

Peatükk 5

Diagnostilised testid I.

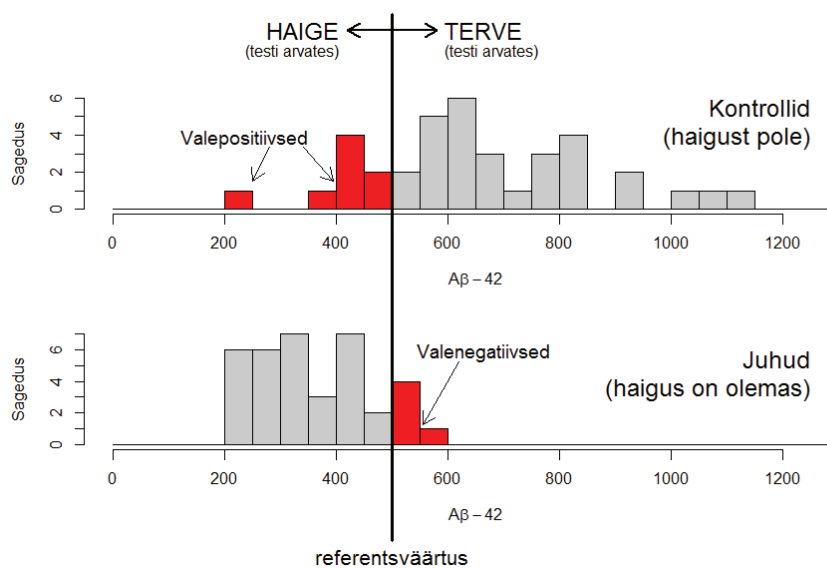
Arstidel on vaja aru saada, kas patsientidel on otsitav haigus või mitte. Diagnostilise paneku kasutamisel kasutatakse sageli diagnostiliste testide abil. Näiteks tehakse vereproov ja vaadatakse, kas mingi verenäitaja väärtus on väiksem referentsväärtusest (või normväärtusest) või mitte. Kui on väiksem lubatud normist siis öeldakse, et test andis positiivse testitulemuse — patsiendil on otsitav haigus. Mõne näitaja korral loetakse muidugi haigeteks hoopis neid inimesi, kelle vereproovi tulemus ületab referentsnäitu.

Meie jaoks on praegu oluline hoopis see, et kõik diagnostilised testid võivad eksida. Vahel arvatakse ekslikult, et inimesel kellel otsitavat haigust tegelikult pole, on see haigus. Saadakse nn valepositiivne tulemus — test näitab haiguse olemasolu, aga päriselt seda haigust antud inimesel pole.

Samuti võidakse vahel saada valenegatiivseid tulemusi. Valenegatiivne testitulemus tekib siis, kui testi arvates peaks inimene terve olema, aga tegelikult on uuritava inimesel siiski otsitav haigus olemas, lihtsalt tehtud test ei suuda seda haigust üles leida. Vaata ka joonist 5.1.

Märkus: andmed sellel joonisel kirjeldavad ühte esimest katset diagnoosida Alzheimeri tõbe tavalise vereproovi abil. Tavaliselt tuleb Alzheimeri diagnoosimiseks võtta ajuvedelikust proov (pistes nõela läbi kolju) või, hädapärast, kõlbab ka magnetresonantstomograafia (mis on ka ebameeldivalt kalline protseduur). Sestap üritatakse leida võimalusi Alzheimeri tõbe diagnoosimiseks tavalise vereproovi abil. Siin proovitakse kasutada amüloid-beeta ($A\beta$) peptiide — määratud tavalise vereproovis — otsustamiseks, kas inimesel on Alzheimer (või on tegemist mingi muu haiguse või probleemiga). Joonisel kujutatud andmed on võetud artiklist Nabers *et al.*, 2018.

Joonis 5.1: Valepositiivsed ja valenegatiivsed testi tulemused. Alzheimeri testimine tavalise vereproovi abil.



5.1 Tundlikkus ja spetsiifilisus

Arstidel on oluline teada, kui sageli diagnostiline test võib eksida. Selle pärast peab iga meditsiinis kasutatava testi jaoks olema teada testi spetsiifilisus ja tundlikkus.

Testi **tundlikkus** (*sensitivity*) on tõenäosus, et test annab testitava haiguse või seisundi olemasolu korral positiivse tulemuse:

$$\begin{aligned}
 \text{tundlikkus} &= P(\text{haigel leitakse haigus}) \\
 &= P(T+ | H+) \\
 &\approx \frac{\text{haigete arv kellel leiti haigus}}{\text{uuritud haigete arv}},
 \end{aligned}$$

Kus $T+$ tähistab positiivset testitulemust ja $H+$ tähistab sündmust, et uuritaval on tegelikult ka otsitav haigus olemas.

Testi **spetsiifilisus** (*specificity*) on tõenäosus, et test annab testitava seisundi puudumisel negatiivse tulemuse:

$$\begin{aligned} \text{spetsiifilisus} &= P(\text{terve loetakse terveks}) \\ &= P(T - |H-) \\ &\approx \frac{\text{negatiivse testitulemuse saanud tervete arv}}{\text{uuritud tervete arv}}. \end{aligned}$$

Mõningate arstide poolt kasutatavate diagnostiliste testide tundlikkused ja spetsiifilisused on ära toodud tabelis 5.1. Tabel on võetud (Maxim, Niebo ja Utell, 2014).

Tabel 5.1: Mõningate meditsiinis kasutatavate testide tundlikkus ja spetsiifilisus, (Maxim, Niebo ja Utell, 2014).

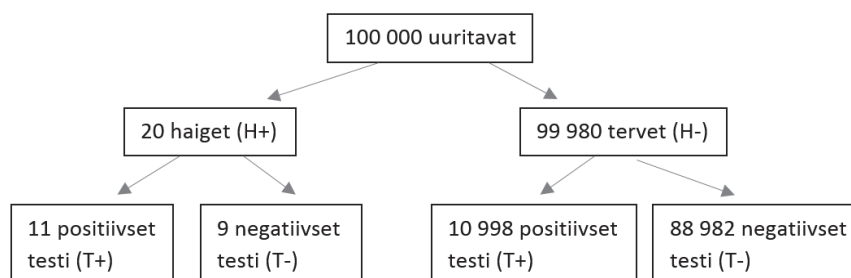
haigus	autor	spetsiifilisus	tundlikkus
Klamüüdia, PCR, emakakael	Mayrand	1	0,965
Klamüüdia, LCR, uriin	Mayrand	1	0,875
Eesnäärmevähk, PSA verest	Catalona	0,59	0,79
Eesnäärmevähk, ultraheli	Catalona	0,27	0,92
Eesnäärmevähk, käega katsumine	Catalona	0,44	0,86
Rinnavähk, mammograaf	Rafferty	0,841	0,655
Rinnavähk, tomosynthesis	Rafferty	0,862	0,627
HIV, ELISA antikehade test	Weiss	0,986	0,973

5.2 Positiivne prognoosiväärtus, PPV

Alustame ühe näitega, mis aitab ehk seletada miks testi spetsiifilisuse ja tundlikkuse jälgimine võib oluline olla.

Kujutame endale ette järgmist olukorda. Eestis on jätkuvalt probleemiks tuberkuloos. Tuberkuloosi esineb küll märksa vähem kui varem, aga jätkuvalt igal aastal haigestub uusi inimesi. Küll oleks tore, kui saaksime tuberkuloosist Eestis lõplikult lahti! Kuidas seda võiks saavutada? Aga näiteks nii: testime kõiki inimesi tuberkuloosi vastu. Haiged ravime terveks, ja olemegi sellest vastikust haigusest lõplikult lahti! Aga kuidas seda teha? Vajame lihtsat ja odavat testi (peame testima ikkagi üle miljoni inimesel), millega saaksime kõiki inimesi testida! Hinna poolest sobiks näiteks PATHOZYME-MYCO IgG, mille tundlikkus on 0,55 ja spetsiifilisus 0,89

(võrdluseks röntgeni võime tuberkuloosi tuvastada: tundlikkus 0,67..0,66; spetsiifilisus: 0,66..0,76). Kuidas selline test Eesti elanike testimisel töötaks? Eestis võiks tuberkuloosi levimus olla umbes 20 juhtu 100 000 inimese kohta. Mis juhtub keskmiselt 100 000 testitavaga, seda kirjeldab alljärgnev skeem:



Näeme, et keskmiselt $(1 - \text{tundlikkus}) * 100000$ haiget jääb meil avastamata. Aga isegi suurem probleem tekib nendega, kellel test annab positiivse testitulemuse. Meil on 11 009 positiivse testitulemuse saanud inimest. Enamik neist on terved, haigete osakaal positiivse testitulemuse saanute seas on väiksem kui 0,001. Peaaegu kindlasti on positiivse testitulemuse saanud inimene terve. Ja kui tahaksime kindluse mõttes neid kõiki kopsuarsti vastuvõtule saata, siis kuluks praeguste võimaluste juures kümme aastat, enne kui nad kõik läbivaadatud/uuritud saaksid. Seega antud diagnostiline test kogu populatsiooni uurimiseks kohe kindlasti ei sobi.

Kaks väga olulist näitajat diagnostiliste testide korral ongi testi testi positiivne prognoosiväärtus $PPV = P(H+|T+)$ (kui tõenäoliselt on positiivse testitulemuse saanud inimene ka tegelikult haige) ja negatiivne prognoosiväärtus $NPV = P(H-|T-)$ (kui tõenäoliselt on negatiivse testitulemuse saanud inimene tegelikult terve). Kui teame testi spetsiifilisust, tundlikkust ja haiguse levimust, siis saame alati testi positiivse prognoosiväärtuse arvutada, kasutades Bayes'i valemit:

$$P(H+|T+) = \frac{P(T+|H+)P(H+)}{P(T+|H+)P(H+) + P(T+|H-)P(H-)}$$

$$PPV = \frac{\text{tundlikkus} \cdot \text{levimus}}{\text{tundlikkus} \cdot \text{levimus} + (1 - \text{spetsiifilisus})(1 - \text{levimus})}$$

ja samuti võime leida negatiivse prognoosiväärtuse:

$$P(H- | T-) = \frac{P(T- | H-)P(H-)}{P(T- | H-)P(H-) + P(T- | H+)P(H+)}$$

$$NPV = \frac{\textit{spetsiifilisus} \cdot (1 - \textit{levimus})}{\textit{spetsiifilisus} \cdot (1 - \textit{levimus}) + (1 - \textit{tundlikkus}) \cdot \textit{levimus}}$$

Testi positiivne ja negatiivne prognoosiväärtus on kliinilises praktikas enamasti märksa olulisemad ja mugavimini kasutatavad näitajad kui tundlikkus ja spetsiifilisus. Miks siis raporteeritakse diagnostiliste testide puhul enamasti testi tundlikkust ja spetsiifilisust, aga mitte positiivset prognoosiväärtust? Põhjus on lihtne — tundlikkus ja spetsiifilisus jäävad enam-vähem samaks, ükskõik kelle peal või kuidas me seda testi kasutame (raskeid haigusjuhte võib olla muidugi kergem tuvastada kui kergeid ja alles algusjärgus haiguseid). Aga testi positiivne ja negatiivne prognoosiväärtus kõiguvad märksa rohkem sõltuvalt testi kasutamise ajast ja sellest, keda me just parasjagu testime. Võtame näiteks ühe gripi kiirtesti (võimalik saada teada testi tulemus koheselt arsti visiidi ajal) Directigen Flu A+B. Selle testi tundlikkus on B-tüüpi gripi tuvastamisel 88% ja spetsiifilisus 97% ((Organization *et al.*, 2005)). Mõlemad numbrid jäävad samaks, ükskõik kas kasutame seda testi gripihooaja alguses või haripunktis. Aga kui me ülemiste hingamisteede ägeda infektsiooniga arsti poole pöördujale teeksime selle testi 10. oktoobril 2019, siis oleks selle testi positiivne prognoosiväärtus umbes 0,03 (peaaegu kindlasti pole positiivse testitulemuse saanud inimesel grippi). Kui sama testi kasutaksime 20. veebruaril 2020, siis oleks sellesama testi positiivne prognoosiväärtus 83% (üsna kindlasti on positiivse testitulemuse saanud inimesele gripp). Erinevus ei ole tingitud testi enda probleemidest, vaid tuleneb sellest, et gripi levimus muutus. Kui me ei testiks iga kõha või nohuga arsti poole pöördijat, vaid nõuaksime täiendavalt mõne gripile omase sümptomi olemasolu, siis tõuseks antud testi positiivne prognoosiväärtus veelgi. Informatsioon gripi levimuse kohta on võetud terviseameti grippi ja gripilaadsete viirusnakkustesse haigestumisele pühendatud veebilehelt.

5.3 Levimuse hindamine ebatäpsete testide korral

Haiguste levimuse hindamiseks ja jälgimiseks (kuidas levimus ajas muutub?) tuleb sageli teha palju diagnostilisi teste. Paljude testide tegemisel on aga kulude kokkuhoid sageli oluline. Kulude kokkuhoidmise vajadus sunnib uurijaid sageli kasutama ebatäpseid teste — mille spetsiifilisus ja tundlikkus on väiksemad kui 100% (päriselus vist ei eksisteerigi sellist testi — kasvõi proovivõtja võimaliku inimliku vea tõttu jäävad kõigi testide näitajad alla 100%).

Kuidas siis ebatäpseid teste kasutades haiguse levimust ikkagi adekvaatselt hinnata?

Appi tuleb täistõenäosuse valem:

$$\begin{aligned} P(T+) &= P(T+|H+)P(H+) + P(T+|H-)P(H-) \\ P(T+) &= \text{tundlikkus} \cdot \text{levimus} + (1 - \text{spetsiifilisus})(1 - \text{levimus}) \\ \text{levimus} &= \frac{P(T+) - 1 + \text{spetsiifilisus}}{\text{tundlikkus} - 1 + \text{spetsiifilisus}} \end{aligned} \quad (5.1)$$

Seega kui kasutame testi mille spetsiifilisus ja tundlikkus on 0,9 haiguse levimuse uurimiseks ja saame, et 12% uuritavatest saavad positiivse testitulemuse, siis on haiguse tegelik levimus $(0,12 - 1 - 0,9)/(0,9 - 1 + 0,9) = 0,025$ ehk sellisel juhul on 2,5% inimestest tegelikult uuritav haigus.

5.4 Hinnangute täpsusest

Ükski hinnang pole kunagi täpne. Nii testi tundlikkust kui spetsiifilisust tuleb andmete pealt hinnata. Kõigi hinnangute täpsust on aga viisakas kirjeldada — kasvõi näiteks usaldusintervalli abil.

Kuidas hinnatakse testi tundlikkust ja spetsiifilisust? Leitakse kõigepealt mingi arv inimesi, kes on tegelikult haiged (testi tundlikkuse hindamiseks) ja mõned inimesed, kellel haigust pole (testi spetsiifilisuse hindamiseks). Kuidas neid leida? Vahel on juba olemas mingi test, mida saame usaldada (mis võib aga olla inimeste kliinilises ravis kasutamiseks kas liiga aeglane — testi tulemus selgub nädala pärast, aga raviotsuse tegemiseks peaksime teadma kohe; täpne test võib olla ka liiga kallis kas mõõdetuna rahas või patsientide kannatuste mõttes — näiteks tõe võime teada saada inimesi operatsioonilaua lahti lõigates). Vahel otsitakse uuringusse näiteks inimesi, kellel haiguse olemasolu või puudumist saab mingitel muudel kaudsetel viisidel tuletada. Näiteks rasedustestide spetsiifilisuse hindamisel on vahel kasutatud nunnade abi. Seejärel leitakse hinnangud lihtsal viisil: kui suur osa tegelikult haigetest said testiga positiivse tulemuse (tundlikkuse hinnang)? Kui suur osa tegelikult tervetest said testiga negatiivse tulemuse (spetsiifilisuse hinnang)?

Hinnangute täpsuse kirjeldamiseks piisab mõistmisest, et hindame tõenäosust. Kui valim on suur ja tundlikkuse/spetsiifilisuse väärtused pole liiga lähedal 1-le, siis on sündmuse toimumise suhteline sagedus ligilähedaselt normaaljaotusega ja me võime kasutada ligikaudset usaldusintervalli valemit (95%-usaldusintervall tõenäosusele p kasutades tõenäosuse hinnangut

\hat{p} :

$$\hat{p} \mp 1,96\sqrt{\hat{p}(1 - \hat{p})/n},$$

Kus n on tõenäosuse hindamiseks kasutatud vaatluste arv — spetsiifilisuse hindamisel siis tegelikult tervete arv, tundlikkuse hindamisel tegelike haigete arv. Kui soovime kasutada teistsugust usaldusnivood (tahame näiteks leida $(1 - \alpha)$ -usaldusintervalli) siis peaksime numbri 1,96 asendama mõne teise sobiva normaaljaotuse kvantiiliga ($\alpha/2$ -kvantiiliga).

