

Tartu Ülikool
Matemaatika-informaatikateaduskond
Matemaatilise statistika instituut

Pille Kool

Sõltuvate uuringute meta-analüüs

Magistritöö

Juhendaja: Märt Möls (TÜ)

Tartu 2010

Sisukord

Sissejuhatus	3
1. Fikseeritud mõjudega meta-analüüsi mudel	5
1.1. Ülevaade	5
1.2. Mudel	5
1.3. Mudeli parameetrite hindamine (üldistatud vähimruutude meetod)	8
2. Juhuslike mõjudega meta-analüüsi mudel	11
2.1. Sissejuhatus	11
2.2. Mudel	12
2.3. Mudeli parameetrite hindamine	14
2.3.1 Suurima tõepära meetod β , σ_{γ}^2 , γ hindamiseks	14
2.3.2. Juhuslike efektidega meta-analüüsi mudeli traditsiooniline esitusviis	18
3. Sõltuvate uuringute juhuslike mõjudega mudel	24
4. Uuringutulemuste sõltuvus geograafilisest asukohast	26
4.1 Sissejuhatus	26
4.2 Mudelid ruumilise sõltuvuse kirjeldamiseks	26
5. Valikulisest publitseerimisest tingitud nihe	32
5.1 Publitseerimisnihe	32
5.2 Lehterjoonis	32
5.3 Lehterjoonise asümmeetria põhjuseid	33
5.4 Lehterjoonise asümmeetriatestid	34
5.4.1. Eggeri test	34
5.4.2. Eggeri test fikseeritud mõjudega mudeli jaoks	35
5.4.3. Eggeri test juhuslike mõjudega mudeli korral	36
5.5. Modifitseeritud Eggeri test	37
5.6. Simuleerimine	38
Kokkuvõte	40
Summary	42
Kasutatud kirjandus	43
Lisa	49

Programm 1	49
Programm 2	49
Programm 3	53
Programm 3	55
Programm 4	59

Sissejuhatus

Iga päev ilmub mitmeid teadusartikleid, maailmas on nüüdseks kümneid tuhandeid teadusajakirju. Uuringutulemused on sageli vastuolulised ja järjest raskem on selles infotulvas orienteeruda.

Kõige esimese meta-analüüsi teostas Karl Pearson 1904. aastal. Ta üritas tõsta oma testi võimsust kombineerides mitmeid väikeseid valimeid kasutanud uuringuid (Pearson, 1904). Esimene avaldatud meta-analüüs ilmus 1940-ndatel ja autoriteks olid J. G. Pratt, J. B. Rhine jt ning see sisaldas 145 uuringu andmeid (osaliselt varem avaldamata) aastatest 1882-1939. Meditsiinialane meta-analüüs aga avaldati esmakordselt alles 1955. aastal. Gene Glass pakkus 1976. a. välja statistilise meetodi uurimistulemuste kombineerimiseks, mille ta nimetas „meta-analüüsiks” (Glass, 1976). Sellest ajast alates on hakatud meta-analüüsi mõistet kasutama erinevate uuringute tulemuste statistilise analüüsina. Meta-analüüs:

1. kasutab juba läbiviidud uuringute tulemusi;
2. baseerub koondstatistikal, mitte originaalandmetel. Näiteks keskmised, standardhälbed ja statistiliste testide tulemused;
3. tunneb huvi töötluse (näiteks ravi) mõju suuruse kirjeldamise vastu, mitte ei piirdu vaid tema olemasolu statistilise tõestamisega;
4. ei piirdu lihtsalt töötluse mõju hindamisega, vaid meta-analüüsi abil üritatakse ka mõista, kuidas mõjutavad töötluse mõju erinevad populatsioonid, katsetingimused.

Enne meta-analüüsi meetodite teket olid laialdaselt levinud artiklite süstemaatilised ülevaated. Meta-analüüsi eeliseks (artiklite ülevaate ees) on see, et saab hinnata üldist faktorite mõju. Tänu tulemuste kombineerimisele on võimalik saavutada suuremat statistilist võimsust, uuringute tulemusi saab üldistada üldkogumile. Lisaks, artiklite ülevaated on muutunud vähem kasutatavateks just seetõttu, et järjest rohkem tuleb juurde uusi uuringutulemusi ja inimene ei suuda kuigi hästi orienteeruda kokku kogutud suures informatsioonihulgas. Sel põhjusel on alates 1980-ndatest analüüsijad järjest rohkem hakanud kasutama meta-analüüsi.

Meta-analüüsi põhisammud on: 1) kirjanduse otsimine – soovitakse teha vähemalt kahel uurijal teineteisest sõltumatult. Nii üritatakse kindlustada võimalikult paljude eelnevate uuringute kaasamine. 2) Uuritavad statistikud ja kasutatavad ravivõtted peavad olema kõigis uuringutes piisavalt sarnased. Tuleb otsustada, millised ilmumata uurimused kaasata meta-analüüsi, et vältida publitseerimisnihet. 3) Meta-analüüsi mudelite abil leitakse hinnangud parameetritele. 4) Publitseerimise nihke kontroll. Käesolevas magistritöös pakutakse täiendavalt välja uudne meetod olukorra jaoks, mil meta-analüüsis kasutatud uuringud on omavahel sõltuvad.

Laialdaselt kasutatakse kahte tüüpi meta-analüüsi mudeleid: fikseeritud ja juhuslike mõjudega mudel. Antud magistritöös antakse neist lühike ülevaate ja täpsemalt tutvustatakse juhuslikke uuringumõjusid sisaldavaid mudeleid. Seni avaldatud juhuslike mõjudega meta-analüüsid on eeldatud, et kõik uuringud käituvad kui väike valim –uuringud on juhuslikult ja teineteisest sõltumatult valitud ühest suurest hüpoteetilisest võimalike uuringute populatsioonist. Praktikas võib sageli näha midagi muud. Magistritöös uuritakse, kuidas analüüsida uuringuid olukorras, kus osad meta-analüüsis kasutatud uuringutulemused võiksid olla sõltuvad. Mõnes piirkorras võib olla olemas (antud) uuringute tegemiseks vajalik kompetents, antud riigis ja tema naaberriikides võib seetõttu olla tehtud palju uuringuid, sageli ka sarnast metodoloogiat kasutades. Näiteks, soovitakse kombineerida Eestis, Lätis, Soomes, Argentiinas, Mehhikos ja Austraalias tehtud uuringutulemusi. Oleks ootuspärane oletada, et Eestis, Lätis ja Soomes tehtud uuringute raviefektid oleksid teineteisele sarnasemad kui näiteks Argentiinas, Mehhikos või Austraalias. Vahel võib sarnasus olla pigem seotud ajaga – 1960-ndatel aastatel nähtud raviefektid võivad olla sarnasemad teineteisele kui 2010-ndal aastal tehtud uuringu raviefekt (kuna toetav ravi on muutunud, jne). Lisaks tegelike andmete uuesti analüüsimisele esitatakse ka simuleeritud andmete analüüsi tulemusi.

1. Fikseeritud mõjudega meta-analüüsi mudel

1.1. Ülevaade

Üks meta-analüüsi eesmärke on soov hinnata üldist või kombineeritud töötluste (näiteks ravi) mõju. Kui kõik uuringud oleks tehtud täpselt ühtemoodi, samadel tingimustel (kõigis uuringutes oleks kasutatud ühesuurust valimit, uuritava tunnuse hajuvus oleks erinevates uuritavates populatsioonides sama suur ning töötluste mõju oleks kõigil samasugune), siis võiks lihtsalt arvutada töötluste mõjuhinnangute keskmise. Kui mõned uuringud on väiksema varieeruvusega (nendest uuringutest saadavad hinnangud on täpsemad) kui teised, siis sooviks nende uuringute tulemustele lisada rohkem kaalu ning just seda tehakse meta-analüüsis.

Meta-analüüsis kasutatakse kahte tüüpi meta-analüüsi mudeleid – fikseeritud (*fixed effect*) ja juhuslike mõjudega (*random effects*) mudelid (Deeks *et al.*, 2001). Nende puhul tehakse uuringute kohta erinevaid eelduseid. Fikseeritud mõjudega mudeli puhul eeldatakse, et kõigis uuringutes on ühesugune ja võrdne tegelik töötluste mõju. Siiski võib vaadelda ka olukordi, kus töötluste mõju sõltub deterministlikult mõnest täiendavast muutujast, näiteks patsientidele antud ravimikogusest. Kuna kõigis uuringutes on sama (ehk fikseeritud) tegelik efekt, siis vaadeldud efekti suurus varieerub erinevates uuringutes ainult hinnanguvea tõttu.

1.2. Mudel

Oletame, et meil on r erinevat uuringut, kus võrreldakse näiteks kontrollgruppi ravigrupiga. Tegelik töötluste mõju tähistatakse parameetriga Θ . Kui igas uuringus oleks kasutatud lõpmatult suurt valimimahtu, siis hinnangud oleksid täpsed (hinnanguvea oleks 0) ja iga uuringu mõõdetud/hinnatud efekt $\hat{\Theta}_i$ oleks võrdne tegeliku efektiga, $\hat{\Theta}_i = \Theta$. Praktikas (loomulikult) pole hinnangud täiesti täpsed ja vaadeldud mõju (ehk uuringus leitud töötluste mõju hinnang) erineb tegelikust mõjust hinnanguvea ε_i võrra, $\hat{\Theta}_i = \Theta + \varepsilon_i$.

Üldine fikseeritud mõjudega mudel on maatrikskujul

$$\hat{\Theta}_{r \times 1} = \mathbf{1}_{r \times 1} \Theta + \boldsymbol{\varepsilon}_{r \times 1} \quad (1.1)$$

ehk komponentide kaupa välja kirjutatult

$$\begin{bmatrix} \hat{\Theta}_1 \\ \hat{\Theta}_2 \\ \dots \\ \hat{\Theta}_r \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} \Theta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_r \end{bmatrix},$$

kus $\hat{\Theta}$ on $r \times 1$ uuringuefektide hinnangute vektor, $\mathbf{1}$ on $r \times 1$ ühtesid sisaldav vektor, Θ on tegelik tundmatu raviefekti suurus, $\boldsymbol{\varepsilon}$ on $r \times 1$ uuringuvigu ehk jääke sisaldav vektor.

Vahel hinnatakse raviefekti veidi erinevates tingimustes, näiteks kasutatakse erinevaid ravidoose. Sel juhul võib i . uuringus ravimi tegelik mõju Θ_i sõltuda ravimikogusest (siin sõltub vaid ühest faktorist), $\Theta_i = \beta_1 + \beta_2 \cdot \text{ravimikogus}_i$. Fikseeritud mõjudega mudeli korral eeldatakse, et antud seos on kõigi uuringute jaoks samasugune ja deterministlik. Kui töötluse tegelik mõju sõltub $p-1$ tunnusest, siis võib **fikseeritud mõjudega mudeli** kirja panna kujul

$$\mathbf{Y}_{r \times 1} = \mathbf{X}_{r \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{r \times 1}, \quad (1.2)$$

$$E\boldsymbol{\varepsilon} = 0, D\boldsymbol{\varepsilon} = \mathbf{R}.$$

Sama mudel komponentide kaupa kirja pandult:

$$\begin{bmatrix} \hat{\Theta}_1 \\ \hat{\Theta}_2 \\ \dots \\ \hat{\Theta}_r \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{p-1,1} \\ 1 & x_{12} & \dots & x_{p-1,2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1r} & \dots & x_{p-1,r} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \dots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_r \end{bmatrix},$$

kus \mathbf{Y} on $r \times 1$ uuringuefektide hinnangute vektor, \mathbf{X} on $r \times p$ fikseeritud efektide disaini- ehk mudelimaatriks, $\boldsymbol{\beta}$ on tundmatu $p \times 1$ parameetervektor (kirjeldab tegeliku raviefekti muutumist sõltuvalt uuringu läbiviimise tingimustest), $\boldsymbol{\varepsilon}$ on $r \times 1$ hinnanguvigu ehk jääke sisaldav vektor.

Eeldatakse, et vigade keskväärtsus on 0 (st kõigis uuringutes on kasutatud nihketa hinnanguid), seega $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$. Kuna uuringud viiakse läbi erineva täpsusega (näiteks erinev valimisuurus), siis iga uuringu korral on uuringuvea dispersioon erinev, $D\varepsilon_i = \sigma_i^2$.

Seega uuringuefektide hinnangud $\hat{\Theta}_i \sim N(\Theta, \sigma_i^2)$. Eeldades, et uuringud on läbi viidud teineteisest sõltumatult (ei kasutata samu andmeid), võiksid hinnangute vead olla ka teineteisest sõltumatud,

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_r^2 \end{bmatrix}.$$

Kuna parameetervektor $\boldsymbol{\beta}$ sisaldab vaid fikseeritud efekte (konstante) on uuringuefektide hinnangute vektori \mathbf{Y} dispersioonimaatriks kujul $\mathbf{V} = D\mathbf{Y} = \mathbf{R}$. Enamasti on parameetrite hinnangud vähemalt asümptootiliselt normaaljaotusega (Schulze, 2004) ning eeldatakse, et:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}).$$

1.3. Mudeli parameetrite hindamine (üldistatud vähimruutude meetod)

Tundmatute parameetrite vektori β väärtust võib hinnata mitmete meetoditega, üks enamlevinud meetod on **üldistatud vähimruutude meetod** (*generalized least-squares* (GLS)) (Littell *et al.*, 2004; Goldstein 1995). Sel juhul leitakse mudeli (1.2) tundmatute parameetrite vektorile β hinnang minimiseerides avaldise $(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$ väärtus. Saadud hinnang on kujul

$$\hat{\beta}_{GLS} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}. \quad (1.3)$$

Üldistatud vähimruutude hinnangut kasutades saab leida parima lineaarse nihketa hinnangu (BLUE – *Best Linear Unbiased Estimator*) mistahes hinnatavale parameeterfunktsioonile (Searle *et al.*, 1992). Üldistatud vähimruutude hinnangu dispersioon avaldub kujul:

$$\begin{aligned} D\hat{\beta}_{GLS} &= D\left((\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}\right) \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} D\mathbf{Y} \left((\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}\right)^T \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{V} \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}. \end{aligned} \quad (1.4)$$

Kui meil on mudel (1.1), siis üldistatud vähimruutude hinnang avaldub kujul:

$$\begin{aligned} \hat{\Theta}_{GLS} &= \left(\begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix}^T \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \sigma_r^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix}^T \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \sigma_r^2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\Theta}_1 \\ \dots \\ \hat{\Theta}_p \end{bmatrix} \\ &= \left(\sum_{i=1}^p \frac{1}{\sigma_i^2} \right)^{-1} \sum_{i=1}^p \left(\frac{1}{\sigma_i^2} \cdot \hat{\Theta}_i \right), \end{aligned}$$

kus täpsed $\sigma_1^2, \dots, \sigma_r^2$ väärtused pole teada. Parameetrite hindamisel kasutatakse uuringutes raporteeritud standardvigasid või dispersioone kui tegelikke väärtuseid. Tähistades $w_i = 1/\sigma_i^2$, saadakse

$$\hat{\Theta}_{GLS} = \frac{\sum_{i=1}^p w_i \hat{\Theta}_i}{\sum_{i=1}^p w_i}. \quad (1.5)$$

Hinnangu dispersioon avaldub kujul:

$$\begin{aligned} D(\hat{\Theta}_{GLS}) &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\ &= \left(\sum_{i=1}^p w_i \right)^{-1}. \end{aligned} \quad (1.6)$$

Saab leida ka $(1 - \alpha)$ -usalduspiirid parameeterhinnangule $\hat{\Theta}_{GLS}$:

$$\hat{\Theta}_{GLS} \pm z_{\alpha/2} \sqrt{D(\hat{\Theta}_{GLS})}, \quad (1.7)$$

kus $z_{\alpha/2}$ on normaaljaotuse $\alpha/2$ kvantiil.

Hinnangut (1.5) tuntakse fikseeritud efektidega meta-analüüsi mudeli hinnanguna (DerSimonian & Laird, 1986; Whitehead, 2002, lk 59).

