}{\exp(-0,12)} = 1,13$ korda lühem, võrreldes esialgse põhidiagnoosiga patsiendist.



Joonis 13. Haigusjuhu kestuse funktsioon erinevate põhidiagnoosi tüüpide korral

Viimane mõjutav faktor on esimese pöördumise viis ehk kuidas toimus esimene kontakt perearstiga, kas käidi perearsti või -õe vastuvõtul või toimus perearsti või -õe telefoni teel nõustamine. (joonis 14). Kui võrrelda omavahel kahte sarnast patsienti, kelle esimeseks visiidiks on perearsti vastuvõtt (koodidega 9001, 9002, 9003 ja 9016), siis patsiendi, kelle esimeseks kontaktiks on pereõe nõustamine telefoni teel (9064), on haigusjuhu kestus keskmiselt $\exp(0,35) = 1,42$ korda pikem, kui patsiendil, kes pöördus esmalt perearsti vastuvõtule.



Joonis 14. Haigusjuhu kestuse funktsioon erinevate pöördumise viiside lõikes

Saadud mudeli järgi on haigusjuhu kestus kõige lühem meespatsiendil vanuses 45–64, kellel on korduv põhidiagnoos rühmast Z00–Z99 (tervise seisundit mõjustavad tegurid ja kontaktid terviseteenistusega) ja tema esimeseks kontaktiks oli perearsti nõustamine telefoni teel.

Haigusjuhu prognoositud pikkus on 1,84 päeva. Kõige pikema haigusjuhu kestusega on naine vanuses 65–84, kellel on esmane põhidiagnoos rühmast S00–T98 (vigastused, mürgistused ja teatavad muud välispõhjuste toime tagajärjed) ja tema esimene kontakt on pereõe nõustamine telefoni teel. Sel juhul on haigusjuhu kestus 24,96 päeva. Nagu eelpool mainitud ei ole Z-koodide näol tegemist haigestumisega ja S–T koodidega diagnoosid eeldavadki pikemat paranemise perioodi.

Kasutades treeningandmeid uurime, kui täpne on saadud mudel. Mudeli RMSE on 14,27 ning MAD on võrdne 7,33-ga.

Tabelis 15 on näha, et ainult 5,78% juhtudest prognoosib mudel haigusjuhu kestuse perioodi täpsusega üks päev. Prognoosi viga 1–5 päeva esineb kõige rohkem – 47,78%-l patsientidest. Prognoosi viga ± 10 päeva esineb 86,57% juhtudest. See on hea tulemus, kuid reaalsuses on suur erinevus, kas haigusjuhu kestus on üks või 10 päeva. Patsiendi haigusjuhu kestusest sõltub perearstiabi töömaht. Mida pikem on haigusjuht, seda rohkem tööd on perearsti meeskonnal.

Tabel 15. Prognoosi vea jaotus

Vea suurus (päevades)	<1	1–5	5–9	9–20	20–29	29–40	40–100	>100
Osatähtsus	5,78%	47,78%	33,01%	7,22%	2,72%	1,24%	1,95%	0,30%

