

Tartu Ülikool  
Loodus- ja täppisteaduste valdkond  
Matemaatika ja statistika instituut

Eneli Maal

# Võrkpallurite õlavöötmevigastuste esinemisriski uurimine elukestusanalüüsi abil

Matemaatilise statistika eriala  
Bakalaureusetöö (9 EAP)

Juhendaja: lektor Märt Möls

Tartu 2018

# Võrkpallurite õlavöötmevigastuste esinemisrisi uurimine elukestusanalüüsi abil

Käesoleva bakalaureusetöö eesmärk on välja selgitada, miks paljud võrkpallurid vaevlevad õlavöötme vigastuste küüsis ning kas vigastuste tekkeriski saab vähendada. Analüüsi läbiviimisel kasutatakse elukestusanalüüsi meetodeid. Esimeses peatükis tutvustatakse töös kasutatavaid meetodeid - Kaplan-Meieri meetod, *log-rank* test ja Coxi mudel elukestuskõverate erinevuse tuvastamiseks. Teises peatükis on eelnevalt nimetatud meetodeid rakendatud mängijate andmetele. Lisaks on arutletud testi võimsuse teemal, kuna antud töö valimimaht on küllaltki väike.

**CERCS teaduseriala:** P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika

**Märksõnad:** elukestusanalüüs, SAS (programmeerimiskeel)

# Survival analysis of volleyball players' shoulder girdle injuries

The aim of this thesis is to find out why a lot of volleyball players suffer from injuries to the shoulder girdle and if the risk of injury can be reduced. The analysis is carried out using survival analysis methods. The first chapter introduces the features of the work methods - Kaplan-Meier method, *log-rank* test and Cox model for identifying the difference in survival curves. In the second chapter, the above methods have been applied to the players data. In addition, the power of the test is reviewed as the sampling size of this work is rather small.

**CERCS research specialisation:** P160 Statistics, operation research, programming, actuarial and financial engineering

**Keywords:** survival analysis, SAS (programming language)

# Sisukord

<b>Sissejuhatus</b> .....	<b>4</b>
<b>1 Teoreetiline osa</b> .....	<b>5</b>
1.1 Elukestusanalüüs.....	5
1.2 Kaplan-Meieri hinnang.....	6
1.3 <i>Log-rank</i> test.....	9
1.4 Riskifunktsioon .....	10
1.5 Võimsus .....	11
1.6 Coxi võrdeliste riskide mudel .....	11
<b>2 Analüüs</b> .....	<b>13</b>
2.1 Algandmete kirjeldus .....	13
2.2 Kaplan-Meieri elukestuskõver.....	13
2.3 <i>Log-rank</i> testi tulemused.....	14
2.4 Mitmese testimise probleem.....	17
2.5 Testi võimsus .....	18
2.6 Coxi mudel .....	20
<b>Kokkuvõte</b> .....	<b>22</b>
<b>Kasutatud kirjandus</b> .....	<b>23</b>
<b>Lisad</b> .....	<b>24</b>

## Sissejuhatus

Sportlaskarjäär nõuab suurt tööd ja pingutusi. Paljud spordialad nõuavad kiireid ja jõulisi liigutusi, mis võivad aeg-ajalt tekitada lihastes mikro- või makrovigastusi. Samuti on ka võrkpalluritega, kelle löögikäsi teeb pidevalt suure kiirusega löögi- või serviliigutusi, et palli võimalikult suure jõuga tabada. Sellest tingituna võib aga tekkida õlavöötmes lihasrebend või muu vigastus, mille ravimine on aeganõudev töö.

Käesolevas bakalaureusetöös uuritakse Eesti meessoost võrkpallurite õlavöötmevigastuste seost kõige üldisemate kehaliste ja treeninguga seotud teguritega, kasutades elukestusanalüüsi. Eesmärgiks on otsida võimalusi, mis alandaks pallurite õlavöötme vigastuste tekkeriski.

Töö esimeses peatükis antakse ülevaade elukestusanalüüsist ning seejärel selgitatakse, kuidas kasutada Kaplan-Meieri meetodit, *log-rank* testi ja Coxi võrdeliste riskide mudelit. Teises peatükis on antud meetodeid kasutatud võrkpallurite andmetel. Esmalt on välja toodud üldine Kaplan-Meieri hinnang elukestuskõverale ning seejärel on kasutatud *log-rank* testi kahe grupi elukestuskõverate võrdlemiseks, lisaks veel Coxi mudelist saadavad tulemused. Samuti on teises peatükis käsitletud testi võimsusega seonduvaid probleeme.

Töö on kirjutatud tekstitöötlusprogrammiga LATEX. Analüüs on läbi viidud statistikatarkvaras SAS.

# 1 Teoreetiline osa

Antud peatükk tugineb Kleinbaumi ja Kleini raamatul „Survival Analysis: A Self-Learning Text” [1], v.a juhul, kui on märgitud teisiti.

## 1.1 Elukestusanalüüs

Elukestusanalüüs on kogum statistilistest protseduuridest, kus huvipakkuvaks tunnuseks on aeg sündmuse toimumiseni. Tunnust aeg mõõdetakse aastates, kuudes, nädalates või päevades alates määratud ajast (näiteks võrkpallitreeningutega alustamisest) kuni sündmuse toimumiseni (näiteks õlavõõtmevigastuse teke). Sündmuste all mõeldakse olenevalt uuringuvaldkonnast surma, haiguse esinemist, tervenemist, seadme rikkimist või muud taolist juhtumit.

Enamik elukestusanalüüsides peavad arvestama ka tsenseerimisega. Olemuselt esineb tsenseeritud andmeid, kui uuringus osalenud subjektidel ei ole uuringu lõpuks uuritavat sündmust toimunud.

Üldiselt on kolm põhjust, miks tsenseeritud andmed esinevad:

- indiviid ei koge sündmust enne uuringu lõppu;
- indiviidiga ei ole uuringu käigus võimalik enam ühendust saada;
- indiviid sureb uuringu toimumise ajal, kuid põhjus ei ole uuringuga seotud.

Antud põhjuste tõttu on enamik elukestusandmeid paremalt tsenseeritud ehk andmete puudumine esineb uuringu lõppjärgus. Vasakult tsenseeritud andmed esinevad juhul, kui sündmus on toimunud juba enne uuringu algust.

Toome näite paremalt tsenseeritud andmetest. Indiviid  $A$  alustas võrkpalli mängimist 15-aastaselt ja on hetkel 21-aastane ning uuringu lõpuks tal õlavigastust tekkinud polnud, seega on tema puhul aeg huvipakkuva sündmuseni tsenseeritud, kusjuures tsenseerimisaeg on 6 aastat, kuna nii kaua on ta võrkpalli harrastanud. Indiviid  $B$  aga alustas võrkpallitreeningutega 11-aastaselt ja on hetkel 19-aastane ning vigastas end 14-aastaselt, seega tema andmed ei ole tsenseeritud (teame täpselt, kui kaua aega kulus huvipakkuva sündmuse toimumiseni), seejuures on vigastustevaba aeg sportlaskarjääri jooksul 3 aastat ehk nii pikk, kui on aeg mänguga alustamisest kuni vigastuseni.

Elukestusanalüüsis tähistatakse aega fikseeritud momendist kuni sündmuse toimumiseni tähega  $T$ , ( $T \geq 0$ ), selle realiseerunud väärtust tähega  $t$ , ( $t \geq 0$ ) ning

tsenseerimise aega tähega  $C$ . Indikaatortunnus  $\delta = I(T < C)$  on võrdne arvuga 1, kui sündmuse toimumist oli võimalik vaadelda (sündmus toimus enne tsenseerimist) ning arvuga 0, kui huvipakkuvat sündmust polnud võimalik vaadelda (toimus tsenseerimine).

**Definitsioon.** Elukestusfunktsioon on defineeritud valemiga

$$S(t) = P(T > t).$$

Elukestusfunktsioonil on järgmised teoreetilised omadused:

- funktsioon on mittekasvav, kui  $t$  väärtus kasvab;
- $S(0) = 1$  ehk elukestusaja tõenäosus olla suurem nullist on võrdne ühega, kuna uuringut alustades ei ole veel kellelgi huvipakkuvat sündmust toimunud;
- $S(\infty) = 0$  ehk kui uuring kestaks lõpmata kaua, siis oleks uuritav sündmus toimunud kõigil uuritavatel ning elukestusfunktsiooni väärtus langeb nulli.

Elukestusfunktsiooni hinnang on piiratud valimi korral üldiselt treppfunktsioon, mis ajalise piirangu (näiteks uuringule seatud ajalise pikkuse) tõttu ei pruugi jõuda väärtuseni null.

## 1.2 Kaplan-Meieri hinnang

Kaplan-Meieri hinnang on mõeldud hindamaks elukestusfunktsiooni juhul, kui osad vaatlused on tsenseeritud. Lähteandmed on järgmisel kujul:

Tabel 1. Kaplan-Meieri andmetabel.

Järjestatud jälgimisajad $t_{(j)}$	Riskigrupi suurus $n_j$	Sündmuste esinemiste arv $m_j$	Tsenseeritud objektide arv vahemikus $[t_{(j)}, t_{(j+1)})$ $q_j$
$t_{(0)}$	$n_0$	$m_0$	$q_0$
$t_{(1)}$	$n_1$	$m_1$	$q_1$
$t_{(2)}$	$n_2$	$m_2$	$q_2$
.	.	.	.
.	.	.	.
.	.	.	.
$t_{(k)}$	$n_k$	$m_k$	$q_k$

Antud tabeli esimeses veerus on kasvavalt järjestatud uuritavate indiviidide jälgimisajad ehk indiviidil esinenenud sündmuse toimumise hetke ja uuringuga liitumise aegade vahed. Riskigrupi suuruse  $n_j$  all on mõeldud nende uuringuga liitunud indiviidide arvu, kellel antud ajahetkeks  $i$  sündmust toimunud pole. Sündmuste esinemiste arv  $m_j$  väljendab ajahetkel  $j$  toimunud sündmuste arvu. Tsenseeritud objektide arv näitab antud jälgimisaja  $t_j$  ja järgmise jälgimisaja  $t_{j+1}$  vahel tsenseeritud objektide arvu (kaasa arvatud uuringust lahkunute arv).

Kaplan-Meieri hinnangut elukestusfunktsioonile saab arvutada järgmise valemiga

$$\hat{S}(t_{(j)}) = \prod_{t_{(i)} \leq t_{(j)}} \frac{n_i - m_i}{n_i}.$$

**Näide 1.1.** Osalegu uuringus 6 võrkpallurit vanuses 17, 17, 18, 19, 22 ja 26 aastat, kes alustasid võrkpallitreeningutega vastavalt vanuses 11, 9, 12, 11, 10 ja 8. Neist teisel, viiendal ja kuuendal mängijal on uuringu lõpuks esinenud vigastus, vastavalt vanuses 14, 17 ja 25 ehk aeg sündmuse esinemiseni on nende mängijate puhul vastavalt 5, 7 ja 17 aastat. Ülejäänud kolme mängija andmed on tsenseeritud, kuna uuringu lõpuks ei olnud neil vigastust tekkinud. Antud sportlased mängisid seega 6, 6 ja 8 aastat ilma, et neil oleks vigastus tekkinud. Uuringu algusajaks on määratud mängija vanus treeningutega alustamise ajal, seega kuuluvad kõik algul riskigruppi, kuna kellelgi polnud enne treeningutega alustamist vigastust esinenud. Aeg uuringu algusest kuni sündmuseni on mõõdetud aastates.

Tabel 2. Kaplan-Meieri andmetabel näites antud andmete korral.

$t_{(j)}$	$n_j$	$m_j$	$q_j$	$\hat{S}(t_{(j)})$
0	6	0	0	1
5	6	1	0	$1 \cdot \frac{6-1}{6} = \frac{5}{6}$
6	5	0	2	$\frac{5}{6} \cdot \frac{5-0}{5} = \frac{5}{6}$
7	3	1	0	$\frac{5}{6} \cdot \frac{3-1}{3} = \frac{5}{9}$
8	2	0	1	$\frac{5}{9} \cdot \frac{2-0}{2} = \frac{5}{9}$
17	1	1	0	$\frac{5}{9} \cdot \frac{1-1}{1} = 0$