Näide 1

Kõige lihtsam moodus meta-analüüsi läbiviimiseks on siis, kui on mitmeid uuringuid, mis sisaldavad täpselt sama kahte gruppi või töötlust, katse on läbi viidud täpselt samades tingimustes. Olgu kaks samadel tingimustel teostatud uuringut, kus võrreldi ravimi mõju vererõhule. Mõõdeti, kas vererõhk vähenes või ei (väljamõeldud juhtum):

Uuring	OR	SE
1	4.1	2.7
2	1.4	1.2

Mudel (1.1) omandab kuju

$$\begin{bmatrix} 4.1 \\ 1.4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Theta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix},$$

kus

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \\ = \begin{bmatrix} 2.7^2 & 0 \\ 0 & 1.2^2 \end{bmatrix}.$$

Üldistatud vähimruutude meetodil (1.3) hinnang $\hat{\beta}_{GLS}$ tegelikule mõjule on:

$$\hat{\Theta}_{GLS} = \underbrace{\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2.7^{-2} & 0 \\ 0 & 1.2^{-2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1}}_{(2.7^{-2}+1.2^{-2})^{-1}} \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2.7^{-2} & 0 \\ 0 & 1.2^{-2} \end{bmatrix} \begin{bmatrix} 4.1 \\ 1.4 \end{bmatrix}}_{(2.7^{-2} \cdot 4.1 + 1.2^{-2} \cdot 1.4)} \\ = 1.845.$$

Hinnangudispersioon leitakse, kasutades valemit (1.4):

$$D\hat{\Theta}_{GLS} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\ = \underbrace{\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2.7^{-2} & 0 \\ 0 & 1.2^{-2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1}}_{(2.7^{-2}+1.2^{-2})^{-1}} \\ = 1.2$$

ja normaaljaotuse $0.05/2$ kvantiil on 1.96. Vähimruutude hinnang tegelikule efektile on $\hat{\Theta}_{FIX} = 1.845$ ja 95% usaldusintervall ($- 0.304 \dots 3.995$).

2. Juhuslike mõjudega meta-analüüsi mudel

2.1. Sissejuhatus

Fikseeritud mõjudega mudeli puhul eeldatakse, et töötluse (tegelik) mõju on kõigis uuringutes sama või muutub samadel alustel sõltuvalt uuringu karakteristikutest (toimeaine kogus, vaatlusperioodi pikkus, vms). Paljudel juhtudel on see eeldus väheusutav. Kui soovitakse metaanalüüsiga analüüsida teatud hulka uuringuid, siis eeldatakse, et antud uuringud on piisavalt sarnased, et on mõtet informatsiooni koos vaadelda. Samas, üldiselt ei saa eeldada, et nad on identsed selles mõttes, et uuringute tegelik mõju suurus kõikides uuringutes on täpselt sama.

Näiteks, soovitakse jälgida haiguse arengut kahes patsientide grupis (vaktsineeritud vs platseebo). Kui ravi mõjub, siis eeldatakse, et kõigis uuringutes on mõju suurus (näiteks riskide suhe) sarnane, kuid mitte identne. Isegi kui kõik katsetingimused oleks samad, siis uuringuid viiakse läbi erinevates populatsioonides ning on võimalik uuritavate/patsientide geneetilisest varieeruvusest tingitud erinevused. Raviefekt võib olla suurem (või väiksem), kui uuritavad on vanemad või rohkem haritud või tervemad kui teised, kui nende eest hoolitsetakse teisiti, jne. Või, töötatakse uuringutega, mis uurivad õppemetoodika mõju. Mõju suurus võib sõltuda muude ressursside kättesaadavusest lastele, klassi suurusest, vanusest, teistest faktoritest, mis suure tõenäosusega on uuringutes (populatsioonides) erinevad.

Kõiki töötluse mõju mõjutada võivaid tunnuseid pole enamasti võimalik mõõta. Teada ei pruugi isegi olla, millised muutujad mõjusuurst tegelikult mõjutavad ning tuleb leppida, et erinevates uuringutes nähtav töötluse tegelik mõju võib uuritava populatsiooni eripäradest lähtuvalt olla erinev. Vaatamata sellele, et töötluse tegelik mõju muutub uuringust uuringusse, võib keskenduda töötluse keskmisele mõjule üle erinevate uuringute (üle erinevate uuritavate populatsioonide). Kui töötluse tegelik mõju muutub erinevates uuringutes raskesti seletatavatel põhjustel, kasutatakse uuringutulemuste kombineerimiseks juhuslike mõjudega meta-analüüsi mudelit (DerSimonian & Laird, 1986).

2.2. Mudel

Juhuslike mõjudega mudelis eeldatakse, et töötuse (näiteks ravi) tegelik mõju juhuslikult valitud i . ($i=1,\dots,r$) uuringupopulatsioonis Θ_i on normaaljaotusega juhuslik suurus, $\Theta_i \sim N(\Theta, \sigma_\gamma^2)$ ehk $\Theta_i = \Theta + \gamma_i$, kus γ_i on juhuslik efekt (ehk populatsiooni eripära), $\gamma_i \sim N(0, \sigma_\gamma^2)$.

Mudel i . uuringul saadud hinnangule oleks siis kujul

$$\begin{aligned}\hat{\Theta}_i &= \Theta_i + \varepsilon_i \\ &= (\Theta + \gamma_i) + \varepsilon_i.\end{aligned}\tag{2.1}$$

Eeldades, et juhuslikud efektid γ_i ja ε_i on sõltumatud (Whitehead, 2002, lk 88), võib kirjutada

$$\hat{\Theta}_i \sim N(\Theta, \sigma_i^2 + \sigma_\gamma^2).$$

Kui uuringu keskmine mõju sõltub mingitest teadaolevatest teguritest/tunnustest (näiteks kasutatud ravimikogus), võib mudeli (2.1) esitada üldisemal kujul:

$$\hat{\Theta}_i = \mathbf{X}_i \boldsymbol{\beta} + \gamma_i + \varepsilon_i.$$

Üldine juhuslike mõjudega meta-analüüsi mudel on maatrikskujul esitatud

$$\mathbf{Y}_{r \times 1} = \mathbf{X}_{r \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\gamma}_{r \times 1} + \boldsymbol{\varepsilon}_{r \times 1}\tag{2.2}$$

ehk

$$\begin{bmatrix} \hat{\Theta}_1 \\ \hat{\Theta}_2 \\ \dots \\ \hat{\Theta}_r \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{p-1,1} \\ 1 & x_{12} & \dots & x_{p-1,2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1r} & \dots & x_{p-1,r} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \dots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \dots \\ \gamma_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_r \end{bmatrix},$$

kus $\mathbf{Y} = [\hat{\Theta}_1, \dots, \hat{\Theta}_r]^T$ on $r \times 1$ uuringuefektide hinnangute vektor, \mathbf{X} on $r \times p$ disaini- ehk mudelimaatriks, $\boldsymbol{\beta}$ on fikseeritud efektide $p \times 1$ parameetervektor, mis kirjeldab tegeliku raviefekti muutumist sõltuvalt teadaolevatest teguritest, $\boldsymbol{\gamma}$ on tundmatute $r \times 1$ juhuslike mõjude parameetervektor, ning $\boldsymbol{\varepsilon}$ on $r \times 1$ prognoosijääkide vektor.

Võetakse kasutusele järgnevad tähistused:

juhuslike kordajate kovariatsioonimaatriks

$$\mathbf{G} := \mathbf{D}\boldsymbol{\gamma},$$

modeli jääkide kovariatsioonimaatriks

$$\mathbf{R} := D\boldsymbol{\varepsilon}.$$

Sel juhul juhusliku suuruse \mathbf{Y} kovariatsioonimaatriks \mathbf{V} avaldub kujul

$$\begin{aligned}\mathbf{V} &= D\mathbf{Y} = D(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}) \\ &= D(\boldsymbol{\gamma} + \boldsymbol{\varepsilon}) \\ &= D(\boldsymbol{\gamma}) + D(\boldsymbol{\varepsilon}) \\ &= \mathbf{G} + \mathbf{R}.\end{aligned}$$

Erijuhul, kui tegelikult asetleidnud uuringuid saab vaadelda kui juhuslikku valimit kõikvõimalike uuringute populatsioonist, kus uuringud satuvad valimisse teineteisest sõltumatult, $Cov(\gamma_i, \gamma_j) = 0$, oleks juhuslike kordajate kovariatsioonimaatriks kujul

$$\mathbf{G} := D\boldsymbol{\gamma} = \begin{bmatrix} \sigma_\gamma^2 & 0 & \dots & 0 \\ 0 & \sigma_\gamma^2 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sigma_\gamma^2 \end{bmatrix}$$

ja mudeli jääkide kovariatsioonimaatriks

$$\mathbf{R} := D\boldsymbol{\varepsilon} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sigma_r^2 \end{bmatrix}.$$

Uuringutulemuste kovariatsioonimaatriks $\mathbf{V}_{r \times r}$ on sellisel juhul kujul:

$$\mathbf{V} = \begin{bmatrix} \sigma_\gamma^2 + \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_\gamma^2 + \sigma_2^2 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sigma_\gamma^2 + \sigma_r^2 \end{bmatrix}.$$

Märkus: üldine juhuslike mõjudega meta-analüüsi mudel (2.2) on vaadeldav lineaarse segamudeli erijuhuna.

2.3. Mudeli parameetrite hindamine

Lineaarsete segamudelite üks eeliseid tavaliste lineaarsete mudelite ees on võimalus hinnata kovariatsioonimaatriksi parameetreid. Kasutades lineaarseid segamudeleid, oleks vaja lisaks uuritava tunnuse keskvaartuse muutumist kirjeldavatele parameetritele β hinnata ka kovariatsioonimaatriksi parameetreid.

2.3.1 Suurima tõepära meetod β , σ_γ^2 , γ hindamiseks

Parameetrid $\sigma_1^2, \dots, \sigma_r^2$ saadakse varem läbi viidud uuringutest, kuid $D\gamma_i = \sigma_\gamma^2$ pole enamasti teada. Eeldades, et γ ja ϵ on normaaljaotusega, siis $\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V})$, kus \mathbf{V} sõltub ühest tundmatust (σ_γ^2) või enamast parameetrist. Sel juhul võib kovariatsiooniparameetrite hindamiseks kasutada **suurima tõepära meetodit** (*ML*) või selle modifikatsiooni – kitsendatud suurima tõepära meetodit (*restricted maximum likelihood, REML*) (Littell *et al.*, lk 498, 1996; Laird & Ware, 1982). Kovariatsioonimaatriksi parameetreid on tänapäeval arvutitarkvara abil võimalik hinnata ka teiste jaotuste korral.

Juhul, kui \mathbf{Y} on normaaljaotusega juhuslik suurus, siis tema jaotusfunktsioon on

$$f(\mathbf{Y}) = |2\pi\mathbf{V}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta)\right).$$

Logaritmiline tõepärafunktsioon avaldub kujul

$$l(\beta, \sigma_\gamma^2) := \ln f = -\frac{1}{2} \ln |2\pi\mathbf{V}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta), \quad (2.3)$$

kus $\mathbf{V} = \mathbf{V}(\sigma_\gamma^2)$ ja huvipakkuvad parameetrid β , σ_γ^2 valitakse selliselt, et valimi tekkimise tõenäosus oleks maksimaalne. Selleks võib leida tõepärafunktsiooni tuletised parameetrite β , σ_γ^2 järgi ja lahendada saadud võrrandisüsteemi (parameetervektoris β on p parameetrit).

Alustatakse β leidmisest:

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= -\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{V}^{-1}(-\mathbf{X}) \\ &= \frac{1}{2}\mathbf{Y}^T \mathbf{V}^{-1}\mathbf{X} - \frac{1}{2}\beta^T \mathbf{X}^T \mathbf{V}^{-1}\mathbf{X}. \end{aligned}$$

Võrdsustatakse saadud tuletis nulliga ja saadakse:

$$\begin{aligned}
 0 &= \frac{1}{2} \mathbf{Y}^T \mathbf{V}^{-1} \mathbf{X} - \frac{1}{2} \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \\
 \frac{1}{2} \mathbf{Y}^T \mathbf{V}^{-1} \mathbf{X} &= \frac{1}{2} \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \\
 \boldsymbol{\beta}^T &= \mathbf{Y}^T \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\
 \boldsymbol{\beta} &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}.
 \end{aligned}$$

Seega, suurima tõepära meetodil jõutakse sama parameetervektori $\boldsymbol{\beta}$ hinnanguni kui vähimruutude meetodi abil (1.3), kuid siin \mathbf{V} on kovariatsioonimaatriksi hinnang suurima tõepära meetodil, mida tuleb veel hinnata. Kui kovariatsioonimaatriks on teada, langeb tulemus muidugi täpselt vähimruutude meetodi hinnanguga kokku. Saamaks hinnangut maatriksile \mathbf{V} , tuleks võtta tuletised ka kõigi tundmatute kovariatsiooniparameetrite järgi ja võrdsustada saadud tuletised nulliga (Lindstrom, Bates, 1990). Seejärel tuleks $\boldsymbol{\beta}$ ja kovariatsiooniparameetrite hinnangute leidmiseks lahendada saadud võrrandisüsteemid. Kuna antud töös vaadeldakse üsna keerukaid kovariatsioonimaatriksi \mathbf{V} struktuure, siis tuletise leidmise asemel maksimiseeritakse numbriliselt funktsiooni (2.3).

Teine võimalus $\boldsymbol{\beta}$, $\boldsymbol{\gamma} \sim N(0, \mathbf{G})$ üheaegseks leidmiseks on kasutada \mathbf{Y} ja $\boldsymbol{\gamma}$ ühisjaotust (Searle *et al.*, lk 275, 1992)

$$\begin{aligned}
 f(\mathbf{Y}, \boldsymbol{\gamma}) &= f(\mathbf{Y} | \boldsymbol{\gamma}) f(\boldsymbol{\gamma}) \\
 &= |2\pi\mathbf{R}|^{-\frac{1}{2}} |2\pi\mathbf{G}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\gamma})^T \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\gamma}) + \boldsymbol{\gamma}^T \mathbf{G}^{-1} \boldsymbol{\gamma} \right]\right). \quad (2.4)
 \end{aligned}$$

Osatuletised $\boldsymbol{\beta}$ ja $\boldsymbol{\gamma}$ järgi võrdsustatakse nulliga. Tähistades hinnangud vastavalt $\hat{\boldsymbol{\beta}}$ ja $\hat{\boldsymbol{\gamma}}$, saadakse võrrandid (2.7), mida nimetatakse segamudelite võrranditeks (*mixed models*) (Robinson, 1991):

$$\mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{X}^T \mathbf{R}^{-1} \boldsymbol{\gamma} = \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Y}, \quad (2.5)$$

$$\mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + (\mathbf{R}^{-1} + \mathbf{G}^{-1}) \boldsymbol{\gamma} = \mathbf{R}^{-1} \mathbf{Y}. \quad (2.6)$$

Saadud võrrandeid (2.5 ja 2.6) võib kompaktsemalt kirjutada kujul

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \\ \mathbf{R}^{-1} \mathbf{X} & \mathbf{R}^{-1} + \hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Y} \\ \mathbf{R}^{-1} \mathbf{Y} \end{bmatrix}, \quad (2.7)$$

mille lahendamisel saadakse

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}, \quad (2.8)$$

$$\hat{\boldsymbol{\gamma}} = \mathbf{G} \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}). \quad (2.9)$$

Hinnangute $\hat{\boldsymbol{\beta}}$ ja $\hat{\boldsymbol{\gamma}}$ kovariatsioonimaatriks on

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \\ \mathbf{R}^{-1} \mathbf{X} & \mathbf{R}^{-1} + \mathbf{G}^{-1} \end{bmatrix}^{-1}, \quad (2.10)$$

kus – märgib üldistatud pöördmaatriksit (Searle, 1971). Kuna maatriksid \mathbf{R} ja \mathbf{G} on tavaliselt tundmatud, siis kovariatsioonimaatriksi hinnangu leidmiseks kasutakse hinnanguid $\hat{\mathbf{R}}$, $\hat{\mathbf{G}}$

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}^T \hat{\mathbf{R}}^{-1} \mathbf{X} & \mathbf{X}^T \hat{\mathbf{R}}^{-1} \\ \hat{\mathbf{R}}^{-1} \mathbf{X} & \hat{\mathbf{R}}^{-1} + \hat{\mathbf{G}}^{-1} \end{bmatrix}^{-1}. \quad (2.11)$$

Kui \mathbf{G} ja \mathbf{R} on teada, siis hinnang $\hat{\boldsymbol{\beta}}$ (2.8) on **parim lineaarne nihketa hinnang** parameetritele $\boldsymbol{\beta}$ (*best linear unbiased estimator* (BLUE)) ja $\hat{\boldsymbol{\gamma}}$ (2.9) on **parim lineaarne nihketa prognoos** parameetritele $\boldsymbol{\beta}$ (*best linear unbiased predictor* (BLUP)) (Littell *et al.*, 1996; Searle 1971; Harville, 1988; Robinson, 1991; Harville, 1988). *Parim* tähendab antud kontekstis, et hinnangutel on minimaalne keskmine ruutviga. Robinson (Robinson, 1991) väidab, et Goldberg (Goldberg, 1962) oli esimene, kes võttis kasutusele termini „parim lineaarne nihketa prognoos“ ja lühendit BLUP hakkas kasutama Henderson aastal 1973.

Goldberg vaatas mudelit

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{D}\mathbf{Y} = \mathbf{V}$$

ja tundis huvi uue vaatluse

$$y_{uus} = \mathbf{x}_{uus}^T \boldsymbol{\beta} + \mathbf{e}_{uus}$$

prognoosimise vastu. Lubatakse vigade omavahelist korreleeritust, $E(\mathbf{e}_{uus} \mathbf{e}^T) = \mathbf{w}^T$ ja eeldatakse, et $E(\mathbf{e}_{uus}) = 0$. Goldberg näitas, et parim lineaarne nihketa prognoos uuele vaatlusele y_{uus} on

$$\mathbf{x}_{uus}^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} + \mathbf{w}^T \mathbf{V}^{-1} \mathbf{Y} - \mathbf{w}^T \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y} \quad (2.12)$$

Magistritöös kirjeldatakse uuringutulemusi aga juhuslike efektidega mudeliga (2.2): $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$. Soovides prognoosida uue uuringu tegeliku raviefekti suurust

$$\Theta_{uus} = \mathbf{x}_{uus}^T \boldsymbol{\beta} + \mathbf{z}_{uus}^T \boldsymbol{\gamma}$$

võib kasutada valemit (2.12). Võetakse $\mathbf{e}_{uus} = \mathbf{z}_{uus}^T \boldsymbol{\gamma}$, $\mathbf{e} = \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$. Mudeli (2.2) jaoks on kovariatsioonimaatriks järgneval kujul:

$$\begin{aligned} \mathbf{D}\mathbf{Y} &= \mathbf{D}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}) \\ &= \overset{\boldsymbol{\gamma} \perp \boldsymbol{\varepsilon}}{\mathbf{D}(\mathbf{X}\boldsymbol{\beta})} + \mathbf{D}(\boldsymbol{\gamma}) + \mathbf{D}\boldsymbol{\varepsilon} \\ &= \mathbf{G} + \mathbf{R}. \end{aligned}$$

Kovariatsioon \mathbf{w}^T omandab kuju:

$$\begin{aligned} \mathbf{w}^T &= \mathbf{E}(\mathbf{e}_{uus} \mathbf{e}^T) \\ &= \mathbf{E}[\mathbf{z}_{uus}^T \boldsymbol{\gamma} (\boldsymbol{\gamma} + \mathbf{e})^T] \\ &= \mathbf{E}[\mathbf{z}_{uus}^T \cdot (\boldsymbol{\gamma}\boldsymbol{\gamma}^T + \boldsymbol{\gamma}\mathbf{e}^T)] \\ &= \mathbf{z}_{uus}^T \left[\underbrace{\mathbf{E}(\boldsymbol{\gamma}\boldsymbol{\gamma}^T)}_{\mathbf{D}(\boldsymbol{\gamma})} + \underbrace{\mathbf{E}(\boldsymbol{\gamma}\mathbf{e}^T)}_{\substack{\mathbf{E}\boldsymbol{\gamma} \cdot \mathbf{E}\mathbf{e}^T \\ \mathbf{0}}} \right] \\ &= \mathbf{z}_{uus}^T \mathbf{D}(\boldsymbol{\gamma}) \\ &= \mathbf{z}_{uus}^T \mathbf{G}. \end{aligned}$$

Seega, asendades saadud kovariatsiooni $\mathbf{w}^T = \mathbf{z}_{uus}^T \mathbf{G}$ avaldisse (2.12), saadakse

$$\begin{aligned} &\mathbf{x}_{uus}^T \underbrace{[\mathbf{X}^T (\mathbf{G} + \mathbf{R})^{-1} \mathbf{X}]^{-1} \mathbf{X}^T (\mathbf{G} + \mathbf{R})^{-1} \mathbf{Y}}_{\hat{\boldsymbol{\beta}}} \\ &+ \mathbf{z}_{uus}^T \mathbf{G} (\mathbf{G} + \mathbf{R})^{-1} \mathbf{Y} \\ &- \mathbf{z}_{uus}^T \mathbf{G} (\mathbf{G} + \mathbf{R})^{-1} \cdot \mathbf{X} \underbrace{[\mathbf{X}^T (\mathbf{G} + \mathbf{R})^{-1} \mathbf{X}]^{-1} \mathbf{X}^T (\mathbf{G} + \mathbf{R})^{-1} \mathbf{Y}}_{\hat{\boldsymbol{\beta}}} \\ &= \mathbf{x}_{uus}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{uus}^T \mathbf{G} (\mathbf{G} + \mathbf{R})^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{x}_{uus}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{uus}^T \mathbf{G}\mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{x}_{uus}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{uus}^T \hat{\boldsymbol{\gamma}}, \end{aligned} \tag{2.13}$$

mis on parim lineaarne nihketa prognoos (BLUP) avaldisele $\mathbf{x}_{uus}^T \hat{\boldsymbol{\beta}} + \mathbf{z}_{uus}^T \hat{\boldsymbol{\gamma}}$. Suuruse $\hat{\boldsymbol{\gamma}}$ parim lineaarne nihketa prognoos saadakse, kui võetakse $\mathbf{x}_{uus}^T = 0$ ja $\mathbf{z}_{uus}^T = \mathbf{I}$:

$$\hat{\boldsymbol{\gamma}} = \mathbf{G}\mathbf{V}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

Seega Goldbergeri prognoos on sama, mis saadi suurima tõepära hinnangutega (2.8, 2.9).

2.3.2. Juhuslike efektidega meta-analüüsi mudeli traditsiooniline esitusviis

Meta-analüüsi kirjanduses antakse hinnangud enamasti maatrikskuju kasutamata ja seetõttu tutvustatakse käesolevas töös hinnanguvalemeid ka elementkujul.

Eeldati, et parameeterhinnangud on normaaljaotusega juhuslikud suurused, $\hat{\Theta}_i \sim N(\Theta, \hat{\sigma}_i^2 + \hat{\sigma}_\gamma^2)$, kus $\hat{\sigma}_i^2$, $\hat{\sigma}_\gamma^2$ on vastavalt σ_i^2 , σ_γ^2 hinnangud. Tähistades

$$\begin{aligned} w_i &= \frac{1}{\hat{\sigma}_i^2} \\ w_i^* &= (\hat{\sigma}_i^2 + \hat{\sigma}_\gamma^2)^{-1} \\ &= (w_i^{-1} + \hat{\sigma}_\gamma^2)^{-1}, \end{aligned}$$

saadakse, et $\hat{\Theta}_i \sim N(\Theta, (w_i^*)^{-1})$. Kui $(w_i^*)^{-1}$ on $\hat{\Theta}_i$ tegelik dispersioon, siis juhuslike mõjudega mudelis fikseeritud mõjude hinnang avaldub analoogilisel kujul kui üldistatud vähimruutude hinnang peatükis 1 (1.5):

$$\hat{\Theta}^* = \frac{\sum_{i=1}^p \hat{\Theta}_i w_i^*}{\sum_{i=1}^p w_i^*}. \quad (2.14)$$

Selle tulemuseni jõuti, kasutades suurima tõepärahinnangu avaldist (2.8), kui eeldatakse, et $\mathbf{G} = \sigma_\gamma^2 \mathbf{I}$. Hinnang $\hat{\Theta}^*$ on asümptootiliselt nihketa tegeliku parameetri Θ jaoks ning saadud hinnangu dispersioon on

$$D(\hat{\Theta}^*) = \frac{1}{\sum_{i=1}^p w_i^*}. \quad (2.15)$$

Ligikaudne 95% usaldusvahemik parameetrile $\hat{\Theta}^*$ on (Whitehead 2002, lk 89)

$$\hat{\Theta}^* \pm z_{\alpha/2} \sqrt{\frac{1}{\sum_{i=1}^p w_i^*}}, \quad (2.16)$$

kus $z_{\alpha/2}$ on normaaljaotuse $\alpha/2$ kvantiil. Kui $\hat{\sigma}_\gamma^2$ on väike, siis modifitseeritud kaalud w_i^* on ligilähedased originaalkaaludele w_i . Sel juhul juhuslike mõjudega mudeli abil saadud

standardviga ja 95% usaldusvahemik on sarnane fikseeritud mudelis saadud tulemustele. Samuti tegeliku efekti mõjuhinnang on mõlemas mudelis sarnane. Kui aga $\hat{\sigma}_\gamma^2$ on suur, siis standardviga ja usalduspiirid on juhuslike mõjudega mudelis laiemad kui fikseeritud mõjudega mudelis. (Whitehead 2002, lk 89)

Järgnevas tutvustatakse parameetri σ_γ^2 hindamiseks lahenduskäiku, mis põhineb momentide meetodil. Fikseeritud mõju Θ hinnangu (1.5) oli kujul:

$$\hat{\Theta}_{\text{FIX}} = \frac{\sum_{i=1}^p w_i \hat{\Theta}_i}{\sum_{i=1}^p w_i},$$

mille keskvärtus on ka juhuslike mõjudega mudeli kehtides Θ , kuid dispersiooni leidmisel tuleb arvestada, et kehtib juhuslike efektidega mudel. Dispersioon on:

$$\begin{aligned} D(\hat{\Theta}_{\text{FIX}}) &= D\left(\frac{\sum_{i=1}^r w_i \hat{\Theta}_i}{\sum_{i=1}^p w_i}\right) \\ &= \frac{\sum_{i=1}^r w_i^2 D(\hat{\Theta}_i)}{\left(\sum_{i=1}^p w_i\right)^2} \\ &= \frac{\sum_{i=1}^r w_i^2 (w_i^{-1} + \hat{\sigma}_\gamma^2)}{\left(\sum_{i=1}^p w_i\right)^2} \\ &= \frac{1}{\sum_{i=1}^r w_i} + \frac{\hat{\sigma}_\gamma^2 \sum_{i=1}^r w_i^2}{\left(\sum_{i=1}^r w_i\right)^2}. \end{aligned}$$

Definieeritakse statististik Q

$$\begin{aligned}
Q &= \sum_{i=1}^r w_i (\hat{\theta}_i - \hat{\theta}_{FIX})^2 \\
&= \sum_{i=1}^r w_i (\hat{\theta}_i - \theta + \theta - \hat{\theta}_{FIX})^2 \\
&= \sum_{i=1}^r w_i \left[(\hat{\theta}_i - \theta)^2 + 2 \cdot (\hat{\theta}_i - \theta)(\theta - \hat{\theta}_{FIX}) + (\theta - \hat{\theta}_{FIX})^2 \right] \\
&= \sum_{i=1}^r w_i (\hat{\theta}_i - \theta)^2 + 2 \cdot (\theta - \hat{\theta}_{FIX}) \sum_{i=1}^r w_i (\hat{\theta}_i - \theta) + (\theta - \hat{\theta}_{FIX})^2 \sum_{i=1}^r w_i \\
&= \sum_{i=1}^r w_i (\hat{\theta}_i - \theta)^2 + 2 \cdot (\theta - \hat{\theta}_{FIX}) \left(\underbrace{\sum_{i=1}^r w_i \hat{\theta}_i}_{\hat{\theta}_{FIX} \sum_{i=1}^r w_i} - \theta \sum_{i=1}^r w_i \right) + (\theta - \hat{\theta}_{FIX})^2 \sum_{i=1}^r w_i \\
&= \sum_{i=1}^r w_i (\hat{\theta}_i - \theta)^2 + 2 \cdot (\theta - \hat{\theta}_{FIX}) (\hat{\theta}_{FIX} - \theta) \sum_{i=1}^r w_i + (\theta - \hat{\theta}_{FIX})^2 \sum_{i=1}^r w_i \\
&= \sum_{i=1}^r w_i (\hat{\theta}_i - \theta)^2 - 2 \cdot (\theta - \hat{\theta}_{FIX})^2 \sum_{i=1}^r w_i + (\theta - \hat{\theta}_{FIX})^2 \sum_{i=1}^r w_i \\
&= \sum_{i=1}^r w_i (\hat{\theta}_i - \theta)^2 - (\theta - \hat{\theta}_{FIX})^2 \sum_{i=1}^r w_i \\
&= \sum_{i=1}^r w_i (\hat{\theta}_i - \theta)^2 - \left(\sum_{i=1}^r w_i \right) (\hat{\theta}_{FIX} - \theta)^2. \tag{2.17}
\end{aligned}$$

Teststatistikut Q (2.17) (nimetatakse ka Cochran'i Q statistikuks) kasutatakse heterogeensuse testimiseks. Kui Q on väga suur, siis uuringute vahel on heterogeensus – peaks kasutama juhuslike mõjudega meta-analüüsi mudelit. Fikseeritud efektidega mudeli kehtides peaks teststatistik olema $Q \approx r - 1$ (Higgins, 2003).

Statistiku Q keskvaartuse $E(Q)$ arvutamisel kasutatakse valemit $DX = E(X - EX)^2$ ja saadakse

$$\begin{aligned}
E(Q) &= E\left(\sum_{i=1}^r w_i (\hat{\Theta}_i - \Theta)^2 - \left(\sum_{i=1}^r w_i\right) (\hat{\Theta}_{FIX} - \Theta)^2\right) \\
&= \sum_{i=1}^r w_i E\left(\hat{\Theta}_i - \underset{E\hat{\Theta}_i}{\Theta}\right)^2 - \left(\sum_{i=1}^r w_i\right) E\left(\hat{\Theta}_{FIX} - \underset{E\hat{\Theta}_{FIX}}{\Theta}\right)^2 \\
&= \sum_{i=1}^r w_i D(\hat{\Theta}_i) - \left(\sum_{i=1}^r w_i\right) D(\hat{\Theta}_{FIX}) \\
&= \sum_{i=1}^r w_i (w_i^{-1} + \sigma_\gamma^2) - \left(\sum_i w_i\right) \left(\frac{1}{\sum_i w_i} + \frac{\sigma_\gamma^2 \sum_i w_i^2}{\left(\sum_i w_i\right)^2}\right) \\
&= \sum_{i=1}^r w_i w_i^{-1} + \sigma_\gamma^2 \sum_{i=1}^r w_i - \sum_i w_i \frac{1}{\sum_i w_i} - \frac{\sigma_\gamma^2 \sum_i w_i^2}{\left(\sum_i w_i\right)} \\
&= \sum_{i=1}^r 1 + \sigma_\gamma^2 \sum_{i=1}^r w_i - 1 - \frac{\sigma_\gamma^2 \sum_i w_i^2}{\left(\sum_i w_i\right)} \\
&= (r-1) + \sigma_\gamma^2 \left(\sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i}\right).
\end{aligned}$$

Siit saab välja kirjutada nihketa hinnangu uuringutevahelise hajuvusele σ_γ^2 :

$$\hat{\sigma}_\gamma^2 = \frac{Q - (r-1)}{\sum_i w_i - \sum_i w_i^2 / \sum_i w_i}. \quad (2.18)$$

Saadud hinnang langeb kokku DerSimonian ja Laird poolt 1986. aastal avaldatud hinnanguga (DerSimonian & Laird 1986). Valemit (2.18) kasutades võib saada ka negatiivse hinnangu (kuid dispersioon ei saa olla negatiivne), siis sellisel juhul võetakse praktikas parameetrihinnanguks $\hat{\sigma}_\gamma^2 = 0$ (Whitehead 2002, lk 88).

Näide 2

Olgu vaatluse all mutatsioon (faktor V Leideni mutatsioon – FVL), mis suurendab vere hüübivust, seeläbi soodustab trombide teket ja sellest tulenevalt tõstavad ajuinsuldi riski. Vaatluse all on 9 uuringut, mis teostati aastatel 1998-2009 (tabel 1).

Tabel 1:

Nr	Autor	Aasta	OR	SE
1	Debus	2004	4.1	0.59
2	Günther	2000	4	0.42
3	Hagstrom	1998	6.3	0.9
4	Heller	1999	3.7	0.67
5	Heller	2003	3.1	0.43
6	Laugesaar	Ilmumata	1	0.78
7	Miller	2006	1.6	0.65
8	Nowak-Göttl	1999	6	0.36
9	Sträter	1999	3.6	0.7

Soovitakse hinnata fikseeritud ja juhuslike mõjudega meta-analüüsi mudeleid. Juhuslike mõjudega mudeli korral leitakse keskmise raviefekti hinnang $\hat{\Theta}^*$ suurima tõepära meetodil ning ka saadud raviefekti standardhälbe (ehk raviefekti varieeruvuse) hinnang $\hat{\sigma}_\gamma^2$. Juhuslike mõjudega mudeli hindamisel antakse ette kovariatsioonimaatriks järgneval kujul:

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 + \sigma_\gamma^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 + \sigma_\gamma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_9^2 + \sigma_\gamma^2 \end{bmatrix}.$$

Fikseeritud efektidega mudeli korral jaoks $\hat{\Theta}_{FIX}=4.05$, 95%UI: 3.70 – 4.40. Juhuslike efektidega mudeli korral aga $\hat{\Theta}^*=3.73$, 95%UI: 2.69 – 4.76, $\hat{\sigma}_\gamma^2=2.13$.

Juhuslike mõjudega meta-analüüsi mudel hinnatakse ka momentide meetodil. Statistiku Q arvutamisel kasutatakse eelnevalt leitud parameetrihinnangut $\hat{\Theta}_{FIX}=4.05$, mis saadi, kui eeldati fikseeritud mõjudega mudeli olemasolu, $Q=70.7$. Asendades saadud väärtuse

valemisse (2.18), siis hinnangu raviefekti varieeruvus on $\hat{\sigma}_\gamma^2=2.37$. Tulemused on esitatud kokkuvõtlikus tabelis 2.

Kõigi uuringute andmeid kasutades prognoositakse, milline on ajuinsuldi saamise šanss (OR) Eestis elaval V Leideni mutatsiooni kandval patsiendil. Parim lineaarne nihketa prognoos saadakse, kasutades valemit (2.13), kusjuures $\hat{\beta} := \Theta$, $\mathbf{X}_{uus} = 1$,

$$z_{uus} = [0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]^T,$$

$$\gamma = [\gamma_1 \ \gamma_2 \ \gamma_3 \ \gamma_4 \ \gamma_5 \ \gamma_6 \ \gamma_7 \ \gamma_8 \ \gamma_9]^T,$$

$$\mathbf{G} := D\gamma = \begin{bmatrix} \sigma_\gamma^2 & 0 & \dots & 0 \\ 0 & \sigma_\gamma^2 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sigma_\gamma^2 \end{bmatrix}.$$

Juhuslike mõjudega mudeli korral $\hat{\gamma}_6 = -2.167$ (kasutades suurima tõepärahinnangut $\sigma_\gamma^2 = 2.129$). Valemi $\hat{\Theta}_6 = \hat{\Theta} + \gamma_6$ abil arvutatakse parim nihketa prognoos Eestis läbiviidud uuringu jaoks.

Tabel 2. V Leideni mutatsiooni mõju insuldiriskile

Mudel	Hindamis-meetod	$\hat{\Theta}$	95%UI	$D\hat{\Theta}$	$\hat{\sigma}_\gamma^2$	$\hat{\Theta}_{EESTI}$	$se(\hat{\Theta}_{EESTI})$	c
Üksik uuring (Laugesaar, Eesti)						1...	0,78	
Fikseeritud		4.051	3.700...4.402	0.032		4.051	0,18	
Juhuslikud	momentide meetod	3.723	2.638...4.808	0.307	2.372	1.556	0.705	
Juhuslikud	STP	3.725	2.689...4.761	0.279	2.129	1.607	0.698	
Juhuslikud mõjud, sõltuvad uuringud	STP	3.671	2.599...4.744	0.2995	2.309	1.557	0.699	2.309

$\hat{\Theta}_{EESTI}$ – prognoos ajuinsuldi šansside suhtele Eestis (V-Leideni mutatsioon vs kontroll).

STP – suurima tõepärameetod

3. Sõltuvate uuringute juhuslike mõjudega mudel

Vaadates näiteks kahte uuringut, mis on läbi viidud erinevates (juhuslikult valitud riikides) erinevate (juhuslikult valitud) uurijate poolt, siis arvatavasti on tegelikud uuringutulemused nendes sõltumatud, $\Theta_i \perp \Theta_j$. Mõned uuringud võivad olla läbi viidud sarnastes oludes või sarnast metoodikat kasutades. Teised aga teistes, väga erinevates oludes, või väga erinevat populatsiooni kasutades. Järgnevas uuritakse, kuidas analüüsida uuringuid olukorras, kus osad nendest võivad olla omavahel seotud.

Näiteks, vaadeldakse 3 uuringut, millest 1. ja 2. on teostatud ühes riigis, 3. uuring aga teises riigis. Sel juhul võiks juhuslike mõjude γ kovariatsioonimaatriks \mathbf{G} olla kujul

$$\mathbf{G} = \begin{bmatrix} \sigma_\gamma^2 & c & 0 \\ c & \sigma_\gamma^2 & 0 \\ 0 & 0 & \sigma_\gamma^2 \end{bmatrix},$$

kus konstant c iseloomustab uuringute vahelise sõltuvuse suurust. Samas riigis teostatud uuringud kasutavad sama uuritavate populatsiooni ja lisaks võib ka uuringut läbi viiva personali seas esineda kattuvust (samad arstid, vms). Uuringutulemuste kovariatsioonimaatriks oleks sellisel juhul kujul

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 + \sigma_\gamma^2 & c & 0 \\ c & \sigma_2^2 + \sigma_\gamma^2 & 0 \\ 0 & 0 & \sigma_3^2 + \sigma_\gamma^2 \end{bmatrix}.$$

Näide 2 (jätkub)

Hinnatakse juhuslike efektidega mudel eeldades uuringute sõltuvust. Kaks vaadeldud uuringut on tehtud sama autori poolt (tabel 1, Heller, 4. ja 5. uuring). Usutav, et autor kasutas sarnast uuringumetoodikat, arvatavasti ka samast piirkonnast pärit inimesi, vms. Seetõttu võib oletada, et tegelikud raviefektid mõlemal juhul on mõnevõrra sarnased. Antakse ette kovariatsioonimaatriks

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 + \sigma_\gamma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 + \sigma_\gamma^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 + \sigma_\gamma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4^2 + \sigma_\gamma^2 & c & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c & \sigma_5^2 + \sigma_\gamma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_6^2 + \sigma_\gamma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_7^2 + \sigma_\gamma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_8^2 + \sigma_\gamma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_9^2 + \sigma_\gamma^2 \end{bmatrix}$$

ja saadakse juhuslike mõjudega mudeli jaoks: $\hat{\Theta} = 3.74$, 95%UI: 2.69...4.78, $D(\hat{\Theta}) = 0.28$, $\hat{\sigma}_\gamma^2 = 2.3$. Uuringute vahelise sõltuvuse suuruse hinnang on $\hat{c} = 2.3$, mis näitab, et seos antud uuringutel oli olemas ja osutus maksimaalseks võimalikest:

$$\begin{aligned} \text{cor}(\gamma_4, \gamma_5) &= \frac{\text{cov}(\gamma_4, \gamma_5)}{\sqrt{D\gamma_4 \cdot D\gamma_5}} \\ &= \frac{2.3}{\sqrt{2.3 \cdot 2.3}} \\ &= 1. \end{aligned}$$

Fikseeritud mõjudega mudel annab küll kõige täpsema tulemuse (tabel 2), kuid kuna uuringutulemused (tabel 1) erinevad üksteisest päris palju, siis võib kahelda fikseeritud mõjudega mudeli sobivuses. Juhuslike mõjudega mudeleid hinnates nägime, et sama autori uuringutulemuste vahel on päris suur sõltuvus, mistõttu võiks kasutada viimast, Helli uuringutulemuste sõltuvust lubavat mudelit. Meta-analüüsi abil leiti lisaks ka täpsem hinnang Eestis läbi viidud uuringutulemusele.

4. Uuringutulemuste sõltuvus geograafilisest asukohast

4.1 Sissejuhatus

Eelmises punktis illustreeriti uuringute vahelist sõltuvust juhul, kui uuringud olid tehtud sama autori poolt. Lisati kovariatsioonimaatriksile lisaparameter c , kuid vahel ei piisa nii lihtsast näitajast uuringute vahelise korrelatsiooni kirjeldamiseks. Mitte alati ei pruugi uuringutulemuste vaheline korrelatsioon olla täpselt ühesugune. Mida lähemal on kahe uuringu toimumiskohad, seda sarnasemad võivad olla uuringutulemused – sarnasus võib tulla meetodikate, uuringute läbiviijate teadmiste, valimite, uuritavate sarnasusest (nii geneetiline, kliimast kui ka käitumisharjumustest tulenev sarnasus).

4.2 Mudelid ruumilise sõltuvuse kirjeldamiseks

Ruumilised korrelatsioonimudelid võeti algselt kasutusele geoloogias. Kõige lihtsamal juhul on antud uuringu- või vaatlustulemused, kusjuures uuringute/mõõtmiste tegemise kohad on teada. Soovitakse hinnata ruumilist korrelatsiooni. Kui vaadelda mudelit (2.2), siis ruumiline korrelatsioon kajastub maatriksis \mathbf{G} , juhuslike mudeli mõjud kovariatsioonimaatriksis γ . Tähistatakse i . uuringu toimumiskoht sümboliga x_i (mis võib olla ka vektor). Uuringute vahelist sõltuvust kirjeldatakse uuringu toimumiskohtade vahelise kauguse funktsioonina,

$$\text{Cor}(\Theta_i, \Theta_j) = f(d(x_i, x_j)).$$

Kahe uuringu toimumiskohtade x_i ja x_j vaheline kaugus tähistatakse sümboliga $d_{ij} := d(x_i, x_j)$. Kuna juhuslike mõjudega mudel oli $\Theta_i = \Theta + \gamma_i$, siis saadakse

$$\begin{aligned} \text{Cor}(\Theta_i, \Theta_j) &= \text{Cor}(\Theta + \gamma_i, \Theta + \gamma_j) \\ &= \text{Cor}(\gamma_i, \gamma_j) \end{aligned}$$

ehk

$$\text{Cor}(\gamma_i, \gamma_j) = f(d_{ij}). \quad (4.1)$$

Kuigi uuringutulemused võivad olla sõltuvad, eeldatakse, et hinnanguvead ε_i on sõltumatud (uuringud on läbi viidud erinevaid valimeid kasutades): $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$. Eeldades, et

juhuslike efektide γ_i dispersioon on samasugune, $D\gamma_i = \sigma_\gamma^2 \quad \forall i$, saadakse **ruumilise korrelatsioonimudeli üldkujuks**

$$\text{Cov}(\gamma_i, \gamma_j) = \sigma_\gamma^2 [f(d_{ij})]. \quad (4.2)$$

Mudelit, milles **kaugusfunktsioon** $f(d_{ij})$ on sama kõigi võrdsetel kaugustel olevate paaride jaoks, nimetatakse **statsionaarseks** mudeliks. Lisaks, kui kaugusfunktsioon $f(d_{ij})$ ei sõltu ka suunast, siis mudel on isotroopne. Juhuslike mõjude γ kovariatsioonimaatriks **G** on:

$$\mathbf{G} = \begin{bmatrix} \sigma_\gamma^2 & \sigma_\gamma^2 f(d_{12}) & \dots & \sigma_\gamma^2 f(d_{1r}) \\ \sigma_\gamma^2 f(d_{21}) & \sigma_\gamma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_\gamma^2 f(d_{r-1,r}) \\ \sigma_\gamma^2 f(d_{r1}) & \dots & \sigma_\gamma^2 f(d_{r,r-1}) & \sigma_\gamma^2 \end{bmatrix},$$

uuringutulemuste kovariatsioonimaatriks on

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 + \sigma_\gamma^2 & \sigma_\gamma^2 f(d_{12}) & \dots & \sigma_\gamma^2 f(d_{1r}) \\ \sigma_\gamma^2 f(d_{21}) & \sigma_2^2 + \sigma_\gamma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_\gamma^2 f(d_{r-1,r}) \\ \sigma_\gamma^2 f(d_{r1}) & \dots & \sigma_\gamma^2 f(d_{r,r-1}) & \sigma_r^2 + \sigma_\gamma^2 \end{bmatrix}.$$

Näiteid sagedamini kasutatavatest kaugusfunktsioonidest ¹ (Littell *et al.*, 1996):

1. sfääriline (*spherical*)

$$f(d_{ij}) = \left[1 - 1.5 \frac{d_{ij}}{\rho} + 0.5 \left(\frac{d_{ij}}{\rho} \right)^3 \right] 1(d_{ij} < \rho) \quad (4.3)$$

2. Gaussi (*Gaussian*)

$$f(d_{ij}) = \exp\left(-\frac{d_{ij}^2}{\rho^2}\right) \quad (4.4)$$

3. lineaarne

$$f(d_{ij}) = (1 - \rho d_{ij}) 1(\rho d_{ij} < 2) \quad (4.5)$$

4. astmeline (*power*)

$$f(d_{ij}) = \rho^{d_{ij}} \quad (4.6)$$

¹ Funktsioon $1(d_{ij} < \rho)$ võrdub ühega, kui $d_{ij} < \rho$, vastasel juhul on 0.

Geostatistikas kutsutakse parameetreid σ_γ^2 ja ρ vastavalt lävend (*sill*) ja autokorrelatsiooniulatus (*range*) (Littell *et al.*, 1996). Kõige enam kasutatav kauguse mõõt on Eukleidiline kaugus

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2}$$

$$= \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2},$$

kus $x_i = (x_{i1}, \dots, x_{ip})$ ja $x_j = (x_{j1}, \dots, x_{jp})$ on kahe mõõdetud objekti asukohad p -mõõtmelises ruumis. Ühemõõtmelise ruumi korral $x_i = (x_{i1})$, $x_j = (x_{j1})$ ning kaugus arvutatakse järgnevalt $d_{ij} = |x_i - x_j|$.

Mõnel juhul ei tööta kovariatsioonimaatriksi mudelid hästi üsna väikeste kauguste korral. Eelnevalt toodud kaugusfunktsioonide korral on kaugusfunktsiooni väärtus alati 1, kui leitakse kaugus iseendaga, $f(d_{ii}) = f(0) = 1$, seega kahes samas riigis läbiviidud uuringus peaksid olema tegelikud raviefektid samasugused (või vähemalt peaks ravitulemuste vaheline korrelatsioon olema 1). Alati see nii ei ole – meetodika erinevuste, ajas toimuvate sündmuste või ajas muutuva elukeskkonna tõttu ei pruugi ka kaks samas kohas tehtud uuringut anda täpselt samasugust tulemust.

Lubamaks samas kohas läbi viidud uuringutulemuste erinevust, tuuakse sisse lisaparameeter – geostatistikas tuntud kui kamakaefekt σ_K^2 :

$$D\gamma_i = \sigma_\gamma^2 f(d_{ii}) + \sigma_K^2$$

$$= \sigma_\gamma^2 + \sigma_K^2,$$

$$\text{Cov}(\gamma_i, \gamma_j) = \sigma_\gamma^2 [f(d_{ij})],$$

kus $f(d_{ij})$ on näiteks mõni eespool toodud kaugusest sõltuv funktsioon. Parameetreid σ_K^2 , $\sigma_\gamma^2 + \sigma_K^2$ ja ρ nimetatakse geostatistikas vastavalt **kamakas**, **lävi/lävend**, **autokorrelatsiooni ulatus** (*nugget*, *sill*, *range*). Seega, mudeleid, kus $D\gamma_i = \sigma_\gamma^2 + \sigma_K^2$, nimetatakse **kamaka efektiga mudeliteks** ja kus $D\gamma_i = \sigma_\gamma^2$, kutsutakse **kamaka efektita** mudeliteks.

Ruumilise sõltuvusega üldine segamudel on samal kujul, mis juhuslike mõjudega meta-analüüsi mudel (2.2), kus ruumiline sõltuvus kirjeldatakse maatriksis \mathbf{G} . Defineerides maatriksi $\mathbf{F}_{n \times n}$, kus ij -s element on kauguse funktsioon $f(d_{ij})$, siis maatriks \mathbf{G} on kujul:

- mudel, kus pole „kamaka“ efekti: $\mathbf{G} = \sigma_\gamma^2 \mathbf{F}$;
- mudel, kus on „kamaka“ efekt: $\mathbf{G} = \mathbf{I}\sigma_K^2 + \sigma_\gamma^2 \mathbf{F}$.

Näide 3. Sõltuvad uuringud (ruumiline korrelatsioonanalüüs)

Artiklis (Hagenau *et al.*, 2009) uuriti D-vitamiini sisaldust seerumis globaalselt. Uurimusi oli 394 ning artikkel sisaldas ka faili kõigi andmetega. Andmetes oli antud kõigi uuringute keskmine D-vitamiini sisaldus seerumis, standardviga ja laiuskraad.

Antakse ette kovariatsioonimaatriks, milles uuringute vahelist seost soovitakse kirjeldada erinevate kaugusfunktsioonide abil. Antud andmete puhul on tegemist ligikaudse kaugusega, kuna on teada vaid üks asukoha koordinaat – laiuskraad. Kontrollitakse, milline hinnatud meta-analüüsi mudel sobib andmetega paremini. Kuna osad vaadeldud mudelid pole üksteise erijuhud (kasutatakse erinevaid kaugusfunktsioone, ühte võrreldavat mudelit ei saada teisest osade parameetrite fikseerimise tulemusel), siis sellisel juhul on tõepärafunktsioonide võrdlemiseks kasutatav Akaike informatsioonikriteerium (AIC – *Akaike Information Criteria*) (Akaike, 1974)

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

kus k – mudelis hinnatavate parameetrite arv, L – tõepärafunktsiooni väärtus. Väiksem AIC vastab sobivamale mudelile. AIC puuduseks on, et see ei võimalda kontrollida hüpoteese ühe või teise mudeli paremast sobivusest, eeliseks aga võimalus omavahel võrrelda vägagi erinevaid mudeleid.

Esimesena vaadeldakse astmelist kaugusfunktsiooni (4.6) ning kovariatsioonimaatriks, mis on välja kirjutatud esimese 4 uuringu jaoks, näeks välja järgmine:

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + \sigma_\gamma^2 & \sigma_\gamma^2 \rho^{d_{12}} & \sigma_\gamma^2 \rho^{d_{13}} & \sigma_\gamma^2 \rho^{d_{14}} \\ \sigma_\gamma^2 \rho^{d_{21}} & \sigma_2^2 + \sigma_\gamma^2 & \sigma_\gamma^2 \rho^{d_{23}} & \sigma_\gamma^2 \rho^{d_{24}} \\ \sigma_\gamma^2 \rho^{d_{31}} & \sigma_\gamma^2 \rho^{d_{32}} & \sigma_3^2 + \sigma_\gamma^2 & \sigma_\gamma^2 \rho^{d_{34}} \\ \sigma_\gamma^2 \rho^{d_{41}} & \sigma_\gamma^2 \rho^{d_{42}} & \sigma_\gamma^2 \rho^{d_{43}} & \sigma_r^2 + \sigma_\gamma^2 \end{pmatrix}.$$

Näitena kirjutatakse välja kovariatsioonimaatriks esimese 4 uuringu jaoks:

$$\mathbf{V} = \begin{pmatrix} 10.2^2 + \sigma_\gamma^2 & \sigma_\gamma^2 \rho^{|7.1-8|} & \sigma_\gamma^2 \rho^{|7.1-8.3|} & \sigma_\gamma^2 \rho^{|7.1-12|} \\ \sigma_\gamma^2 \rho^{|8-7.1|} & 2.3^2 + \sigma_\gamma^2 & \sigma_\gamma^2 \rho^{|8-8.3|} & \sigma_\gamma^2 \rho^{|8-12|} \\ \sigma_\gamma^2 \rho^{|8.3-7.1|} & \sigma_\gamma^2 \rho^{|8.3-8|} & 4.2^2 + \sigma_\gamma^2 & \sigma_\gamma^2 \rho^{|8.3-12|} \\ \sigma_\gamma^2 \rho^{|12-7.1|} & \sigma_\gamma^2 \rho^{|12-8|} & \sigma_\gamma^2 \rho^{|12-8.3|} & 2.6^2 + \sigma_\gamma^2 \end{pmatrix}$$

Tabel 3.

Mudel	$\hat{\Theta}$	$D\hat{\Theta}$	AIC	mLOGTOEP	$\hat{\sigma}_\gamma^2, \hat{\rho}, \hat{\sigma}_K^2$
Fikseeritud mõjud	36.75	<0.01	857423.2	428710.6	
Juhuslike mõjudega mudel	33.87	0.97	1245.46	620.73	$\hat{\sigma}_\gamma^2=368.69$
Gaussi funktsioon, kamakas	40.94	0.86	1232.674	612.337	$\hat{\sigma}_\gamma^2=323.43$ $\hat{\rho}=9.64$ $\hat{\sigma}_K^2=343.17$

Antud andmete korral sobis andmetega kõige paremini Gaussi kaugusfunktsiooni ja kamakaefektiga mudel.

5. Valikulisest publitseerimisest tingitud nihe

5.1 Publitseerimisnihe

Viimastel aastatel on eriti kasvanud meta-analüüsi kasutamine terviseuuringutes, sest see võimaldab kirjeldada kokkuvõtlikult seni avaldatud tulemusi. Üheks peamiseks miinuseks on meta-analüüsi tundlikkus uuringute valikulisest publitseerimisest tuleneva nihke suhtes (Duval & Tweedie, 2000).

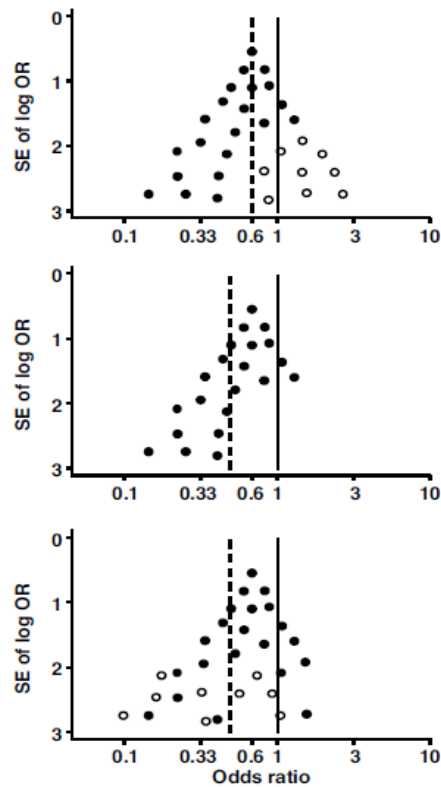
Enamasti avaldatakse artikkel sel juhul, kui jõutakse üsna suure raviefektini (*significance bias*) (Easterbrook *et al.*, 1991; Stern & Simes, 1997). Samuti eelistatakse suuremaid valimeid kasutanud uuringuid (*size bias*). Ka kõikide artiklite leidmine mingil teemal on ebatõenäoline, kuna puudub ühtne andmebaas. Sageli piirduakse vaid inglise keeles avaldatud artiklitega, mistõttu teatud piirkondades tehtud uuringud võivad osutuda alaesindatuks (Rothman *et al.*, 2008). Tulemuste valikulist avaldamist – publitseerimise tõenäosus on seotud uuringutulemuste statistilise olulisusega – hakati nimetama **publitseerimisnihkeks** (*publication bias*) (Rosenthal, 1979). Meditsiiniga seotud valdkondades esineb tulemuste valikulist publitseerimist kahjuks sageli.

5.2 Lehterjoonis

Ilmselt kõige tuntum meetod publitseerimisnihe olemasolu kontrollimiseks on nn **lehterjoonis** (*funnel plot*) (Light & Pillemer 1984; Egger *et al.* 1997). Lehterjoonis on lihtne hajuvusgraafik, kus X-teljel on töötluse mõju hinnangud (saadud artiklitest) ja Y-teljel on mingi kaalumõõt, näiteks standardviga, valimimaht, standardvea pöördväärtus (Light & Pillemer 1984; Egger *et al.*, 1997; Tang & Liu, 2000; Sterne & Egger 2001).

Light ja Pillemer kirjutasid (1984): „Kui kõik valimid/uuringud oleksid võetud samast üldkogumist, siis peaks joonis välja nägema kui lehter, kus valimimahu suurenedes mõjusuurused lähenevad tegelikule väärtusele. Kui tegemist on publitseerimisnihkega, siis on joonis asümmeetriline (joonisel peaks olema auk)“ (Light & Pillemer, 1984). Väiksemate valimite hinnangud on enamasti laiemalt hajunud lehtri allosas. Kui nihet pole, siis joonis peaks olema sümmeetriline (joonis 1, üleval) (Sterne & Harbord, 2004). Kui aga tegemist on nihkega, näiteks väiksemad valimid, kus oli mitteoluline mõju (tühjad ringid joonisel 1), on

jäänud avaldamata, siis sel juhul nähakse mittestümmeetrilist joonist. Näitena toodud joonisel on paremal all (OR=1) ümber – tühjus (joonis 1, keskel) ja sel juhul meta-analüüsiga hinnatud kombineeritud efekt ülehindab töötluse mõju (Egger *et al.* 1997).



Joonis 1 (Sterne & Harbord, 2004).

Hüpoteetilised lehterjoonised: (üleval) sümmeetriline joonis, kus pole nihet (tühjad ringid tähistavad siin väiksemaid uuringuid, mis ei näidanud soovitud efekti); (keskel) asümmeetriline joonis, kus on publitseerimisnihe (väiksemad uuringud, mis ei näidanud soovitud efekti, on puudu); (alumine) asümmeetriline joonis, kus nihe on tingitud/põhjustatud väiksemate uuringute madalalammast metodoloogilisest kvaliteedist (tühjad ringid tähistavad väikeste uuringute tulemusi, mis on kallutatud suuremate efektide suunas).

5.3 Lehterjoonise asümmeetria põhjuseid

Publitseerimisnihe võib põhjustada asümmeetrilisi lehterjooniseid, kuid joonise asümmeetria ei ole meta-analüüsis nihke tõestamiseks piisav. Publitseerimisnihe võib olla vaid üks põhjustest, mis mõjutavad lehterjoonise kuju (Egger *et al.*, 1997).

Mida väiksem on uuring, seda suurem peab tulema mõjuhinnang, et see oleks ka statistiliselt oluline. Suuremate uuringute läbiviimisel on rohkem raha ja ka aega, mis peaks tähendama

ka seda, et uuring on kõrge metoodilise kvaliteediga ja avaldatakse isegi siis, kui tulemus polnud oodatud (Sterne *et al.*, 2000). Asümmeetria võib olla põhjustatud ka töötluse mõju ülehindamisest madalama metoodilise kvaliteediga väiksemates uuringutes (alumine graafik, joonis 1).

Egger *et al.* (1997) tõi välja võimalikud põhjused, miks lehterjoonis on asümmeetriline – erinevad nihked (publitseerimis-, asukoha-, keele-, viidetenihked), heterogeensus, andmete kogumises on erinevused (pettused, vale analüüs, väikeste uuringute kehv metoodika) ja juhus. Lehterjoonise asümmeetriat võib põhjustada näiteks uuringute raviefektide heterogeensus. Vähem täpsed uuringud võivad saada oluliselt teistsuguse efekti võrreldes suuremate uuringutega, kuna nad erinevad omavahel ka uuringudisaini, osalejate valimise, ravivõtete, jne poolest. Seega ei pruugita üldsegi hinnata sama mõju (Berlin *et al.*, 1989; Stuck *et al.*, 1998).

Lisaks eelpool nimetatud põhjustele arvab autor, et lehterjoonise asümmeetriat võib põhjustada ka uuringute omavaheline sõltuvus.

5.4 Lehterjoonise asümmeetriatestid

Lehterjoonise kasutamise puuduseks on selle interpretatsiooni subjektiivsus. Kirjandusest on teada mitmeid meetodeid lehterjoonise asümmeetria testimiseks: Begg'i astak-korrelatsiooni test (Begg & Mazumbar 1994), regressioonil põhinevad testid (Egger *et al.* 1997; Macaskill *et al.*, 2001; Sterne & Egger 2001; Tang, Liu 2000; Peters *et al.*, 2006), jt. Nendest tuntum on Eggeri test (Egger *et al.*, 1997), mida järgnevas tutvustame.

5.4.1. Eggeri test

Lehterjoonise asümmeetria testimiseks soovitasid Egger *et al.* (1997) kasutada lineaarsel regressiooni mudelil põhinevat testi. Esmalt leitakse **standardiseeritud efektid (STE)** (ka *standard normal deviate*) järgnevalt:

$$STE_i = \frac{\hat{\Theta}_i}{\sigma_i} \quad (5.1)$$

ja hinnatakse (vähimruutude meetodiga) regressioonimudel

$$E(STE_i) = a + b \times \frac{1}{\sigma_i},$$

kus σ_i on (artiklis toodud) hinnangu standardviga.

Soovitas lehterjoonise asümmeetria kontrollimiseks testida hüpoteesi: $H_0 : a = 0$. Kui joonistada välja hajuvusgraafik standardiseeritud efektide ja täpsuse jaoks, siis väikeste valimite hinnangud oleks X -teljel nulli lähedal, sest täpsus ($1/\sigma_i$) sõltub päris palju valimimahust. Väikesed uuringud võivad tulemuseks saada ühest erineva šansside suhte või nullist erineva keskmiste erinevuse, aga kuna standardviga on suur, siis standardiseeritud efekt on nulli lähedal. Väikese valimimahuga uuringud on mõlemal teljel nulli lähedal – seega, lähedal nullpunktile. Samas, suurte valimite tulemused on täpsemad ja seetõttu on vastavad STE_i väärtused suuremad. Kui pole publitseerimisnihet, siis punktid asetsevad joonisel joone lähedal, mis läbib nullpunkti ($a=0$) ning tõus b näitab efekti suurust ja suunda (Galbraith, 1988). Sel juhul on lehterjoonis sümmeetriline. Vabaliige a on **asümmeetria mõõde** – mida rohkem vabaliige erineb nullist, seda erinevamad on väikeste uuringute tulemused suurtest uuringutest (Tang & Liu, 2000).

Eelmine test on algebraliselt identne regressioonimudeliga (Sterne *et al.*, 2000), mis saadakse, kui jagada avaldise mõlemad pooled läbi uuringute standardveaga σ_i

$$STE_i = a + b \cdot \frac{1}{\sigma_i} \Big| \sigma_i$$

$$\hat{\Theta}_i = b_0 + b_1 \cdot \sigma_i + \varepsilon_i.$$

Lehterjoonise asümmeetria testimiseks kasutatakse **asümmeetria koefitsenti** (Sterne & Egger, 2001), vahel nimetatakse ka nihkekoefitsendiks (Sterne *et al.*, 2000). Kui $b_1=0$, siis lehterjoonis on sümmeetriline. Mida suurem on asümmeetria koefitsient b_1 , seda suurem on väikeste valimite mõju. Kuna asümmeetriatesti võimsus on väike, siis soovitatakse kasutada olulisusenivood 0.1 (Sterne & Egger, 2001). Samuti on näidatud, et regressioonil põhinev test on võimsam kui Begg'i astak-korrelatsioonitest (Sterne *et al.*, 2000).

5.4.2. Eggeri test fikseeritud mõjudega mudeli jaoks

Järgnevas tutvustatakse Eggeri testi tööpõhimõtet fikseeritud efektidega mudeli (1.1) korral:

$$\hat{\Theta}_i = \Theta + \varepsilon_i.$$

Jagades avaldise mõlemad pooled läbi hajuvusega σ_i , saadakse:

$$\frac{\hat{\Theta}_i}{\sigma_i} = \frac{\Theta}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}.$$

Kasutades Eggeri *et al.* (1997) tähistusi (5.1) ja tähistades $b := \Theta$, omandab mudel kuju :

$$STE_i = a + b \times \frac{1}{\sigma_i} + jääk_i, \quad (5.2)$$

kus $a=0$, kui on fikseeritud mõjudega mudel, $jääk_i := \frac{\varepsilon_i}{\sigma_i}$. Saadud mudelit (5.2) saab hinnata

tavalise vähimruutude meetodiga, sest jäägid on sõltumatud ja sama dispersiooniga:

$$\begin{aligned} D\left(\frac{\varepsilon_i}{\sigma_i}\right) &= \frac{1}{\sigma_i^2} D\varepsilon_i \\ &= \frac{1}{\sigma_i^2} \cdot \sigma_i^2 \\ &= 1, \forall i. \end{aligned} \quad (5.3)$$

Kui aga avaldatud hinnang sõltub uuringu täpsusest, näiteks $E\hat{\Theta}_i = C_0 + C_1 \cdot \sigma_i$, siis mudel (1.1) omandab kuju

$$\hat{\Theta}_i = C_0 + C_1 \cdot \sigma_i + \varepsilon_i.$$

Jagades läbi uuringuhinnangu varieeruvusega σ_i , saadakse

$$\begin{aligned} \frac{\hat{\Theta}_i}{\sigma_i} &= \frac{C_0}{\sigma_i} + \frac{C_1 \sigma_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \\ STE_i &= C_0 \cdot \frac{1}{\sigma_i} + C_1 + jääk_i. \end{aligned}$$

Saadud mudelit saab hinnata taas vähimruutude meetodil, sest jäägid on sõltumatud ja sama dispersiooniga iga i jaoks (5.3). Eggeri testis on tähistused vastavalt: $a := C_1$ ja $b := C_0$.

5.4.3. Eggeri test juhuslike mõjudega mudeli korral

Kui on tegemist juhuslike mõjudega meta-analüüsi mudeliga, siis lisandub mudelisse ka juhuslik efekt γ_i :

$$\begin{aligned}\hat{\Theta}_i &= C_0 + C_1 \cdot \sigma_i + \gamma_i + \varepsilon_i \cdot \sigma_i & (5.4) \\ \frac{\hat{\Theta}_i}{\sigma_i} &= \frac{C_0}{\sigma_i} + C_1 + \frac{\gamma_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \\ STE_i &= C_0 \cdot \frac{1}{\sigma_i} + C_1 + j\ddot{a}k_i.\end{aligned}$$

Heterogeensete uuringute korral pole seega jääkide dispersioon enam konstantne:

$$\begin{aligned}D(j\ddot{a}k_i) &= D\left(\frac{\gamma_i + \varepsilon_i}{\sigma_i}\right) \\ &= \frac{1}{\sigma_i^2} D\gamma_i + \frac{1}{\sigma_i^2} D\varepsilon_i \\ &= \frac{1}{\sigma_i^2} (\sigma_\gamma^2 + \sigma_i^2) \\ &= \frac{\sigma_\gamma^2}{\sigma_i^2} + 1\end{aligned}$$

ning need, kes kasutavad publitseerimisnihke olemasolu/puudumise tõestamiseks Eggeri testi (mis eeldab jääkide sõltumatust ja konstantset dispersiooni), võivad saada valepositiivseid tulemusi või kaotada testi võimsuses. Sarnased probleemid esineksid ka siis, kui osad uuringud on omavahel sõltuvad.

5.5. Modifitseeritud Eggeri test

Kui oleks tegemist juhuslike mõjudega meta-analüüsi mudeliga, siis võiks publitseerimisnihke kontrollimiseks kasutada mudelit (5.4). Publitseerimisnihke olemasolu korral oleks uuringu sõltuvus nähtav hinnangu standardveast (σ_i) ja vastav kordaja oleks nullist erinev, $C_1 \neq 0$. Mudeli parameetrite C_0 , C_1 hindamisel tuleb arvestada uuringute sõltuvuse ja muutuva dispersiooniga ehk tuleks leida suurima tõepära (või üldistatud vähimruutude) hinnang, kasutades andmetega sobivat kovariatsioonimaatriksit \mathbf{V} .

Sõltumatute juhuslike mõjudega mudeli korral oleks:

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 + \sigma_\gamma^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 + \sigma_\gamma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_r^2 + \sigma_\gamma^2 \end{bmatrix}.$$

Kui uuringud on sõltuvad toimumiskoha tõttu, siis võiks kasutada näiteks maatriksit, kuhu on lisatud kamakaefekt:

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 + \sigma_\gamma^2 + \sigma_K^2 & \sigma_\gamma^2 f(d_{12}) & \dots & \sigma_\gamma^2 f(d_{1,r}) \\ \sigma_\gamma^2 f(d_{21}) & \sigma_2^2 + \sigma_\gamma^2 + \sigma_K^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_\gamma^2 f(d_{r-1,r}) \\ \sigma_\gamma^2 f(d_{r1}) & \dots & \sigma_\gamma^2 f(d_{r,r-1}) & \sigma_r^2 + \sigma_\gamma^2 + \sigma_K^2 \end{bmatrix}.$$

Fikseeritud ja juhuslike mõjudega parameetrite olulisuse testimiseks saab kasutada tulemusi lineaarsete segamudelite teooriast. Hüpooteesi

$$H_0 : \boldsymbol{\lambda}^T \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} = 0$$

kontrollimiseks võib kasutada t -statistikut:

$$t = \frac{\boldsymbol{\lambda}^T \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \boldsymbol{\gamma} \end{bmatrix}}{\sqrt{\boldsymbol{\lambda}^T \hat{\mathbf{C}} \boldsymbol{\lambda}}},$$

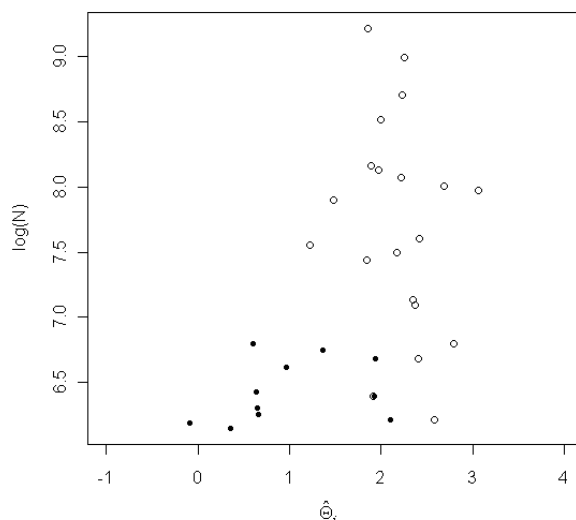
kus $\hat{\mathbf{C}}$ avaldub kujul (2.11) (McLean & Sanders 1988; Stroup, 1989). Kui $\boldsymbol{\gamma}$ ja $\boldsymbol{\varepsilon}$ on normaaljaotusega, siis on statistik t ligikaudu t -jaotusega. Modifitseeritud Eggeri testi korral

$$\boldsymbol{\beta} = \begin{pmatrix} C_0 \\ C_1 \end{pmatrix} \text{ ja } \boldsymbol{\lambda}^T \text{ valitakse selliselt, et } \boldsymbol{\lambda}^T \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} = C_1.$$

Antud t -statistik võimaldab kontrollida hüpooteese juhuslike mõjude meta-analüüsi mudeli jaoks ja võtab arvesse etteantud kovariatsioonimaatriksi. Alternatiivne võimalus on χ^2 -statistik, mis võrdleb kahte fikseeritud mõjudega mudelit, üks on teise erijuht. Viimane arvutatakse just siis, kui võrreldakse erinevate kovariatsioonidega mudeleid (Littell *et al.*, lk 491-502, 1996).

5.6. Simuleerimine

Käesolevas magistritöös simuleeritakse 31 uuringut (vt lisa), millest 11 on omavahel sõltuvad. Antakse ette tegelik keskmine raviefekt ja lisatakse juhuslike efektidega mudelisse juhuslikud mõjud $\boldsymbol{\gamma}$ nii, et 11 suhteliselt väikest valimit kasutatavat uuringut oleks omavahel sõltuvad.



Joonis 2. Simuleeritud andmete lehterjoonis.

Kontrollitakse Eggeri testi abil, kas lehterjoonis on sümmeetriline ehk kas andmetes esineb publitseerimisnihet. Eelnevalt näidati magistritöös, et juhuslike mõjudega mudeli korral ei pruugi Eggeri test õigesti töötada, sest ta jätab arvestamata, et jäägid pole enam konstantsed. Kui test annab olulisuse nivool 0.05 enamal kui 5% juhtudest statistiliselt olulise tulemuse, siis Eggeri test ei tööta juhuslike mõjudega meta-analüüsi mudeli korral, sest andmed genereerisiti ilma publitseerimisnihketa. Kontrollitakse tulemusi ka autori poolt välja töötatud nn modifitseeritud Eggeri testi abil, mis võtab arvesse, et mudeli jääkide hajuvus pole enam konstantne ja jäägid pole enam sõltumatud.

Simuleeritakse andmed 10 000 meta-analüüsi jaoks. Fikseeritud mõjudega mudeli korral võttis Eggeri test ekslikult vastu alternatiivse hüpoteesi 5.2% (95% UI: 4.8...5.7%), juhuslike mõjudega mudeli korral 29% (95% UI: 28...30%), modifitseeritud Eggeri test 4.4% (95% UI: 4.0...4.8%) juhtudest.

Seega, Eggeri test publikatsiooninihke kontrollimiseks ei tööta õigesti, kui on tegemist (sõltuvate) juhuslike mõjudega mudeliga. Käesoleva töö autori poolt välja pakutud test aga ei teinud I-liiki viga lubatust sagedamini.

Kokkuvõte

Meta-analüüs on järjest enam kasutamist leidev meetod juba teostatud uuringute tulemuste üldistamiseks. Maailmas ilmub kümneid tuhandeid teadusajakirju, nendes avaldatakse vahel artikleid, mis käsitlevad sama teemat, mõõdavad sama ravi efektiivsust vms. Uuringutulemused on sageli vastuolulised ja järjest raskem on selles infotulvas orienteeruda. Enne meta-analüüsi meetodite kasutuselevõttu olid levinud artiklite süstemaatilised ülevaated. Meta-analüüsi eeliseks artiklite ülevaate ees on see, et saab hinnata üldist faktorite mõju, tänu tulemuste kombineerimisele on võimalik saavutada suuremat statistilist võimsust ning üldistusvõimet.

Laialdaselt kasutatakse kahte tüüpi meta-analüüsi mudeleid: fikseeritud mõjudega ja juhuslike mõjudega mudelid. Seni avaldatud juhuslike mõjudega meta-analüüsides on eeldatud, et kõik uuringud on juhuslikult ja teineteisest sõltumatult valitud ühest suurest hüpoteetilisest võimalike uuringute populatsioonist. Praktikas aga ei pruugi see nii olla. Mõned uuringud võivad olla läbi viidud sarnastes oludes või sarnast metoodikat kasutades, mõnda populatsiooni on uuritud mitmeid kordi. Käesolevas magistritöös uuritakse, kuidas analüüsida uuringuid olukorras, kus osad meta-analüüsis kasutatud uuringutulemused võiksid olla sõltuvad. Mida lähemal on kahe uuringu toimumiskohad, seda sarnasemad võivad tulemused olla. See võib tulla metoodikate, uuringute läbiviijate teadmiste, valimite, uuritavate (nii geneetilisest, kliimast kui ka käitumisharjumustest tulenevast) sarnasusest. Lisaks uuritakse, kas ruumiline sõltuvus suudaks kirjeldada antud olukorda paremini. Käesolevas magistritöös läbiviidud meta-analüüsi puhul sobis andmetega kõige paremini uuringute ruumilist sõltuvust arvestav mudel.

Meta-analüüs põhineb varem ilmunud artiklite analüüsil. Üheks peamiseks miinuseks on selle tundlikkus uuringute valikulisest publitseerimisest tuleneva nihke suhtes. Enamasti avaldatakse artikkel sel juhul, kui jõutakse üsna suure raviefektini. Ilmselt kõige tuntum meetod publitseerimisnihe olemasolu kontrollimiseks on lehterjoonis. Publitseerimisnihe võib põhjustada asümmeetrilisi lehterjooniseid, kuid joonise asümmeetria ei ole meta-analüüsis nihke tõestamiseks piisav. Publitseerimisnihe võib olla vaid üks põhjustest, mis mõjutavad lehterjoonise kuju. Lehterjoonise asümmeetriat võivad põhjustada erinevad nihked (publitseerimis-, asukoha-, keele-, viidetenihked), uuringute raviefektide

heterogeensus, erinevused andmete kogumises. Magistritöös pakutakse välja, et asümmeetriat võib põhjustada ka uuringute omavaheline sõltuvus. Lehterjoonise kasutamise puuduseks on selle interpretatsiooni subjektiivsus.

Tuntuim meetod lehterjoonise asümmeetria kontrollimiseks on Eggeri test, mis aga ei võta arvesse, et juhuslike mõjudega meta-analüüsi korral pole jäägid alati sõltumatud ja konstantsed. Nii võib saada otsuse, et lehterjoonis on asümmeetriline ehk esineb publitseerimisnihe, kuigi asümmeetria oli põhjustatud hoopis uuringute vahelisest sõltuvusest. Magistritöös töötati välja nn modifitseeritud Eggeri test juhuslike mõjudega mudeli jaoks, mis võtab arvesse, et mudeli jääkide hajuvus pole enam konstantne ja jäägid pole sõltumatud. Simuleeritud andmetega näidati, et Eggeri test publikatsiooninihke kontrollimiseks ei tööta õigesti, kui on tegemist (sõltuvate) juhuslike mõjudega mudeliga. Käesoleva töö autori poolt välja pakutud test publitseerimisnihke kontrollimiseks aga ei teinud I-liiki viga lubatust sagedamini. Kui uuringutulemusi publitseeritakse valikuliselt, siis ei tohiks meta-analüüsi kasutada. Sel juhul oleks meta-analüüsi tulemus nihkega ja võiks viia eksitavate järeldusteni.

Pille Kool

Meta-analysis of dependent studies

Master Thesis

Summary

Meta-analysis is a statistical technique for combining estimated treatment effects from independent comparable clinical trials (studies). Every day a number of articles are published in thousands of journals. As research results accumulate, it becomes increasingly difficult to understand what they tell us. In 1976, Gene Glass proposed a method to integrate and summarize the findings from a body of research. He called the method “meta-analysis”. Meta-analysis is the statistical analysis of a collection of individual studies. Although the term “meta-analysis” was coined relatively recently, statisticians have been using these methods for about 100 years. Such analyses have become increasingly popular in medical research, where information about treatment efficacy is available from a number of clinical trials with inconclusive or inconsistent results.

There are two types of models in meta-analysis: the fixed effect model and the random effects model. The present thesis provides an introduction to both methods. Previously published meta-analyses with random effects assumes that all studies are selected randomly and independently of each other from a large hypothetical population of possible studies. In practice, however, that may not be the case. Some research may be conducted in similar conditions or using similar methodology, and populations have been studied on a number of occasions. The thesis examines how to analyze studies in situations where components used in the meta-analysis of trial results can be correlated.

The correlation between the results of the studies may not always be exactly the same. The closer the two locations of the study are situated, the more similar to each other the results of the study may be. Similarities in the results of two studies may come from methodologies, knowledge, research investigators, and samples (both genetic and climate as well as behavior patterns resulting from the similarity). In addition, we examined whether spatial correlation is able to describe the situation better. The thesis conducted a meta-analysis of the data that best fits the data research model, taking into account spatial dependence.

Meta-analysis is based on a re-analysis of previously published articles. One major disadvantage of meta-analysis is the sensitivity of publication bias – the association of publication probability with the statistical significance of study results.

Funnel plots are a visual tool for investigating publication and other biases in meta-analysis. Publication bias may lead to asymmetrical funnel plots. It is important to realize that funnel-plot asymmetry need not result from bias. The studies displayed in a funnel plot may not always estimate the same underlying effect of the same intervention, and such heterogeneity in results may lead to asymmetry in funnel plots if the true treatment effect is larger in the smaller studies. The present thesis proposes that asymmetry could be caused by dependent studies.

The best-known method to test the asymmetry of funnel plot is Egger's test, but it does not take into account that residuals of the meta-analysis of random effects are not always independent and constant. We have developed the so-called "modified Egger's test" for models of random effects, which takes into account that the variation of residuals is no longer constant and residuals are no longer independent.

The simulated data showed that the Egger's test to detect publication bias is not working correctly if we have a random effects model. A test proposed by the author to test the asymmetry of the funnel plot did not make any more mistakes than those allowed. If the results of studies published on a selective basis are an indication, you should not use meta-analysis. In this case, the meta-analysis of results could lead to biased and misleading conclusions.

Kasutatud kirjandus

1. Akaike H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6): 716–723.
2. Begg C.B., Mazumdar M. (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics*; 50:1088-99.
3. Berlin J.A., Begg C.B., Louis T.A. (1989) An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 84, pp. 381–391.

4. Copas J., Shi J.Q. (2000) Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*. Sep;1(3):247-62.
5. Deeks J.J., Altman D., Bradburn M.J. (2001) Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M., Davey Smith G., Altman D.G., editors. *Systematic reviews in health care: Meta-analysis in context*. 2nd ed. London (UK): BMJ., p. 285–312.
6. DerSimonian R., Kacker, R. (2007) Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials* 28, 105–114.
7. DerSimonian R., Laird N. (1986) Meta-analysis in clinical trials. *Control Clin Trials*; 7 : 177–87.
8. Duval S., Tweedie R. (2000) Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *BIOMETRICS*, 56, 455-463.
9. Easterbrook P.J., Berlin J.A., Gopalan R., Matthews D.R. (1991) Publication bias in clinical research. *Lancet*. 1991 Apr 13;337(8746):867-72.
10. Egger M, Smith G.D., Schneider M., Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315: 629-34.
11. Egger M., Smith G.D. (1998) Bias in location and selection of studies. *BMJ* 316, pp. 61–66.
12. Egger M., Smith G.D. (1995) Misleading meta-analysis. Lessons from „an effective, safe, simple” intervention that wasn't. *Br Med J* 310, pp. 752–754.
13. Egger M., Jüni P., Bartlett C., Holenstein F., Sterne J. (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment* 7: 1–68.
14. Elvik R. (1998) Evaluating the statistical conclusion validity of weighted mean results in meta-analysis by analysing funnel graph diagrams. *Accid Anal Prev* 30, pp. 255–266.
15. Galbraith R. (1988) A note on graphical presentation of estimated odds ratios from several clinical trials. *Stat Med*;7:889-94.
16. Glass G.V. (1976) *Primary, Secondary and Meta-analysis of Research*, Educational Research, 5, 3-8.

17. Goldberg A.S. (1962) Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Assoc.* 57 369-375.
18. Goldstein H. (1995) *Multilevel Statistical Models*, 2nd edn. Kendall's Library of Statistics 3. Edward Arnold: London, U.K.
19. Gonser M., Vetter K., Noack F. (1995) Meta-analyses of interventional trials done in populations with different risks. *Lancet* 345, p. 1304.
20. Graybill FA. (1976) *Theory and Application of the Linear Model*. Wadsworth & Brooks.
21. Hagenau T., Vest R., Gissel T. N., Poulsen C. S., Erlandsen M., Mosekilde L., Vestergaard P. (2009) Global vitamin D levels in relation to age, gender, skin pigmentation and latitude: an ecologic meta-regression analysis. *Osteoporos International* 20:133–140.
<http://www.springerlink.com/content/27735r23x6265162/fulltext.pdf> (01.05.2010)
http://www.springerlink.com/content/27735r23x6265162/MediaObjects/198_2008_6_26_MOESM1_ESM.doc (01.05.2010).
22. Harville D.A. (1977) Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, 72, 320 -338.
23. Harville D.A. (1988) Mixed-Model Methodology: Theoretical Justifications and Future Directions. *Proceedings of the Statistical Computing Section, American Statistical Association, New Orleans*, 41 -49.
24. Harville D.A. (1990) BLUP (Best Linear Unbiased Prediction), and Beyond, in *Advances in Statistical Methods for Genetic Improvement of Livestock*, Springer-Verlag, 239 -276.
25. Higgins J.P., Thompson S.G., Deeks J.J., Altman D.G. (2003) Measuring inconsistency in meta-analyses. *BMJ*. Sep 6;327(7414):557-60.
26. Ioannidis J.P.A., Cappelleri J.C., Lau J. (1998) Meta-analyses and large randomized, controlled trials. *N Engl J Med* 338, p. 59.
27. Irwig L., Macaskill P., Berry G., Glasziou P. (1998) Bias in meta-analysis detected by a simple, graphical test: graphical test is itself biased. *BMJ* 316, p. 740.

28. Iyengar S., Greenhouse, J. B. (1988) Selection models and the file drawer problem (with discussion). *Statistical Science* 3, 109-135.
29. Laird N.M., Ware J.H. (1982) Random effects models for longitudinal data, *Biometrics*, 38, 963-974.
30. Laugesaar R., Kahre T., Kolk A., Uustalu Ü., Kool P., Talvik T. (2010) Factor V Leiden and prothrombin 20210G>A mutation and paediatric ischemic stroke: a case-control study and two meta-analyses. *Acta Paediatrica* 2010; in press. Published Online: March 2, 2010.
31. Light R.J., Pillemer D.B. (1984) Summing up. The science of reviewing research. Cambridge, MA: Harvard University Press.
32. Lindstrom, M. J., Bates, D.M. (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46, 673-687.
33. Littell R. C, Pendergast, J., Natarajan, R. (2000) Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*. 19: 1793-1819.
34. Littell, R. C., Milliken, G.A., Stroup W.W., Wolfinger R.D. (1996) SAS® System for Mixed Models. SAS Institute Inc. Cary, NC.
35. Macaskill P, Walter SD, Irwig L. (2001) A comparison of methods to detect publication bias in meta-analysis. *Stat Med*;20: 641-54.
36. McLean R.A. and Sanders, W.L. (1988) Approximating Degrees of Freedom for Standard Errors in Mixed Linear Models. *Proceedings of the Statistical Computing Section, American Statistical Association, New Orleans*, 50 -59.
37. McLean R.A., Sanders, W.L., and Stroup, W.W. (1991) A Unified Approach to Mixed Linear Models. *The American Statistician*, 45, 54 -64.
38. Pearson K. (1904) Report on certain enteric fever inoculation statistics. *Br Med J*;3:1246.
39. Peters J.L., Sutton A.J., Jones D.R., Abram K.R., Rushton L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *JAMA*;295: 676-80.
40. Prevost T. C., Abrams K. R., Jones D. R. (2000) Hierarchical models in generalized synthesis of evidence: An example based on studies of breast cancer screening. *Statistics in Medicine*. 19: 3359 – 3376.

41. Robinson, G.K. (1991) That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science*, 6, 15 -51.
42. Rosenthal R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin* 86, 85-97.
43. Rothman K.J., Greenland S., Lash T.L. (2008) *Modern epidemiology*. Lippincott Williams & Wilkins, 1k 652-682,
44. Schmid C.H., Lau J., McIntosh M., Cappelleri J.C. (1998) An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med* 17, pp. 1923–1942.
45. Schulz K. F., Chalmers I., Hayes R. J., Altman D. G. (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association* 273(5): 408–412.
46. Searle S.R. (1971). *Linear models*, New York: John Wiley & Sons, Inc.
47. Searle S. R., Casella, G., and McCulloch, C.E. (1992) *Variance Components*, New York: John Wiley & Sons, Inc.
48. Searle S.R. (1997). "The matrix handling of BLUE and BLUP in the mixed linear model". *Linear Algebra and its Applications* 264: 291–311.
49. Stern J. M., Simes R. J. (1997) Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal* 315(7109): 640–645.
50. Sterne J.A.C., Harbord Roger M. (2004) Funnel plots in meta-analysis. *Stata Journal*, StataCorp LP, vol. 4(2), pages 127-141, June.
51. Sterne J.A.C., Egger M. (2001) Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis *Journal of Clinical Epidemiology*, 54 (10), pp. 1046-1055.]
52. Sterne J.A.C., Gavaghan D.J., Egger M. (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 53, pp. 1119–1129.
53. Stuck A.E., Rubenstein L.Z. and Wieland D. (1998) Asymmetry detected in funnel plot was probably due to true heterogeneity. *BMJ* 316, p. 469.

54. Tang J.L., Liu J.L. (2000) Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol*; 53: 477-84.
55. Vandembroucke J.P. (1988) Passive smoking and lung cancer: a publication bias. *BMJ* 296, pp. 391–392.
56. Whitehead A. (2002) *Meta-analysis of Controlled Trials. Statistics in Practice*. John Wiley & Sons, Ltd.

Lisa

Programm 1

Parameetrite hindamiseks kasutame numbrilist meetodit funktsiooni maksimumi leidmiseks. Fikseeritud mudelite korral leiame parameetrite Θ hinnangu $\hat{\Theta}_{GLS}$ üldistatud vähimruutude meetodil (1.3)

#Näide 1 jaoks

```
Y=c(4.1, 1.4)          # efektide vektor
se=c(2.7,1.2)         # hinnangute standardvead
r=length(Y)
# 1) fikseeritud mõjudega mudel. Leian üldistatud vähimruutude meetodi abil
# parameetrite hinnangu:
var=se**2             # ravieffektide hinnangute dispersioonide vektor
V=diag(var)          # diagonaalmaatriksi moodustamine
V_1=solve(V)          # pöördmaatriksi leidmine
X=rep(1,r)            # ühikmaatriks

# vähimruutude hinnang  $\hat{\Theta}_{GLS}$  tegelikule efektile
mu=solve(t(X)**V_1**X)**t(X)**V_1**Y)
# vähimruutude hinnangu dispersioon
d_mu=solve(t(X)**V_1**X)

#  $\hat{\Theta}_{GLS}$  ja 95% usaldusintervall
cat("mu=",round(mu,4)," 95%UI==( ",round(mu- qnorm(1-
0.025)*sqrt(d_mu),4)," ",round(mu+ qnorm(1-0.025)*sqrt(d_mu),4)," ")

#-----
# teine võimalus  $\hat{\Theta}_{GLS}$  arvutamiseks (kasutades elementkuju)
w=1/var              # so sama , mis V_1 peadiagonaalil asuvad arvud.
# Selle saab maatriksist kätte nii, kui
# korrutada t(X)**V_1          w=t(X)**V_1
sum(w*Y)/(sum(w))    # fikseeritud mõjudega mudeli hinnang (valem 1.5)
1/sum(w)              # fikseeritud mõjudega mudeli dispersioon (valem 1.6)
```

Programm 2

Sõltumatute valimite korral juhuslike mõjudega mudelis on vajalik parameetrite hindamisel eeldada, et dispersioon ei ole negatiivne ja lisatakse piirang, mis ei luba proovida negatiivseid dispersioonihinnanguid $\sigma_{\gamma}^2 \geq 0$.

```
Andmed=read.table("naide2.csv",header=T,sep=";", dec=",")
attach(Andmed)
Y=OR                  #so efektide vektor Y
```

```

var=SE**2      #var - ravieffektide hinnangute dispersioonide vektor
r=length(Y)
#OR=c(6.3, 1.6, 4.1, 4.0, 3.7, 3.1, 6.0, 3.6, 1.0)
#SE=c(0.90, 0.65, 0.59,0.42, 0.67, 0.43, 0.36, 0.70, 0.78)

#1) fikseeritud efektidega mudel, sõltumatud valimid. Hinnang  $\hat{\Theta}_{GLS}$  leitakse
kasutades üldistatud vähimruutude meetodit (1.3)
#-----
V=diag(var)      #diagonaalmaatriks
V_1=solve(V)     #pöördmaatriksi leidmine
X=rep(1,r)       #so ühikmaatrik
solve(t(X)**solve(V)**X)**t(X)**solve(V)**(Y)   #so  $\hat{\Theta}_{GLS}$ 
#1*) sama asi kasutades elementide kuju (1.5)

$$w_i = \frac{1}{\sigma_i^2}$$


$$w_i^* = (\hat{\sigma}_i^2 + \hat{\sigma}_\gamma^2)^{-1}$$


$$= (w_i^{-1} + \hat{\sigma}_\gamma^2)^{-1}$$

#-----
w=1/var
DEETA_FIX=(sum(w*Y))/(sum(w))      # fikseeritud mõjudega mudeli hinnang
DEETA_FIX
D_DEETA_FIX =1/(sum(w))            # fikseeritud mõjudega mudeli dispersioon

# väljastab ühes reas hinnangu  $\hat{\Theta}_{FIX}$ , dispersiooni  $D(\hat{\Theta}_{GLS})$  ja usaldusintervalli
cat("DEETA_FIX=",round(DEETA_FIX,4),"D(DEETA_FIX)=",round(D_DEETA_FIX,4),"
95%UI==( ",round(DEETA_FIX -qnorm(1-0.025)*sqrt(D_DEETA_FIX),4)," ",round(DEETA_FIX
+qnorm(1-0.025)*sqrt(D_DEETA_FIX),4)," ")")

#-----prognoosi leidmine-----
d_gamma2=0
G=diag(rep(d_gamma2, r))
V=diag(var) +G      #diagonaalmaatriks
V_1=solve(V)       #pöördmaatriksi leidmine
X=rep(1, r)
BLUP=G**V_1**Y-X*DEETA_FIX
BLUP[6,1]          #Eesti

DEETA_FIX+BLUP[6,1]      #

```

Leiame juhuslike mõjudega mudeli hinnang (momentide meetod)hinnangu Θ^* . Selleks aga peame enne leidma statistiku Q ja ravieffekti varieeruvuse hinnangu $\hat{\sigma}_\gamma^2$.

```

# 2.1) leiame statistiku Q väärtuse (2.14)
Q=sum(w*(Y- DEETA_FIX)**2)

# 2.2) arvutame valemi (2.15) abil hinnangu ravieffekti varieeruvusele
d_gamma2=( Q-(r-1))/( sum(w)-sum(w**2)/sum(w) )   #so  $\hat{\sigma}_\gamma^2$ 
d_gamma2
# 2.3) juhuslike mõjudega mudeli hinnang  $\Theta^*$ 

```

```

wtarn=1/((1/w)+d_gamma2)          # w*
DEETAtarn=sum(Y*wtarn)/sum(wtarn)  # Θ*

# 2.4) hinnangu Θ* dispersioon (2.11): D(DEETAtarn)
D_DEETAtarn=1/sum(wtarn)

# väljastab ühes reas Θ*, D(Θ*), 95% usaldusintervalli hinnangule
cat("DEETAtarn=",round(DEETAtarn,4),"D(DEETAtarn)=",round(D_DEETAtarn,4),
"d_gamma2=",round(d_gamma2,4),"Q=",round(Q,4)," 95%UI==( ",round(DEETAtarn-
qnorm(1-0.025)*sqrt(D_DEETAtarn),4)," ",round(DEETAtarn+qnorm(1-
0.025)*sqrt(D_DEETAtarn),4)," ")

#-----prognosi leidmine-----

G=diag(rep(d_gamma2, r))
V=diag(var) +G          #diagonaalmaatriks
V_1=solve(V)           #pöördmaatriksi leidmine
X=rep(1, r)
BLUP=G%*%V_1%*(Y-X*DEETAtarn)
BLUP[6,1]              #Eesti

DEETAtarn+BLUP[6,1]

#-----standardvea leidmine prognoosile-----
R=diag(var)
k1=t(X)%*%solve(R)%*%X
k2=t(X)%*%solve(R)
k3=solve(R)%*%X
k4=solve(R)+solve(G)

C_hat=solve(rbind(
cbind(k1,k2),
cbind(k3,k4)))

parameeterv=c(DEETAtarn, BLUP[,1])
lambda=c(1,0,0,0,0,0,1,0,0,0)
t(lambda)%*%parameeterv
t(lambda)%*%C_hat%*%lambda
sqrt(t(lambda)%*%C_hat%*%lambda)

```

Juhuslike mõjudega mudeli hinnang (numbriline meetod)

```

logtoep=function(S, mu, eff, var){
var=SE**2
V=diag(S+var)
V_1=solve(V)
l=(-0.5)*log(det(2*pi*V))-0.5*t(eff-mu)%*%V_1%*(eff-mu)
return(l)
}
logtoep(S=0,mu=1,eff, var) #S=0 kui sõltumatud uuringud.
                          #mu=1 anname ette, et üldkeskmise raviefekt on 1
#-----
# R'is on vahend, mis minimiseerib funktsiooni väärtust.
# Sestap minimiseerime miinus log-tõepära väärtust log-tõepära
# väärtuste maksimiseerimise asemel.
mlogtoep=function(arg, eff, var){
S=arg[1]
mu=arg[2]
return(-logtoep(S, mu, eff=eff, var=var))
}

```

```

}
mlogtoep(c(0,1),eff, var)      # ise annan ette, et S=0
#-----
# Funktsioon, mis leiab suurima tõepära hinnangu.
# argumentideks hinnatud raviefektid ja hinnangute dispersioonid
# Soovi korral võib ette anda algväärtused (tähistame theta) parameetritele.
theta=c(1,1)
names(theta)=c("S","mu")
STH=function(eff, var, theta){

# dispersioon ei saa olla negatiivne. Sestap lisame piirangu, mis
# ei luba proovida negatiivseid dispersioonihinnanguid.
# selleks defineerime matriksi ui ja ci, vaata ka käsu constrOptim
# abiinfot!

ui=c(1,0)      #ui*theta - ci>=0→see võrdus kehtib, kui  $\sigma_y^2 > 0$ 
ci=matrix(0)
tulem=constrOptim(theta, mlogtoep, eff=eff, ui, ci, , grad=NULL)

if (tulem$convergence!=0)
print("Vastus pole usaldusväärne - maksimumi ei leitud!!!")
if (tulem$convergence==0)
print("Tõepärafunktsiooni maksimum leiti")
return(tulem)
}
STH(eff=Y, var=SE**2, theta)
U= STH(Y, var, theta)
U$par
d_gamma2_n =U$par[1]
DEETAtarn_n =U$par[2] #juhuslike mõjudega mudeli hinnang  $\Theta^*$ 
w=1/var
wtarn_n=1/((1/w)+d_gamma2_n)
# 4)dispersioon D(DEETAtarn_n)
D_DEETAtarn_n=1/sum(wtarn_n)

# väljastab ühes reas hinnangu OR-e ehk meie tähistustes  $\Theta^*$ , dispersiooni
 $D(\Theta^*)$  ja OR jaoks usaldusintervalli
cat("DEETAtarn numbr=",round(DEETAtarn_n,5),"D(DEETAtarn
numbr=",round(D_DEETAtarn_n,5), "d_gamma2 numbr=",round(d_gamma2_n,4),"
95%UI==( ",round(DEETAtarn_n-qnorm(1-
0.025)*sqrt(D_DEETAtarn_n),4)," ",round(DEETAtarn_n+qnorm(1-
0.025)*sqrt(D_DEETAtarn_n),4)," ")")

#-----prognoosi leidmine-----
G=diag(rep(d_gamma2_n, r))
V=diag(var) +G      #diagonaalmaatriks
V_1=solve(V)      #pöördmaatriksi leidmine
X=rep(1, r)
BLUP=G%%V_1%%(Y-X*DEETAtarn)
BLUP[6,1]      #Eesti

DEETAtarn_n+BLUP[6,1]      #Eestis peaks olema OR=seeeee
d_gamma2_n =2.129
G=diag(rep(d_gamma2_n, r))
X=rep(1, r)
R=diag(var)
k1=t(X)%%solve(R)%%X
k2=t(X)%%solve(R)
k3=solve(R)%%X
k4=solve(R)+solve(G)

```

```

C_hat=solve(rbind(
cbind(k1,k2),
cbind(k3,k4)))
lambda=c(1,0,0,0,0,0,1,0,0,0)
t(lambda)%*%C_hat%*%lambda
sqrt(t(lambda)%*%C_hat%*%lambda)

```

Programm 3

Hindamisel on meil vaja eeldada:

- 1) $\sigma_\gamma^2 \geq 0$;
- 2) $konst < \sigma_\gamma^2$.
- 3) $konst > 0$

Kõik on sarnane eelmisele näitele, ainult anname ette uue maatriksi V.

```

logtoep=function(S, mu, eff, var, konst){
var=SE**2
#konst=0.1
#S=0
mat=matrix(0, ncol=r, nrow=r)      #moodustan nullide maatriksi
for (i in 1:r) {
  for (j in 1:r) {
    if (!(i==j)&(Autor[i]==Autor[j]))
      mat[i,j]=konst
  }
}

V= diag(S+var) + mat
V_1=solve(V)
l=(-0.5)*log(det(2*pi*V))-0.5*t(eff-mu)%*%V_1%*(eff-mu)
return(l)
}
logtoep(S=0,mu=1,eff, var=SE**2, konst=0.1) #S=0 kui sõltumatud uuringud.
#mu=1 arvame, et üldkeskmise rafiefekt on 1,

ise pakume
#-----
# R'is on vahend, mis minimiseerib funktsiooni väärtust.
# Sestap minimiseerime miinus log-tõepära väärtust log-tõepära
# väärtuste maksimiseerimise asemel.
mlogtoep=function(arg, eff, var){
S=arg[1]
mu=arg[2]
konst=arg[3]
return(-logtoep(S, mu, eff=eff, var=var, konst=konst))
}
mlogtoep(c(0,1,0.1),eff, var=SE**2)      # ise annan ette, et S=0
#-----
# Funktsioon, mis leiab suurima tõepära hinnangu.
# argumentideks hinnatud raviefektid ja hinnangute dispersioonid
# Soovi korral võib ette anda algväärtused (tähistame theta)