Kasutatud kirjandus

1. Ambulatoorse epikriisi täitmise juhend (2014). http://pub.e-tervis.ee/manuals/Ambulatoorse%20epikriisi%20%C3%A4itmise%20juhend/5/Ambulatoorse_epikriisi_taitmise_juhend.pdf
2. RHK-10 instruksioonide käsiraamat (1996). http://pub.e-tervis.ee/classifications/RHK-10/3/RHK-10_k%C3%A4siraamat.pdf
3. Tiit EM, Tooding LM. Statistikalleksikon. Tartu: Tartu Ülikooli Kirjastus; 2019
4. Üldarstiabi rahastamise lepingud, lisa 1, Eesti Haigekassa https://www.haigekassa.ee/sites/default/files/RRL/%C3%9CRL_lisa_1.pdf
5. Tervise infosüsteemi edastatavate dokumentide andmekoosseisud ning nende säilitamise tingimused ja kord (17.09.2008). RT I, 12.03.2019. <https://www.riigiteataja.ee/akt/13349775?leiaKehtiv>
6. 2018. aasta majandusaasta aruanne, Eesti Haigekassa, 2019 https://www.haigekassa.ee/sites/default/files/uuringud_aruanded/Eesti%20Haigekassa_majandusaruanne_2018.pdf
7. Idavain, J., Anderson, E., Rätsep, M. (2019). Tervise infosüsteemi andmete kasutamine. [ettekanne]. Esitatud tervishoiustatistika aruannete ja registriteatiste koolituses, Tallinn, Tervise Arengu Instituut, detsember 2019. https://intra.tai.ee/images/3_Tervishoiustatistika_koolitus2019_tervise_infosysteemi_vordlus_Anderson_Idavain_Ratsep.pdf
8. Anderson E., Panov L. Perearstiabiastutuste ambulatoorsed visiidid tervise infosüsteemi andmetes 2015. aastal. Tallinn: Tervise Arengu Instituut; 2017. https://intra.tai.ee/images/prints/documents/149872658710_Perearstiabiastutuste_ambulatoorsed_visiidid_tervise%20infos%C3%BCsteemi%20andmetes_2015_a.pdf
9. Anderson, E., Eigo, N., Kirpu, V., Panov, L., Rätsep, M., Sokurova, D., Väärsi, K. Perearstiabisse pöördumise põhjused, 2017. Tallinn: Tervise Arengu Instituut; 2018. <https://www.tai.ee/et/terviseandmed/uuringud/download/468>
10. Käärrik M. Üldistatud lineaarsed mudelid. Loengukonspekt; 2019. <https://courses.ms.ut.ee/2019/glm/spring/Main/Lectures>
11. Käärrik E. Andmeanalüüs II. Loengukonspekt; 2013. <http://dSPACE.ut.ee/bitstream/handle/10062/35401/AndmeanaluusII.pdf>
12. Käärrik M. Elukestusmudelid. Loengukonspekt; 2018. <https://courses.ms.ut.ee/2018/survmod/fall/Main/Lectures>
13. Eesti Haigekassa tervishoiuteenuste loetelu. Riigi Teataja. Vastu võetud 27.04.2017 nr 82. <https://www.riigiteataja.ee/akt/129042017007>

LISA

Rahvusvahelise haiguste ja tervisega seotud probleemide statistilise klassifikatsiooni 10. versiooni peatükid

Kood	Nimetus
A00–B99	Teatavad nakkus- ja parasiithaigused
C00–D48	Kasvajad
D50–D89	Vere- ja vereloomeelundite haigused ja teatavad immuunmehhanismiga seotud haigusseisundid
E00–E90	Sisesekretsiooni, toitumis- ja ainevahetusehaigused
F00–F99	Psüühika- ja käitumishäired
G00–G99	Närvisüsteemihaigused
H00–H59	Silma- ja silmamanuste haigused
H60–H95	Kõrva- ja nibujätkehaigused
I00–I99	Vereringeelundite haigused
J00–J99	Hingamiselundite haigused
K00–K93	Seedeelundite haigused
L00–L99	Naha- ja nahaaluskoe haigused
M00–M99	Lihaskonna ja sidekoe haigused
N00–N99	Kuse- ja suguelundite haigused
O00–O99	Raseduse ja sünnitusega seotud tüsistused
P00–P96	Sünniperioodis tekkivad teatavad seisundid
Q00–Q99	Kaasasündinud väärarendid, deformatsioonid ja kromosoomianomaaliad
R00–R99	Mujal klassifitseerimata sümptomid, tunnused, kliiniliste ja laboratoorsete leidude hälbed
S00–T98	Vigastused, mürgistused ja teatavad muud välispõhjuste toime tagajärjed
V01–Y98*	Haigestumise ja surma välispõhjused
Z00–Z99	Tervise seisundit mõjustavad tegurid ja kontaktid tervise teenistusega

* Andmestikus puudus V01–Y98 rühm.

Tervisestatistilised andmed ja informatsioon:

- **Tervisestatistika ja terviseuuringute andmebaas**
<http://www.tai.ee/tstua>
- **Tervise Arengu Instituudi tervisestatistika osakonna veebileht**
<http://www.tai.ee/tegevused/tervisestatistika>
- **Andmepäring Tervise Arengu Instituudile**
tai@tai.ee
- **Statistikaameti andmebaas**
<http://www.stat.ee/>
- **Euroopa Liidu statistika**
<http://ec.europa.eu/eurostat>
- **Maaailma Terviseorganisatsiooni (WHO) Euroopa esinduse tervise andmebaas**
<http://data.euro.who.int/hfadbf/>
- **Majandusliku Koostöö ja Arengu Organisatsiooni (OECD) andmebaas**
http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT