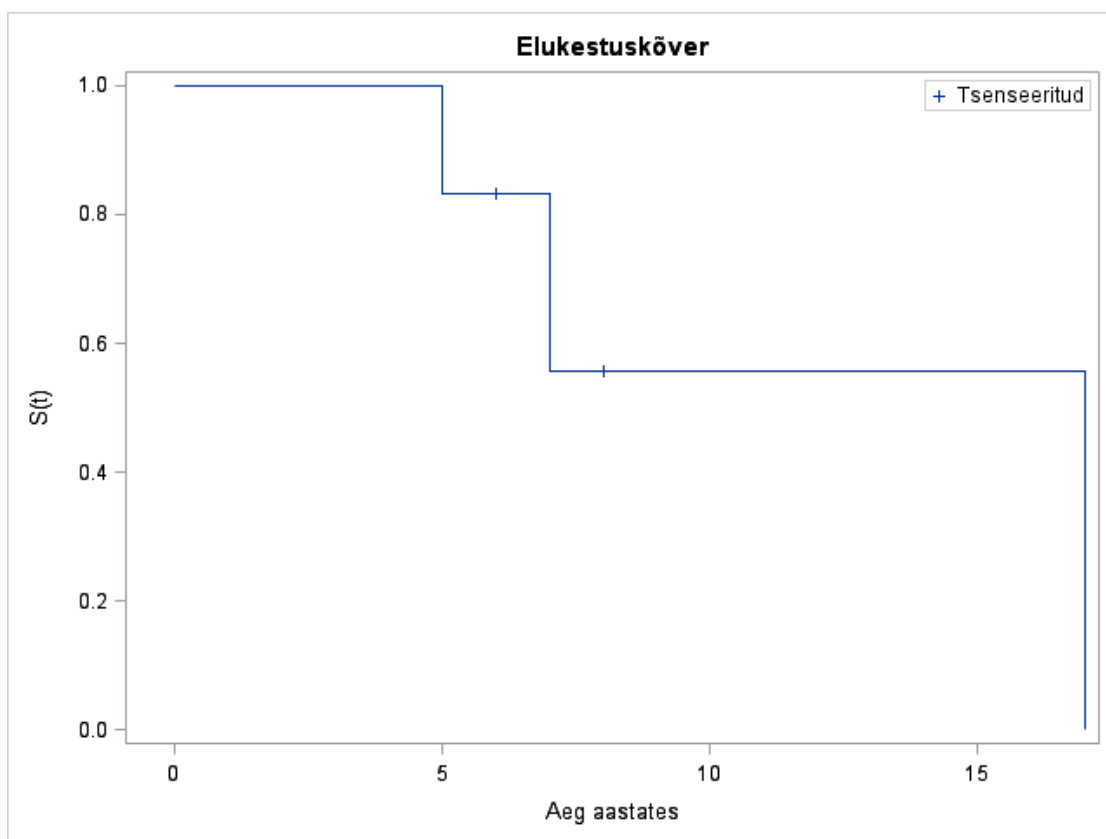
Mängijate jälgimisajad järjestatakse kasvavalt ja lisatakse, mitu vigastust või tsenseerimist vastaval ajahetkel toimus. Andmetest näeme, et kõige lühem jälgimisaeg oli 5 aastat ning antud mängijal on selle aja jooksul tekkinud vigastus. Elukestusfunktsiooni hinnangu väärtus langeb väärtuseni  $5/6$  just seetõttu, et ühel uuritavatest on sündmus toimunud. Edasi näeme, et riskigruppi kuulub nüüdseks 4 inimest, sest vigastatud mängijat enam ei jälgita. Järgmise kahe palluri jälgimisajaks oli 6 aastat ning mõlema andmed on tsenseeritud, kuna uuringu lõpuks neil vigastust tekkinud polnud. Selle tulemusena ei lange ka elukestusfunktsiooni hinnangu väärtus.

Järele on jäänud 3 uuritavat, kelle jälgimisaeg on suurem kui 6 aastat. Neist esimene, keda jälgiti kokku 7 aastat, kuulub vigastatute hulka ja seega kahaneb tema tulemuse põhjal ka elukestusfunktsiooni hinnang väärtuseni  $5/9$ . Sportlase andmed, keda jälgiti 8 aastat, on tsenseeritud, seega tema puhul hinnang jääb eelmisega samaks. Kuna aga viimane mängija, keda jälgiti kokku 17 aastat, on saanud vigastuse, siis langeb tema tulemuste põhjal elukestusfunktsiooni hinnang nulli vastavalt valemile 1.1 järgmiselt:

$$\frac{6-1}{6} \cdot \frac{5-0}{5} \cdot \frac{3-1}{3} \cdot \frac{2-0}{2} \cdot \frac{1-1}{1} = 0,$$

kuna pikema jälgimisajaga indiviide meil uuringus polnud. Toome näite jaoks ka Kaplan-Meieri hinnangu elukestusfunktsioonile (joonis 1).





Joonis 1. Kaplan-Meieri hinnang vigastuse tekkele.

Järgneva peatüki jaoks toome sisse uued tähistused riskigrupi suuruse ja sündmuste esinemiste arvu jaoks. Kuna siiani kasutasime muutujaid vaid ühe grupi jaoks, siis kahe grupi võrdluseks tähistame riskigrupi suuruse tähega  $n_{ij}$  ning sündmuste esinemiste arvu  $m_{ij}$ , kus antud töö puhul  $i = 1, 2$  tähistab gruppe.

### 1.3 *Log-rank* test

Elukestuskõverate võrdlemiseks kahes või enamas grupis saab kasutada *log-rank* testi. Testi statistik kasutab hii-ruut testile omaselt sündmuste esinemiste tegelikke ja oodatud tulemusi. Meie vaatleme kahe grupi võrdlust, kus oodatavad sündmuste esinemiste arvud leitakse kahes grupis igal jälgimisajal järgmiste valemitega:

$$\begin{aligned}
 e_{1j} &= \frac{n_{1j}}{n_{1j} + n_{2j}} \cdot (m_{1j} + m_{2j}); \\
 e_{2j} &= \frac{n_{2j}}{n_{1j} + n_{2j}} \cdot (m_{1j} + m_{2j}),
 \end{aligned}
 \tag{1}$$

kus  $j$  tähistab jälgimisaja järjekorranumbrit,  $\frac{n_{1j}}{n_{1j} + n_{2j}}$  ja  $\frac{n_{2j}}{n_{1j} + n_{2j}}$  väljendavad riskigrupi proportsioone  $j$ -ndal jälgimisajal ning  $m_{1j} + m_{2j}$  on esinenud sündmuste koguarv kahe grupi peale kokku  $j$ -ndal jälgimisajal.

Kui võrdluses on kaks gruppi, siis formuleeritakse *log-rank* statistiku jaoks tegelikest tulemustest  $m_{ij}$  lahutatud oodatavate tulemuste  $e_{ij}$  (vaata 1) summa ehk

$$O_i - E_i = \sum_{j=1}^n (m_{ij} - e_{ij}),$$

kus  $i = 1, 2$  ja  $n$  on erinevate jälgimisaegade koguarv. Hinnang vahe  $O_i - E_i$  dispersioonile arvutatakse valemiga:

$$Var(O_i - E_i) = \sum_{j=1}^n \frac{n_{1j} \cdot n_{2j} \cdot (m_{1j} + m_{2j}) \cdot (n_{1j} + n_{2j} - m_{1j} - m_{2j})}{(n_{1j} + n_{2j})^2 \cdot (n_{1j} + n_{2j} - 1)},$$

kus  $i = 1, 2$ . Dispersioonkordaja valemis on kasutatud kummagi grupi riskigrupi suurust ( $n_{ij}$ ) ja sündmuse toimumise arvu kummaski grupis ( $m_{ij}$ ) ajahetkel  $j$ . Arvutusvalem *log-rank* testi statistikule avaldub kahe eelneva tulemuse abil, kusjuures tulemus ei erine grupiti (sama teststatistiku väärtus saadakse nii valiku  $i = 1$  kui valiku  $i = 2$  korral):

$$L = \frac{(O_i - E_i)^2}{Var(O_i - E_i)}.$$

Antud statistikuga kontrollime hüpoteesipaari, kus

$H_0$ : elukestuskõverad pole erinevad;

$H_1$ : elukestuskõverad on erinevad.

Kui kehtib nullhüpotees ehk elukestuskõveratel ei ole erinevust ja kui valimid on suured, siis on *log-rank* statistik ligikaudu hii-ruut jaotusega vabadusastmete arvuga 1.