```

```

# parameetritele.
theta=c(1, 1, 0.5) # 0<kov <  $\sigma_\gamma^2$ 
names(theta)=c("S", "mu", "konst") #või ikka c("s2", "mu")????????
STH=function(eff, var, theta){

# dispersioon ei saa olla negatiivne. Sestap lisame piirangu, mis
# ei luba proovida negatiivseid dispersioonihinnanguid. Kui tahame
# hinnata ka konstanti peame defineerime matriksid ui ja ci nii, et
# kehtiks ui*theta - ci ≥ 0, vaata ka käsu constrOptim abiinfot!
ui=rbind(c(1,0,0),          #      1)  $\sigma_\gamma^2 > 0$ 
          c(1,0,-1),        #      2) konst <  $\sigma_\gamma^2$ 
          c(0,0,1)          #      3) konst > 0
          )
ci=rbind(c(0),c(0),c(0))
tulem=constrOptim(theta, mlogtoep, eff=eff, ui, ci, grad=NULL)

if (tulem$convergence!=0)
print("Vastus pole usaldusväärne - maksimumi ei leitud!!!")
if (tulem$convergence==0)
print("Tõepärafunktsiooni maksimum leiti")
return(tulem)
}
U=STH(eff, var=SE**2, theta)
mlogtoep(c(2.309, 3.671, 2.309),eff, var=SE**2)

U$par
d_gamma2_s =U$par[1]
DEETAtarn_s =U$par[2] #juhuslike mõjudega mudeli hinnang  $\Theta^*$ , s-sõltuvad
uuringud
konst=U$par[3]

#-----95%UI-----
w=1/var
wtarn_s=1/((1/w)+d_gamma2_s)          #w_tärn
# dispersioon D(DEETAtarn_s)
D_DEETAtarn_s=1/sum(wtarn_s)          #dispersioon (valem 2.11)
D_DEETAtarn_s

# väljastab ühes reas hinnangu  $\hat{\Theta}_s^*$ , dispersiooni, 95% usaldusintervalli
cat("DEETAtarn sõltuvad=",round(DEETAtarn_s,5),"D(DEETAtarn
s)=",round(D_DEETAtarn_s,5), "d_gamma2 s=",round(d_gamma2_s,4),"
95%UI==( ",round(DEETAtarn_s-qnorm(1-
0.025)*sqrt(D_DEETAtarn_s),4),",",round(DEETAtarn_s+qnorm(1-
0.025)*sqrt(D_DEETAtarn_s),4),")")

#-----prognoosi leidmine-----
G=diag(rep(d_gamma2_s, r))
V=diag(var) +G          #diagonaalmaatriks
V_l=solve(V)          #pöördmaatriksi leidmine
X=rep(1, r)
BLUP=G%*%V_l%*%(Y-X*DEETAtarn_s)
BLUP[6,1]          #Eesti

DEETAtarn_s+BLUP[6,1]          #Eestis peaks olema OR=seeeee

```

```

#-----standardviga prognoosile-----

d_gamma2_n =2.129

mat=matrix(0, ncol=r, nrow=r)      #moodustan nullide maatriksi
for (i in 1:r) {
  for (j in 1:r) {
    if (!(i==j)&(Autor[i]==Autor[j]))
    mat[i,j]=2.309
  }
}

G=diag(rep(d_gamma2_n, r))+mat
X=rep(1, r)

R=diag(var)
k1=t(X)%*%solve(R)%*%X
k2=t(X)%*%solve(R)
k3=solve(R)%*%X
k4=solve(R)+solve(G)

C_hat=solve(rbind(
cbind(k1,k2),
cbind(k3,k4)))

lambda=c(1,0,0,0,0,0,1,0,0,0)

t(lambda)%*%C_hat%*%lambda
sqrt(t(lambda)%*%C_hat%*%lambda)

#FUNNEL PLOTI JOONISTAMINE
punkte=r
varvid=rep("", punkte)
varvid[1:3]="blue"
varvid[4:5]="yellow"
varvid[6:9]="pink"
plot(Y, 1/var, xlim=c(-2,5), pch=21, col=varvid)
punkte=r
varvid=rep(NA, punkte)
varvid[1:3]=1
varvid[4:5]=20
varvid[6:9]=1
plot(Y, 1/var, xlim=c(-2,5), pch=varvid)

```

Programm 3

```

Andmed=read.table("andmed.csv",header=T,sep=";", dec=",")
attach(Andmed)
Y=keskmine      #tähistame ümber, sest andmestik on tunnuse nimeks KESKMINE
var=SE**2
r=length(Y)

```


Programm 3.1 juhuslike mõjudega mudel

```
#Järgnevas teisendatakse skaala ümber, et R arvutaks välja:
  Andmed=read.table("andmed.csv",header=T,sep=";", dec=",")
attach(Andmed)

Y=keskmine
var=SE**2
r=length(Y)

eff=keskmine
logtoep=function(S, mu, eff, var, skaala=1){
  # S - sigma**2[gamma] - juhuslike efektide dispersioon
  # mu - Theta, keskmine raviefekt
  # roo - juhuslike efektide sõltuvust kirjeldav parameeter
  # eff - uuringutes saadud ravieffektide hinnangud, Theta_hat_i-d
  # var - uuringutes raporteeritud hinnangute standardvead (sigma**2_i)
  # koht - uuringu läbiviimise koht
  # skaala - vaatluste teisendus enne tõepära arvutamist - aitab vältida
numbrilisest ebatäpsusest tingitud lõpmatusi

  eff2=eff*skaala
  var2=var*skaala**2
  S2=S*skaala**2
  mu2=mu*skaala

  r=length(eff2)
  V= diag(S2+var2)
  #V= diag(var2)
  V_1=solve(V)
  l=(-0.5)*log(det(2*pi*V))-0.5*t(eff2-mu2)%*%V_1%*(eff2-mu2)
  return(l)
}
#-----
mlogtoep=function(arg, eff, var, skaala=1){
  S=arg[1]
  mu=arg[2]
  return(-logtoep(S, mu, eff=eff, var=var, skaala=skaala))
}
#-----
STH=function(eff, var, skaala, theta){
  ui=c(1,0)      #ui*theta - ci>=0 see võrdus kehtib, kui
                # st ui=c(S DEETAtarn_n)
  ci=matrix(0)

  tulem=constrOptim(theta, mlogtoep, eff=eff, var=var, skaala=skaala, ui,
ci, grad=NULL)
  if (tulem$convergence!=0)
    print("Vastus pole usaldusväärne - maksimumi ei leitud!!!")
  if (tulem$convergence==0)
    print("Tõepärafunktsiooni maksimum leiti")
  return(tulem)
}
```

```

theta=c(200, 33) #algväärtused leiame nii, et r=50

names(theta)=c("S","mu") #roo - kaugus
U=STH(keskmine, SE**2, 1/20, theta)

mlogtoep(c(S=200,mu=33), keskmine, SE, koht=Latitude, skaala=1/20)
k=2 #mudelis hinnatavate parameetrite arv
AIC=2*k+2*U$value

```

Programm 3.2 juhuslike mõjudega mudel, astmeline kaugusfunktsioon

```

logtoep=function(S, mu, roo, eff, var, koht, skaala=1){
  # S - sigma**2[gamma] - juhuslike efektide dispersioon
  # mu - Theta, keskmine raviefekt
  # roo - juhuslike efektide sõltuvust kirjeldav parameeter
  # eff - uuringutes saadud ravieffektide hinnangud, Theta_hat_i-d
  # var - uuringutes raporteeritud hinnangute standardvead (sigma**2_i)
  # koht - uuringu läbiviimise koht
  # skaala - vaatluste teisendus enne tõepära arvutamist - aitab vältida
  numbrilisest ebatäpsusest tingitud lõpmatusi

  eff2=eff*skaala
  var2=var*skaala**2
  S2=S*skaala**2
  mu2=mu*skaala

  r=length(eff2) #artiklite/uuringute arv
  mat=matrix(0, ncol=r, nrow=r) #moodustatakse nullide maatriks
  kaugus=matrix(0, ncol=r, nrow=r) #moodustatakse abimaatriks
  for (i in 1:r) {
    for (j in 1:r) {
      kaugus[j,i]=abs(koht[j]-koht[i])
      if (!(i==j)) {
        mat[i,j]=S2*roo**kaugus[j,i]
      }
    }
  }
  #so astmeline funktsioon
  # mat[i,j]=S2*exp((-kaugus[j,i]**2)/(roo**2) )
  #so Gaussi funktsioon
  V= diag(S2+var2)+mat
  V_1=solve(V)
  l=(-0.5)*log(det(2*pi*V))-0.5*t(eff2-mu2)%*%V_1%*(eff2-mu2)
  return(l)
}

#-----
mlogtoep=function(arg, eff, var, koht, skaala=1){
  S=arg[1]
  mu=arg[2]
  roo=arg[3]
  return(-logtoep(S, mu, roo=roo, eff=eff, var=var, koht=koht,
skaala=skaala))
}

```

```

#-----
STH=function(eff, var, koht, skaala, theta){

ui=rbind(c(1,0,0),
         c(0,0,1),
         c(0,0,-1)
        )
ci=rbind(c(0),c(0),c(-1))

  tulem=constrOptim(theta, mlogtoep, eff=eff, var=var, koht=koht,
skaala=skaala, ui, ci, grad=NULL)
  if (tulem$convergence!=0)
    print("Vastus pole usaldusväärne - maksimumi ei leitud!!!")
  if (tulem$convergence==0)
    print("Tõepärafunktsiooni maksimum leiti")
  return(tulem)
}

theta=c(400, 35, 0.5) #algväärtused leiame nii, et r=50
names(theta)=c("S","mu","roo") #roo - kaugus
U=STH(keskmine, SE**2, Latitude, 1/20, theta)

```

Programm 3.3 juhuslike mõjudega mudel, astmeline kaugusfunktsioon, kamaka efekt

```

#astmeline kaugusfunktsioon, roo ja kamakaefekti hindamine
logtoep=function(S, mu, roo, kamakas, eff, var, koht, skaala=1){
  # S - sigma**2[gamma] - juhuslike efektide dispersioon
  # mu - Theta, keskmine ravieffekt
  # roo - juhuslike efektide sõltuvust kirjeldav parameeter
  # eff - uuringutes saadud ravieffektide hinnangud, Theta_hat_i-d
  # var - uuringutes raporteeritud hinnangute standardvead (sigma**2_i)
  # koht - uuringu läbiviimise koht
  # skaala - vaatluste teisendus enne tõepära arvutamist - aitab vältida
numbrilisest ebatäpsusest tingitud lõpmatusi

  eff2=eff*skaala
  var2=var*skaala**2
  S2=S*skaala**2
  mu2=mu*skaala
  kamakas2=kamakas*skaala**2

  r=length(eff2) #artiklite/uuringute arv
  mat=matrix(0, ncol=r, nrow=r) #moodustan nullide maatriksi
  kaugus=matrix(0, ncol=r, nrow=r) #moodustan abimaatriksi
  for (i in 1:r) {
    for (j in 1:r) {
      kaugus[j,i]=abs(koht[j]-koht[i])
      if (!(i==j)) {
        mat[i,j]=S2*roo**kaugus[j,i]
      }
    }
  }
#so astmeline funktsioon
  #mat[i,j]=S2*exp((-kaugus[j,i]**2)/(roo**2) )
#so Gaussi funktsioon
}

```

```

}
V= diag(S2+var2+kamakas2)+mat
V_1=solve(V)
l=(-0.5)*log(det(2*pi*V))-0.5*t(eff2-mu2)%*%V_1%*(eff2-mu2)
return(l)
}
mlogtoep=function(arg, eff, var, koht, skaala=1){
  S=arg[1]
  mu=arg[2]
  roo=arg[3]
  kamakas=arg[4]
  return(-logtoep(S, mu, roo=roo, kamakas=kamakas, eff=eff, var=var,
koht=koht, skaala=skaala))
}

#-----
STH=function(eff, var, koht, skaala, theta){
  ui=rbind(c(1,0,0,0),      # 1)
           c(0,0,1,0),    # 2)
           c(0,0,-1,0),   # 3)
           c(0,0,0,1)     # 4) kamakas>=0
          )
  ci=rbind(c(0),c(0),c(-1), c(0))
  tulem=constrOptim(theta, mlogtoep, eff=eff, var=var, koht=koht,
skaala=skaala, ui, ci, grad=NULL)
  if (tulem$convergence!=0)
    print("Vastus pole usaldusväärne - maksimumi ei leitud!!!")
  if (tulem$convergence==0)
    print("Tõepärafunktsiooni maksimum leiti")
  return(tulem)
}

theta=c(400, 35, 0.05, 1) #algväärtused leiame nii, et r=50
names(theta)=c("S", "mu", "roo", "kamakas") #roo - kaugus
STH(keskmine, SE, Latitude, 1/20, theta)

```

Programm 4

Andmete simuleerimine, modifitseeritud Eggeri test

```

m=1000 #korduste arv
p=rep(NA,m) #p-value vektor (Egger'i test)
p_modif=rep(NA,m) #p-value vektor (modifitseeritud test)
N1=c(10000,8000,6000,5000,3000,3500,2900,2700,2000,3200,3400,1800,1700,190
0,1250,1200,900,800,600,500)
N2=c(800,850,900,750,500,550,520,490,470,620,600)
N=c(N1,N2)
r1=length(N1)
r2=length(N)-r1

```

```

r=r1+r2
G=diag(rep(0.5**2, r))          # valitud  $\sigma_i^2$ 
  for (k in (r1+1):r) {
    for (h in (r1+1):r)
      {
        G[k,h]=0.5**2
      }
  }
G=G+diag(rep(1e-10,r))
for (i in 1:m) {
gamma=c(rnorm(r1, sd=0.5), rep(rnorm(1, sd=0.5), r2) )
Theta=2                          #keskmine efekt  $\Theta$ 
Theta_i=Theta+gamma              #tegelikud  $\Theta_i = \Theta + \gamma_i$ 
s=10                              # uuritava tunnuse standardhälve
standardviga=s/sqrt(N)
hinnanguviga=rnorm(r, sd=standardviga)      # so epsilon.
theta_hat_i=Theta_i+hinnanguviga
tapsus=1/standardviga
STE=theta_hat_i/standardviga
p[i]=coef(summary(lm(STE~tapsus)))[1,4]      #Egger'i testi p-value

#=====

G=diag(rep(0.5**2, r))          #vabalt valitud hajuvus
  for (k in (r1+1):r) {
    for (h in (r1+1):r)
      {
        G[k,h]=0.5**2
      }
  }
G=G+diag(rep(1e-10,r))

for (i in 1:m) {
gamma=c(rnorm(r1, sd=0.5), rep(rnorm(1, sd=0.5), r2) )
Theta=2                          #keskmine efekt
#Theta_i=Theta                  #kui on fikseeritud mõjudega mudel
Theta_i=Theta+gamma              #tegelikud deeta_i, mida me kunagi ei näe
                                  #lisame mõõtmisvea epsilon_i
                                  #juh efektidega mudel
s=10                              # uuritava tunnuse standardhälve
standardviga=s/sqrt(N)
hinnanguviga=rnorm(r, sd=standardviga)
theta_hat_i=Theta_i+hinnanguviga
tapsus=1/standardviga
STE=theta_hat_i/standardviga
p[i]=coef(summary(lm(STE~tapsus)))[1,4]      #Egger'i testi p-value

#=====
#modifitseeritud Eggeri test

lambda=c(0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0 )

```

```

X=cbind(rep(1, r),standardviga)
R=diag(standardviga**2)

V=R +G
V_1=solve(V)
k1=t(X)%*%solve(R)%*%X
k2=t(X)%*%solve(R)
k3=solve(R)%*%X
k4=solve(R)+solve(G)

C_hat=solve((rbind(
cbind(k1,k2),
cbind(k3,k4))))

beeta_hat=solve(t(X)%*%solve(V)%*%X)%*%t(X)%*%solve(V)%*%(theta_hat_i)
BLUP=G%*%V_1%*(theta_hat_i-X%*%beeta_hat)
parameetrid=c(beeta_hat,BLUP)
C1_hat=t(lambda)%*%parameetrid
C1_hat
t=C1_hat/(sqrt(t(lambda)%*%C_hat%*%lambda))
p_modif[i]=2*pt(-abs(t), df=r-2)
}
table(p<0.05)
table(p_modif<0.05)

```