Võimaluse korral tasuks muidugi eelistada täpset usaldusintervalli tõenäosusele, mida näiteks R-is saab leida käsuga `binom.test`. Kui uuriti näiteks 100 tegelikku nakatunut, kellest 96-l leidis test ka haiguse üles, siis ligikaudne 95%-usaldusintervall on (0,921...0,998), täpne 95%-usaldusintervall on aga (0,901...0,989). Täpse usaldusintervalli saame leida nii:

```
> binom.test(96,100)

      Exact binomial test

data: 96 and 100
number of successes = 96, number of trials = 100,
      p-value < 2.2e-16
alternative hypothesis: true probability of
      success is not equal to 0.5
95 percent confidence interval:
 0.9007428 0.9889955
sample estimates:
probability of success
      0.96
```

Ka levimuse hinnangule (vaata peatükki 5.3) tuleb vahel usaldusintervall leida. Kui saame spetsiifilisuse ja tundlikkuse väärtused täpseks lugeda (nende väärtuste väljaselgitamiseks on uuritud väga palju tegelikult terveid ja tegelikult haigeid inimesi. Vähemalt palju rohkem kui meil on inimesi oma valimis), siis võib haiguse tegeliku levimuse hindamiseks kasutatava valemi 5.1 tulemus vigane tulla vaid sellepärast, et positiivse testitulemuse saamise tõenäosus on valesti hinnatud. Sellisel juhul võime leida usalduspiirid positiivse testitulemuse saamise tõenäosusele ja panna saadud usalduspiirid valemisse 5.1. Tulemuseks saame usalduspiirid haiguse tegelikule levimusele.

Kui spetsiifilisuse ja tundlikkuse hinnangud võivad olla suhteliselt ebapäpset (nende hindamiseks kasutatud valimite suurused pole suurusjärke

suuremad meie valimist mida kasutame levimuse hindamiseks), siis võib usaldusintervalli leidmisel olla kasu näiteks simulatsioonidest või Monte-Carlo meetoditest. Näiteks kui kõik kasutatud valimid (tundlikkuse, spetsiifilisuse ja levimuse hindamiseks kasutatud valimid) on piisavalt suured selleks, et hinnangute jaotus võiks juba läheneda normaaljaotusele ($n \cdot \hat{p} > 5$ ja $n \cdot (1 - \hat{p}) > 5$, siis võime usaldusintervalli haiguse tegelikule levimusele leida näiteks järgmise simulatsiooni abil:

```
# Testi tundlikkuse hinnang ja tundlikkuse hindamiseks kasutatud
# tegelikult haigete arv
tundlikkus=0.9
n_tundlikkus=934

# Testi spetsiifilisuse hinnang ja spetsiifilisuse hindamiseks
# kasutatud tegelikult tervete arv
spetsiifilisus=0.9
n_spetsiifilisus=1478

# Levimuse hindamine. Positiivsete testitulemuste osakaal ja testitute
# koguarv meie populatsioonis
pos_testi_osakaal=0.12
n_teste=420

# Simulatsioon:
abi_tund=rnorm(10000, mean=tundlikkus,
  sd=sqrt(tundlikkus*(1-tundlikkus)/n_tundlikkus))
abi_spets=rnorm(10000, mean=spetsiifilisus,
  sd=sqrt(spetsiifilisus*(1-spetsiifilisus)/n_spetsiifilisus))
abi_pos_test=rnorm(10000, mean=pos_testi_osakaal,
  sd=sqrt(pos_testi_osakaal*(1-pos_testi_osakaal)/n_teste))

# Arvutame hinnangud levimusele
levimused=(abi_pos_test-1+abi_spets)/(abi_tund-1+abi_spets)

# Leiame 0,95-usaldusintervalli tegelikule levimusele:
intervall=quantile(levimused, c(0.025, 0.975))
intervall
```

Kui kasutada ligikaudseid meetodeid usaldusintervallide leidmiseks, siis võib mõni usalduspiiridest (tundlikkusele/spetsiifilisusele/levimusele) tulla

kas negatiivne (alumine piir) või suurem ühest (ülemine piir). Ka näites toodud simulatsiooniprogramm tagastab negatiivse alumise usalduspiiri väärtuse. Aga tõenäosused ei saa olla ei negatiivsed ega ühest suuremad. Negatiivse alumise usalduspiiri korral on kombeks asendada see 0-ga; ka ühest suurem ülemine usalduspiir asendatakse tavaliselt 1-ga.

Ülesanne

Paari aasta pärast on elu peale COVID-19 paanikat tagasi pöödnud tava-pärasesse rütmi. Oled edukalt lõpetanud ülikooli ja noore andeka spetsialistina kutsutakse sind osalema rahataotluste hindamisel. Pandeemiajärgne majanduskriis on ju juba unustatud ning raha jätkub ka muudeks vajalikeks ja kasulikeks tegemisteks! Paraku pole raha kunagi piisavalt, et seda rumalustele kulutada. Leia järgmisest kahest ettepanekutest üles nõrk koht — midagi, mida suudab tuvastada eelkõige hea matemaatika/statistikaharidusega inimene — ja too see välja (nii et ka mittematemaatik/mittestatistik aru saaks, mis on viltu)! Samas põhjenda oma arvamust asjakohaste ja täpsete arvutustega!

1. Paar aastat on Botswanas ja Lõuna-Aafrika Vabariigis massiliselt tehtud HIV-teste kasutades suhteliselt odavat Quixi kiirtesti (tundlikkus 96,2%; spetsiifilisus 98,9%). Mõlemas riigis on massilise testimise programme ka väliseksperptide poolt hinnatud ja tõhusaks kuulutatud.

Suhteliselt vaeses Laoses on naaberriigis Tais leviva HIV-epideemia tõttu tekkinud AIDS-i paanika ja poliitikud on huvitatud oma elanikkonna massilisest testimisest. Odava Quixi kiirtesti abil oleks see ka rahaliselt põhimõtteliselt mõeldav (kui ka Eesti riik õige-tsipakene arenguabi annaks). Kuna Eesti riigil on järsku tekkinud soov Laosega sõbraks saada, siis oleks mõne hästi valitud (rahaliselt odava ja palju positiivset vastukaja tekitava) abiprojekti toetamine Eesti poolt igati mõeldav.

Teilt kui haritud (ülikoolis käinud) inimeselt küsitakse igaks juhuks ka arvamust – kas tasub antud projekti toetada või mitte. Põhjendage oma arvamust (ülemused ilmselt loodavad, et ehk tulete mõne hea väitega välja, mida nad ise saaksid projekti läbisurumisel ära kasutada).

Vihjeks: ülesande lahendamisel võib abi olla näiteks sellisest veebilehest:

https://en.wikipedia.org/wiki/List_of_countries_by_HIV/AIDS_adult_prevalence_rate

2. Enamasti avastatakse maksavähk hilises staadiumis. Sümptomite ilmnemisel on enamasti vähk arenenud juba kaugemale ja andnud metastaase, mistõttu ravi enamasti ei õnnestu. Ühendriikides näiteks sureb 2/3 patsientidest peale sümptomite ilmnemist kahe kuu jooksul, 5 aastat või kauem elavad kõigest 6% maksavähi diagnoosi saanud haigetest. Ka Eestis sureb maksavähki keskel läbi 100 inimest aastas (peaaegu kõik maksavähi all kannatajad on üle 60 aasta vanad).

Meie ettepanek oleks viia sisse lausaline maksavähi testimine perearstikeskustes inimestele vanuses 60 aastat või enam (kõik mistahes põhjusel perearsti poole pöördunud vanurid). Selle proovimiseks uurisime 10 000 vanurit, kellest 210-lt leiti varases staadiumis maksavähk. Leitud maksavähijuhtude korral soovitasime inimestele kemoterapiat (üks patsientidest keeldus ravist). Varases staadiumis maksavähi leidmiseks kasutasime uut kõrge tundlikkusega (98%) ja spetsiifilisusega (98%) diagnostilist testi.

209-st ravi saanud patsiendist tervenesisid täielikult (nende surmapõhjuseks polnud vähk ja 5 järgneva aasta jooksul ei tekkinud maksavähile iseloomulikke kaebuseid) 198. Tänu varasele diagnoosile tõusis patsientide 5 aasta elulemus elulemus 6%-lt 95%-le. Soovitame tungivalt võtta kasutusele skriinimine (läbi viia vanema elanikkonna regulaarseid sõeluuringuid) maksavähi vastu. Maksavähi varasel avastamisel on tohutu mõju patsientide hilisemale elulemusele! Kuna meie poolt soovitatav test on imeodav, siis peamiseks kuluks on perearstide vaeva kompenseerimiseks mõeldud lisatasud-preemiad (projekti kogukulu on täiesti võrreldav 1-2 maksavähi all kannatava inimese ravikuluga)!

bibliograafia

- [1] L Daniel Maxim, Ron Niebo ja Mark J Utell. “Screening tests: a review with examples”. *Inhalation toxicology* 26.13 (2014), lk. 811–828.
- [2] Andreas Nabers *et al.* “Amyloid blood biomarker detects Alzheimer’s disease”. *EMBO molecular medicine* 10.5 (2018).
- [3] World Health Organization *et al.* “WHO recommendations on the use of rapid testing for influenza diagnosis”. *Geneva, Switzerland: WHO* (2005).