## 1.4 Riskifunktsioon

Tunnuse  $T$  (elukestuse) jaotust saab kirjeldada ka riskifunktsiooni abil. Riskifunktsioon defineeritakse järgmiselt:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t},$$

kus lugejas on sündmuse toimumise tinglik tõenäosus intervallis  $[t, t + \Delta t)$  tingimusel, et sündmus pole toimunud enne ajahetke  $t$  ja nimetajas on vaadeldud aja intervalli pikkus  $\Delta t$ . Nende jagatisena saame seega sündmuse toimumise hinnangu ajajuhiku kohta ning piirväärtuse võtmise järel oleme saanud riski hetke  $t$  jaoks. [7]

Paneme tähele, et kui elukestusfunktsioon kirjeldab sündmuse toimumata jäämistel, siis riskifunktsioon iseloomustab just nimelt sündmuse toimumist. Riskifunktsioonil on järgmised kaks omadust:

- $h(t) \geq 0$ ;
- $h(t)$  ei ole ülevalt tõkestatud.

Need omadused tulenevad otseselt riskifunktsiooni definitsioonist, kuna nii lugeja kui nimetaja on mittenegatiivsed ning  $\Delta t$  on mis tahes nullilähedane väärtus.

## 1.5 Võimsus

**Definitsioon.** Testi võimsuseks  $P$  nimetatakse tõenäosust lükata ümber nullhüpoteesi, kui see on vale ehk  $P = 1 - \beta = \mathbf{P}(\text{lükata ümber } H_0 | H_0 \text{ on vale})$ , kus  $\beta$  on II liiki vea tegemise tõenäosus ehk  $\beta = \mathbf{P}(\text{ei lükata ümber } H_0, \text{ kui ta on vale})$  [3].

Lane [4] on kirja pannud viis tegurit, millest võimsus sõltub:

- valimi suurus - mida suurem on valimi suurus, seda parem on testi võimsus;
- standardhälve - võimsus on suurem, kui standardhälve on väike;
- testitava parameetri tõeline väärtus - mida suurem on vahe tõelise väärtuse ja nullhüpoteesis seatud väärtuse vahel, seda kõrgem on võimsus;
- olulisuse nivoo  $\alpha$  - mida suurem on olulisuse nivoo, seda kõrgem on testi võimsus;
- ühe- või kahepoolsed testid - võimsus on ühepoolse testi korral suurem.

## 1.6 Coxi võrdeliste riskide mudel

Ühe samuti laialt levinud elukestusanalüüsi mudelina kasutatakse Coxi võrdeliste riskide mudelit. Coxi mudel uurib elukestuse ja kirjeldavate tunnuste vahelist seost ning püüab nende põhjal meid huvitava sündmuse toimumise riski hinnata. Coxi võrdeliste riskide mudel esitatakse tihti just riskifunktsiooni kaudu:

$$h(t, \mathbf{X}) = h_0(t) \cdot e^{\sum_{i=1}^p \beta_i X_i},$$

kus  $\mathbf{X} = (X_1, X_2, \dots, X_p)$  on seletavate tunnuste vektor, mis on modelleeritud ennustamiseks indiviidi riski sündmuse esinemisele ning  $\beta = (\beta_1, \beta_2, \dots, \beta_p)$  on parameetrite ehk regressioonikordajate vektor.

Näeme, et mudeli baasriski funktsioon  $h_0(t)$  sisaldab aega  $t$ , kuid mitte seletavaid tunnuseid  $X_i$ . Seevastu eksponentsiaalosas  $e^{\sum_{i=1}^p \beta_i X_i}$  ei arvestata ajaga  $t$ , kuid see sisaldab seletavaid tunnuseid, mis on ajast sõltumatud.

Coxi mudelisse kuuluv baasriski funktsioon on määratlemata funktsioon, seega on Coxi mudel poolparameetiline. Coxi mudel annab siiski üsna häid hinnanguid regressioonikordajatele ja riskitiheduste suhetele, seega on tulemused ligilähedased nendele tulemustele, mis tuleksid kasutades õiget parameetrilist mudelit.

**Näide 1.2.** Võtame Coxi mudeli selgitamiseks vaatluse alla kaks mängijat, kellest üks venitas end treeningutel regulaarselt ( $X = 0$ ), kuid teine mitte ( $X = 1$ ). Kui mängija ei venita treeningutel end regulaarselt ehk seletava tunnuse väärtus on null, siis jääb Coxi riskide mudelisse alles vaid baasriski funktsioon. Kui mängijal on tavaks end treeningutel venitada, siis jääb eksponentosa mudelisse sisse:

$$\begin{aligned}h(t|X = 0) &= h_0(t) \cdot e^{\beta \cdot 0} = h_0(t) \cdot e^0 = h_0(t); \\h(t|X = 1) &= h_0(t) \cdot e^{\beta \cdot 1} = h_0(t) \cdot e^\beta.\end{aligned}$$

Seega mängijal, kes oma lihaseid treeningul ei venita, on risk sündmuse toimimiseks  $e^\beta$  korda suurem kui end venitava mängijal. Arv  $\beta$  on tunnusele vastav regressioonikordaja.

## 2 Analüüs

### 2.1 Alandmete kirjeldus

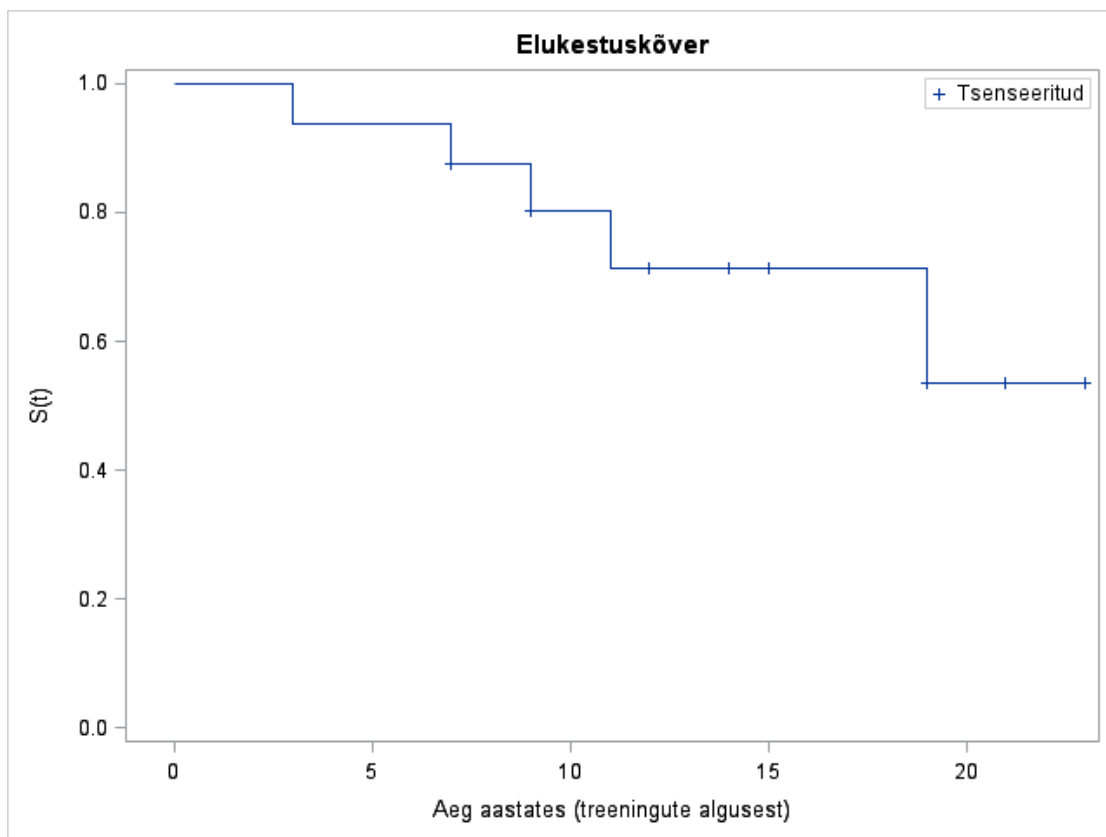
Andmed on kogutud meessoost mängijatelt, kelle hulka kuuluvad Eesti rahvusmeeskonna liikmed ning kahe Eesti mesitriliiga klubi mängijad. Alandmestikus on 14 mängija andmed 10 huvipakkuva tunnusega:

- õlavöötme vigastus, 0 - vigastus puudub, 1 - vigastus toimunud;
- vanus esmase vigastuse tekkimisel;
- vanus praegu;
- pikkus;
- domineeriva käe pikkus;
- vanus võrkpallitreeningute alustamisel;
- jõusaalitreeningute arv nädalas enne vigastuse tekkimist;
- pallitreeningute arv nädalas enne vigastuse tekkimist;
- lihaste venitamine enne ja pärast treeningut vigastuse tekkimise eel;
- regulaarselt lihasmassaažide saamine enne vigastuse tekkimist.

Analüüsi jaoks moodustame veel lisaks uue tunnuse, milleks on vahe praeguse vanuse ja vanuse vahel, mil tekkis esmane õlavigastus või kui vigastust uuringu jooksul toimunud pole, siis praeguse vanuse ja treeningutega alustamise vanuse vahe. Samuti on analüüsi tarvis mängijad iga tunnuse puhul kahte gruppi jaotatud. Jaotamise aluseks võtsime mediaani.

### 2.2 Kaplan-Meieri elukestuskõver

Vaatleme, millise tulemuse annab elukestuskõver, mis iseloomustab vigastuseta mängijate osakaalu teatud aja möödumisel peale treeningutega alustamist. Pane me tähele, et andmestik on seisuga 17.04.2017 ning seetõttu mängijad, kellel pole selleks hetkeks vigastust esinenud, kuuluvad tsenseeritud uuritavate hulka. Vaatleme meie andmete põhjal saadavat elukestuskõverat, kus on kõik andmestikus olevad mängijad ning mis iseloomustab vigastuseta mängijate osakaalu (joonis 2).



Joonis 2. Kaplan-Meieri hinnang vigastusteta mängijate osakaalule.

Jooniselt on näha, et andmetes leidis 8 mängijat, kellel uuringu lõpuks vigastust ei olnud tekkinud ning nende tulemused on tähistatud pluss-märgiga. Seega hinnanguliselt 50% uuritutest ei ole saanud õlavigastust. Samuti näeme, et hinnanguliselt pooltel enam kui 20 aastat treeninud mängijatel esineb õlavigastus.

### 2.3 *Log-rank* testi tulemused

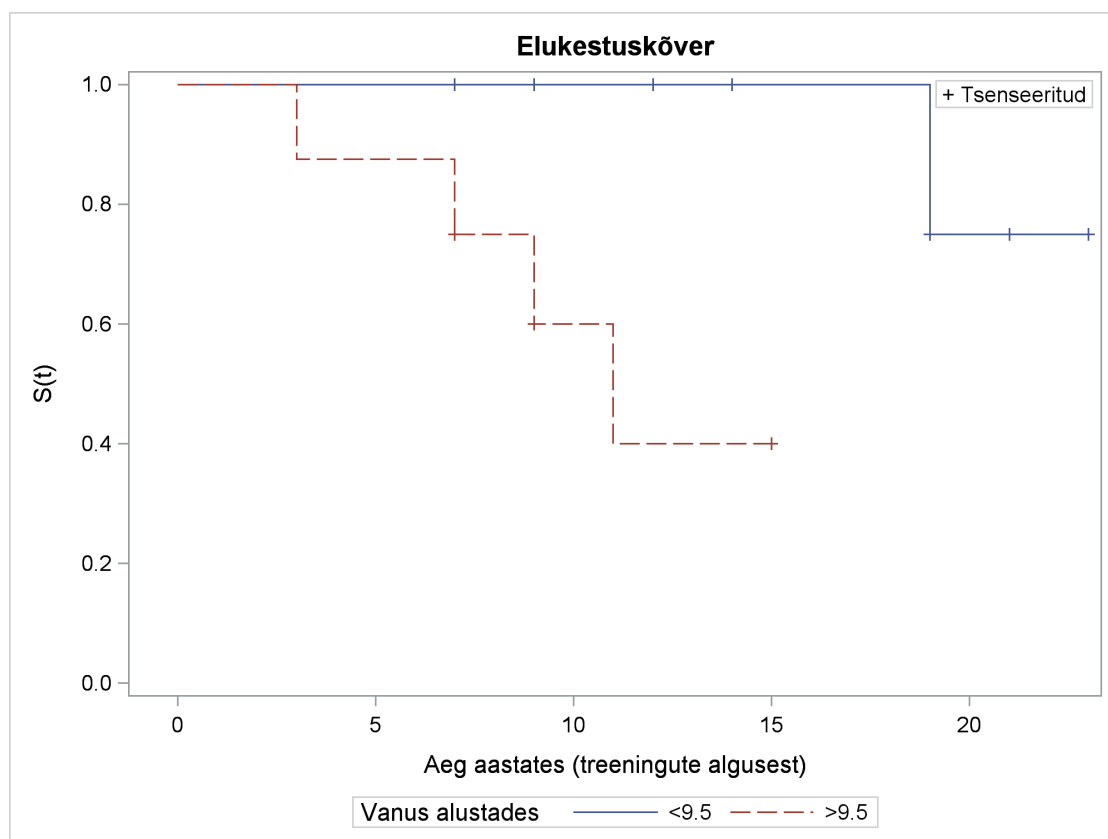
Toome alustuseks kõigi tunnuste kohta tabeli (tabel 3) koos *log-rank* testi tulemuste p-väärtustega.

Tabel 3. *Log-rank* testi tulemused.

Tunnus	p-väärtus
Regulaarne massaaž	0,5699
Regulaarne venitamine	0,5951
Vanus	0,8359
Pikkus	0,5345
Käe pikkus	0,1875
Vanus trennidega alustades	0,0201
Jõusaalitrennide arv nädalas	0,9959
Pallitrennide arv nädalas	0,0793

Näeme, et kaheksa erineva tunnuse seas on vaid ühe korral gruppide võrdluses kõverad erinevad. Antud tunnuseks on vanus võrkpallitreeningutega alustades, mille korral jäi p-väärtus ainsana väiksemaks, kui olulisuse nivoo  $\alpha = 0,05$ .

Vaatleme jooniselt, kuidas erinevad omavahel kuni 9-aastaselt alustanute ja pärast 9-ndat eluaastat alustanud mängijate kohta käivad hinnangulised elukestuskõverad. Välja on toodud vaid antud tunnus, kuna see oli ainus, mille gruppide kõveratel sai tõestada erinevust.



Joonis 3. Treeningutega alustamise vanusgruppide elukestuskõverate hinnangud.

Näeme jooniselt, et enne kümnendat eluaastat alustanute seas on tsenseeritud andmetega mängijaid hulgaliselt rohkem kui teises grupis. Samuti näeme, et esimese grupi korral on kõver vaid astme võrra madalamale langenud, mistõttu püsib vigastusest pääsemise tõenäosus üsna kõrge. Vastupidiselt nooremale grupile on vanemas eas alustanute grupis vigastusi rohkem esinenud ning ka hinnanguline kõver on madalamale langenud ehk vigastuse esinemise tõenäosuse hinnang on tõusnud üpris kõrgeks.

Antud tunnus võis osutada oluliseks seetõttu, et mängijad, kes alustasid trenniga nooremas eas, said individuaalsemat ja põhjalikumat baastreeningut kui need, kes alustasid hilisemas vanuses. Hilisemas etapis on raskem treenida lapsi, kes on väga ebaühtlase teadmiste ja oskuste taustaga ning seetõttu võivad kohati algajad jääda tagaplaanile ning sooritavad löögiliigutust valesti.



## 2.4 Mitmese testimise probleem

Statistiliste testide korral kummutatakse nullhüpotees, kui seose saamise tõenäosus jääb alla määratud olulisuse nivoo  $\alpha$ . Kui aga on tarvis kontrollida mitut hüpoteeside paari, siis puututakse kokku mitmese testimise probleemiga.

Kui uuringu käigus viiakse läbi  $k$  sõltumatut testi ühe ja sama olulisuse nivoo  $\alpha$ , siis tõenäosus teha viga vähemalt ühe hüpoteesipaari korral on

$$P(\text{vähemalt ühe hüpoteesipaari korral tehti I liiki viga}) = 1 - (1 - \alpha)^k.$$

Näiteks 8 sõltumatu testi korral, millel igal ühel on olulisuse nivooks  $\alpha = 0,05$ , saame vähemalt ühe valepositiivse testi saamise tõenäosuseks ligikaudu 33,7%. [2]

Lahenduseks oleks kasutada Bonferroni meetodit. Antud meetod seisneb  $k$  üksikvõrdluse puhul ühe või enama vea tegemise tõenäosuse piiramist olulisuse nivoo  $\alpha$ . Selleks tuleb üksikvõrdlustel võtta uueks olulisuse nivooks  $\alpha^* = \alpha/k$ . See tähendab, et antud meetodi puhul saab tõestada sisukat hüpoteesi vaid juhul, kui testi  $p$ -väärtus on väiksem kui  $\alpha^*$ . [6]

Toome uuesti välja *log-rank* testi tulemustega tabeli, kuhu on lisatud ka Bonferroni meetodil korrigeeritud  $p$ -väärtused,  $p_{Bonferroni} = \min\{p\text{-väärtus} \cdot k; 1\}$  (tabel 4).

Tabel 4. *Log-rank* testi tulemused koos Bonferroni meetodiga korrigeeritud tulemustega.

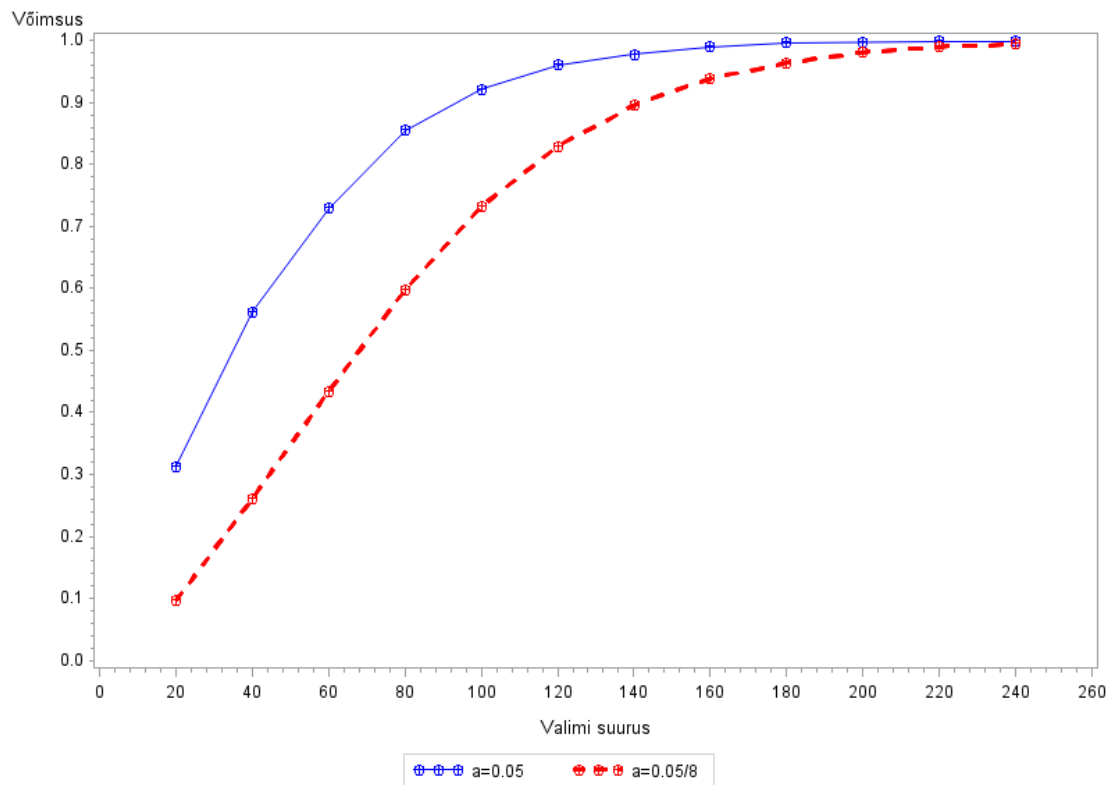
Tunnus	p-väärtus	$p_{Bonferroni}$
Regulaarne massaaž	0,5699	1
Regulaarne venitamine	0,5951	1
Vanus	0,8359	1
Pikkus	0,5345	1
Käe pikkus	0,1875	1
Vanus trennidega alustades	0,0201	0,1608
Jõusaalitrennide arv nädalas	0,9959	1
Pallitrennide arv nädalas	0,0793	0,6344

Näeme, et korrigeeritud p-väärtustest ei ole ükski tulemus väiksem kui tavaliselt kasutatavast olulisuse nivoost 0,05. Seetõttu oleks paslik mõelda muudatustele uuringu läbiviimisel. Selle jaoks uurime testi võimsust.

## 2.5 Testi võimsus

Väikese valimiga uuringute korral ei pruugi tulemused olla piisavalt täpsed ega anda uurijale palju infot. Kui tõenäoline on näha statistiliselt olulist tulemust, seda kirjeldab testi võimsus. Kas meie valim oli piisavalt suur avastamiseks alternatiivset hüpoteesi, kui see oleks tegelikult kehtinud? Leidsime testi võimsuse erineva suurustega valimite korral nii üksiktesti ( $\alpha = 0,05$ ) kui ka kõigi kaheksa testi ja Bonferroni meetodi kasutamise korral (üksiktesti olulisuse nivoo  $\alpha = 0,05/8$ ). Testi võimsuse leidsime simulatsioonide abil (lisa 4).

### Mitmese testimise probleemi graafik

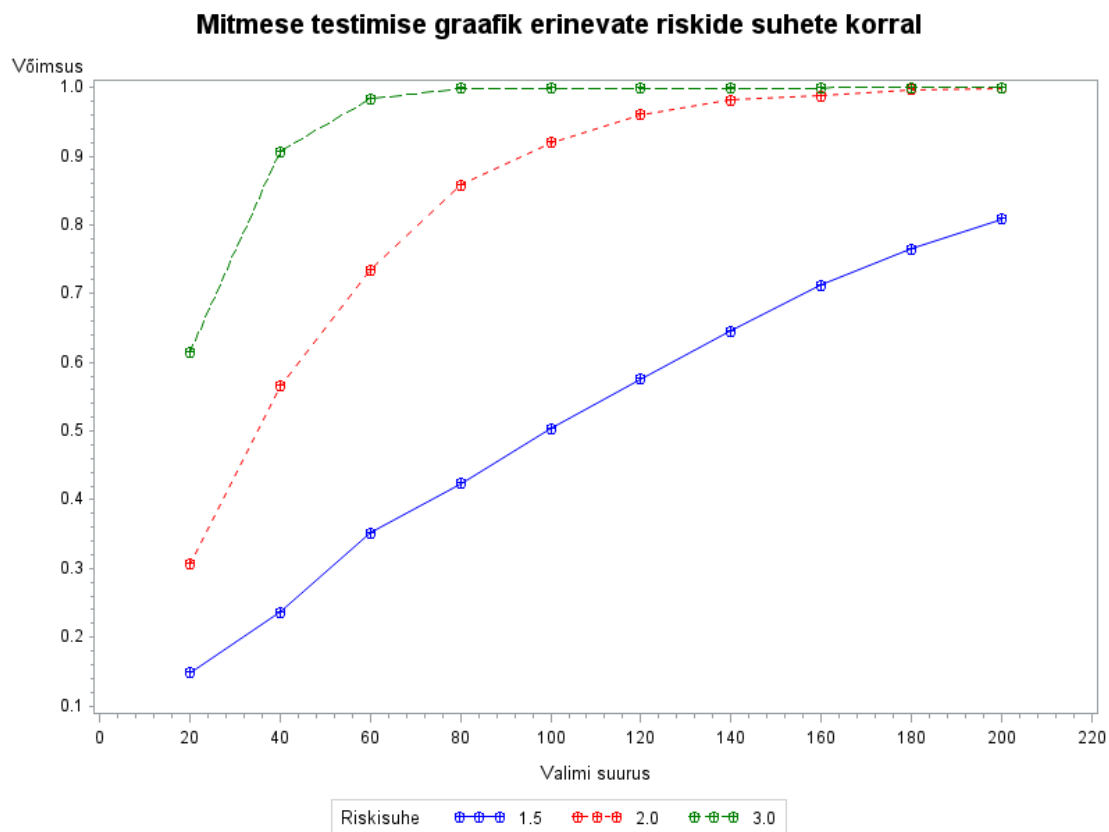


Joonis 4. Valimi suurus ja võimsus, et tuvastada kahekordset riskide erinevust.

Joonisel 4 on kujutatud graafik, kus on näha valimi suurus ja võimsus nii algse  $\alpha$  kui ka Bonferroni meetodi korral. Mainime, et joonisel on kasutatud riskisuhtet suurusega 2, mis tähendab, et uuritava sündmuse toimumise sagedus on ühes grupis 2 korda suurem kui teises grupis. Näeme, et Bonferroni meetodi korral on suure võimsuse saavutamiseks vaja suuremat valimimahtu kui üksiktesti korral. Alles valimi suurusega 200 korral hakkavad võimsused ühtlustuma.

Kui vaatame joonist, siis näeme, et kui valimi suurus on 20, siis Bonferroni meetodi korral on võimsus vaid umbkaudu 0,1. See tähendab, et kui meil on kaks gruppi, kus ühes on sündmuse toimumise sagedus kaks korda suurem, siis vaid tõenäosusega 0,1 tuvastab test gruppide erinevuse. Kuna aga antud tõenäosus on väga väike, siis tuleb püüelda suurema testi võimsuse poole, mis võiks olla vähemalt 0,8.

Toome joonise ka mitmese testimise kohta, milles on kasutatud erinevaid riskisuhteid (joonis 5).



Joonis 5. Valimi suuruse ja võimsuse seos erinevate riskisuhete korral.

Antud joonisel on näha, et riskisuhe mõjutab oluliselt testi võimsust. Kõige kiiremini kasvab võimsus valimi suurenedes, kui riskisuhe on 3 ning kõige kehvemini, kui suhe on 1,5. Seega, kui riskisuhe on 3, võib valimi suurus olla üsna väike, isegi vaid 40, et võimsus ulatuks juba suuruseni 0,9. See aga tähendab, et valimis peaks olema 3 korda rohkem inimesi, kellel on sündmus toimunud, kui on sündmuseta inimesi. Võrkpallurite korral peaks 40 palluri uurimisel 30-l olema õlavõõtme vigastus toimunud ning vastukaaluks 10 tervet mängijat.

Sportlaste uurimisel on üsna raske saavutada riskisuhet suurusega 3, seega tuleks keskenduda pigem väiksele riskisuhetele ning sellega kaasnevale suurele valimile. Sellise meetodi korral on küll töömaht suurem, kuid tulemus on kindlamini saavutatav.

## 2.6 Coxi mudel

Toome välja tabeli (tabel 5), kus on kahe tunnuse kohta käivad Coxi mudelist saadud p-väärtused, seal hulgas on pallitrennide arvu nädalas uuritud nii pideva

kui binaarse tunnusena. Ülejäänud tunnused jätsime uurimata, kuna nende korral olid *log-rank* testi p-väärtused juba küllaltki suured, et muud testid nende olulisust näidata võiks.

Tabel 5. Coxi mudeli tulemused.

Tunnus	p-väärtus
Vanus trennidega alustades	0,0088
Pallitrennide arv nädalas (0/1 tunnus)	0,0825
Pallitrennide arv nädalas (pidev tunnus)	0,0062

Sarnaselt *log-rank* testile tuli ka Coxi mudeliga tunnus vanus trennidega alustades oluline ehk see mõjutab õlavöötme vigastuse teket. Kuna Coxi mudel toimib ka pideva tunnuse korral, siis vaatasime tunnust pallitrennide arv nädalas nii binaarse kui pideva tunnusena. Nagu tulemustest näha, siis on pallitrennide arv nädalas pideva tunnusena statistiliselt oluline, kuid 0/1 tunnusena jääb siiski antud valimi korral olulisuse nivoost suuremaks.

Bonferroni korrektsiooni korral saame pideva tunnuse pallitrennide arv nädalas samuti oluliseks, p-väärtuseks on 0,0496, seevastu ülejäänud kaks välja toodud tunnust Bonferroni korral oluliseks ei osutu. Sellise tulemuse saame korrutades Coxi mudeli p-väärtust suurusega 8. Kuna meil oli algselt erinevaid teste just 8, siis korrutame ka seekord p-väärtused sama suurusega, olenemata sellest, et välja tõime vaid kahe erineva tunnuse tulemused.

## Kokkuvõte

Antud bakalaureusetöös käsitlesime Eesti meesvõrkpallurite andmeid ning analüüsisime nende õlavöötme vigastuse teket elukestusanalüüsi abil, kaasates Kaplan-Meieri meetodit, *log-rank* testi ning Coxi mudelit. Samuti tõime töös välja testi võimsusega seonduvad probleemid.

Analüüsi läbi viies saime teada, et *log-rank* testi puhul on ainus oluline tunnus võrkpallurite õlavöötme vigastuse tekkimisele vanus võrkpallitreeningutega alustades. Täpsemalt, kui mängija alustas treeningutega varasemas eas, enne kümnendat eluaastat, siis on tõenäosus saada vigastus väiksem, kui hilisemas eas treeningutega alustades. Sealjuures uurisime, et kui korrigeerida p-väärtuseid Bonferroni meetodil, siis ei tohiks ükski tunnus pallurite õlavigastuse teket mõjutada. Coxi mudeli korral aga saime, et nii vanus trennidega alustades kui ka pallitrennide arv nädalas pideva tunnusena võivad mõjutada õlavigastuse tekkeriski.

Analüüsi teises osas vaatlesime ka testi võimsusega seonduvaid probleeme, millest tuli välja, et kui  $\alpha = 0,05$ , siis võimsuse üle 0,8 saavutame valimiga, mille suurus on 80 mängijat ning riskisuhtega 2. Vaatlesime mitmest testimist ka erinevate riskisuhete korral, millest tuli välja, et kui riskisuhe on kõrge, siis võib valim olla üsna väike. Kuna aga selle saavutamine on küllaltki keeruline, siis madala riskisuhte korral peaks olema valimi suurus umbkaudu 200 ning riskisuhe vähemalt 1,5. Seetõttu oleks kindlasti hea, kui uuringut teostataks uuesti suurema valimiga, et võrkpalluritele paremat tagasisidet saaks anda.

## Kasutatud kirjandus

- [1] Kleinbaum, D. G., Klein, M. (2005). *Survival Analysis: A Self-Learning Text, Second Edition*.
- [2] Koppel, K. „PheWAS ja selle praktiline läbiviimine Tartu Ülikooli Eesti Geenivaramu andmete põhjal”
- [3] Käärrik, E. (2016) *Andmeanalüüs II*. Loengukonspekt.
- [4] Lane, David M „Factors Affecting Power”. Kättesaadav aadressil:  
<http://onlinestatbook.com/2/power/factors.html>
- [5] Lepik, N. (2016). *Tõenäosusteooria ja matemaatiline statistika I*. Loengukonspekt
- [6] Möls, M. (2017) *Biostatistika*. Loengukonspekt.
- [7] Rodriguez, G. (2010). *Survival Models*. Kättesaadav aadressil:  
<http://data.princeton.edu/wws509/notes/c7.pdf>
- [8] SAS/STAT(R) 9.3 User's Guide

# Lisad

## Lisa 1. Küsitlus võrkpalluritele

Tere! Olen Tartu Ülikooli matemaatilise statistika eriala bakalaureuseõppe tudeng ning soovin antud küsitlusega välja selgitada, millest võivad sõltuda õlavöötme vigastused ning kuidas edaspidi vigastust ennetada.

1. Nimi:
2. Sünniaeg:
3. Pikkus:
4. Domineeriva käe pikkus:
5. Mis vanuses alustasite võrkpallitreeninguid?
6. Mitu korda nädalas toimuvad jõusaalitreeningud?
7. Mitu korda nädalas toimuvad pallitreeningud?
8. Kas venitate lihaseid treeningu alguses ja lõpus regulaarselt?
9. Kas saate regulaarselt lihasmassaaže?
10. Kas Teil on esinenud õlavöötme vigastusi?

Kui vastus viimasele küsimusele oli „ei“, siis järgnevatele küsimustele Te vastama ei pea. Kui vastus oli „jah“, siis vastake järgnevatele küsimustele esmasele vigastusele vahetult eelnenud ajaperioodi kohta.

11. Mis vanuses tekkis Teil esimest korda õlavöötme vigastus?
12. Mitu korda nädalas toimusid jõusaalitreeningud?
13. Mitu korda nädalas toimusid pallitreeningud?
14. Kas venitasite lihaseid treeningu alguses ja lõpus regulaarselt?
15. Kas saite regulaarselt lihasmassaaže?

Täna vastamast! Jõudu treeningutel!



**Lisa 2.** SAS kood *log-rank* testile ja elukestusfunktsiooni hinnangule

```
proc lifetest data=andmed1 plots=(s);
time vahe*vigastus(0);
strata joukas;
run;
```

**Lisa 3.** SAS kood Coxi mudelile

```
proc phreg data=andmed1;
model vahe*vigastus(0)=mass_e;
run;
```

**Lisa 4.** Testi võimsuse leidmine simulatsiooni abil

```
data andmed;
vigastus=1;
do a = 1 to 10;
    do j = 1 to 5000;
        n_a=10*a;
        n_b=10*a;
        valimisuurus=n_a+n_b;
        hr=2;
        do i = 1 to n_a+n_b;
            p1=(hr*n_a)/(n_a*hr+n_b);
            x=ranuni(-90);
            if x < p1 then do; y='A'; n_a=n_a
                -1; end; else do; y='B'; n_b=n_b
                -1; end;
            output;
        end;
    end;
end;
run;

proc lifetest data=andmed;
time i*vigastus(0);
strata y;
by a j;
ods output HomTests=minuasjad1;
run;

data minuasjad_hr;
merge minuasjad1 minuasjad2 minuasjad3;
```

```

    by hr;
run;

data minuasjad_hr;
set minuasjad_hr;
if ProbChiSq<0.05 then u=1;
else u=0;
run;

proc sql;
create table tulemus as select mean(u) as v6imsus, 20*a as
    valimisuurus, hr as hr from minuasjad_hr group a, hr;
quit;

symbol1 interpol=join value=+ color=blue line=1;
symbol2 interpol=join value=+ color=red line=2;
symbol3 interpol=join value=+ color=green line=3;
axis1 order=(0 to 220 by 20);
title "Mitmese testimise graafik erinevate riskide suhete
    korral";
legend1 frame;
proc gplot data=tulemus;
plot v6imsus*valimisuurus=hr/overlay haxis=axis1 legend=
    legend1;
label v6imsus="Võimsus" valimisuurus="Valimi suurus" hr="
    Riskisuhe ";
run;
quit;

```

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Eneli Maal (sünnikuupäev 02.06.1995)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Võrkpallurite õlavõtmevigastuste esinemisriski uurimine elukestusanalüüsi abil", mille juhendaja on Märt Möls,
  - i. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
  - ii. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **13.02.2018**