

TARTU ÜLIKOOL

Sotsiaalteaduste valdkond

Ühiskonnateaduste instituut

Ühiskonna ja infoprotsesside analüüs õppekava

Karlis Põldoja

**Järeldava analüüsi eeldustele vastavus ja statistiliste järelduste adekvaatsus
valitud Haridus- ja Teadusministeeriumi tellitud rakendusuringutes aastatel
2013–2024**

Magistritöö

Juhendaja: Indrek Soidla, MA

Tartu 2025

AUTORIDEKLARATSIOON

Olen koostanud töö iseseisvalt. Kõik töös kasutatud teiste autorite tööd, põhimõttelised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

Karlis Põldoja

27.05.2025

SISUKORD

SISSEJUHATUS	5
1. KIRJANDUSE ÜLEVAADE	9
1.1 Üldine teoreetiline raamistik	9
1.2 Järeldava analüüsi eeldused.....	10
1.2.1 Kaetuse viga	10
1.2.2 Valikumeetodid ja valiku viga	12
1.2.3 Mittevastamise viga.....	14
1.2.4 Kohandamise viga	16
1.3 Statistilised järeldused	18
1.3.1 P-väärtus ja NHST.....	18
1.3.2 P-väärtusega seotud levinud väärtõlgendused.....	19
1.3.3 Usaldusvahemikud	22
2. PROBLEEMISEADE.....	25
3. METOODIKA	27
3.1 Valim.....	27
3.2 Meetod.....	29
3.3 Kodeerimisjuhendi koostamine	31
4. TULEMUSED	34
4.1 Analüüsiüksused	34
4.2 Järeldava analüüsi eeldused.....	35
4.2.1 Valimimeetodi sõnastus	37
4.2.2 Valimi suurus ja valimiraam	37
4.2.3 Planeeritud ja saavutatud valim	38
4.2.4 Osalejatega kontakteerumine ja valikukriteeriumid	39
4.2.5 Sihtpopulatsiooni määratlus ja valimi esinduslikkus	40
4.2.6 Kaalumise ja andmete imputeerimine	41
4.2.7 Osalusmäära suurendamise strateegiad	41
4.2.8 Kvoodid ja paneelid.....	42
4.2.9 Vastamis- ja osalemismäärad.....	43
4.2.10 Uuringudisaini piirangud.....	43
4.2.11 Järeldava analüüsi eelduste kokkuvõte	44

4.3 Järeldav analüüs.....	46
4.3.1 Statistiliste tulemuste esitamise täpsus	47
4.3.2 P-väärtuste väärtõlgendamise vältimine	48
4.3.3 Usaldusvahemike esitamine võtmetulemustes	49
4.3.4 Usaldusvahemike väärtõlgendamise vältimine	49
4.3.5 Statistiliste järelduste ja analüüsitulemuste kooskõla.....	50
4.3.6 Statistiliselt mitteoluliste tulemuste tõlgendamine	50
4.3.7 Statistilise ja sisulise olulisuse eristamine	51
4.3.8 Mudeli muutujate kaasamise ja väljajätmise põhjendatus.....	52
4.3.9 Koefitsientide esitamise selgus.....	53
4.3.10 Sisuline ja praktiline olulisus.....	53
4.3.11 Statistiliste olulisuste võrdlemise vältimine	54
4.3.12 Järeldava analüüsi kokkuvõte	55
5. ARUTELU JA JÄRELDUSED	56
5.1 Järeldava analüüsi eelduste hinnangud.....	58
5.2 Järeldava analüüsi hinnangud	61
5.3 Praktilised soovitusel	65
5.4 Töö piirangud	67
5.5 Edasised uurimissuunad	68
KOKKUVÕTE	69
SUMMARY	70
KASUTATUD KIRJANDUS	71
LISAD	74
Lisa 1. Valim.....	74
Lisa 2. Kodeerimisjuhend.....	77
Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks	84

SISSEJUHATUS

Statistiline analüüs on andmepõhise teadustöö keskpunkt, pakkudes süsteemset viisi keeruliste sotsiaalsete nähtuste mõtestamiseks. Statistiline analüüs jaguneb kirjeldavaks ja järeldavaks statistikaks. Järeldav statistika hõlmab meetodeid, mis võimaldavad valimist saadud tulemuste põhjal teha üldistusi kogu populatsiooni kohta. Selle kaudu püütakse otsustada, kas andmetes nähtavad mustrid on populatsioonile laiendatavad või lihtsalt juhuslikud.

Eriti oluline on järeldava statistika korrektne rakendamine rakendusuringutes, mille tulemusi kasutatakse poliitika kujundamises. Sellistes kontekstides tehakse järelduste alusel otsuseid, mis mõjutavad laiemalt ühiskonda. Seetõttu peavad järeldava statistika tulemused põhinema metoodiliselt korrektsetel ja läbipaistvalt esitatud andmetel, et tagada analüüsi usaldusväärsus ja üldistatavus.

Käesoleva uurimistöö eesmärk on hinnata valitud Haridus- ja Teadusministeeriumi tellitud rakendusuringute, milles on kasutatud järeldavat analüüsi, metoodilist kvaliteeti, keskendudes sellele, kuivõrd on nendes täidetud järeldava statistika kasutamiseks vajalikud eeldused ning kas statistilised järeldused on esitatud selgelt ja sisukalt, võimaldades teha usaldusväärseid ja põhjendatud üldistusi.

See uurimistöö keskendub just Haridus- ja Teadusministeeriumi tellitud rakendusuringutele, kuna need on otseselt seotud ühiskondlikult kriitiliste aspektidega nagu Eesti haridus-, teadus- ja noortepoliitika kujundamisega. HTM (2025) on rõhutanud oma eesmärki toetada teadmispõhist poliitikakujundamist, milleks tegeletakse järjepidevalt analüüsimise, uurimise ja hindamisega ning tellitakse vajadusel uuringuid ka ministeeriumist väljastpoolt. Sellises kontekstis on oluline mõista, kuivõrd vastavad nende uuringute raportite metoodilised ja analüütilised praktikad kaasaegsete järeldava statistika nõuete ja usaldusväärse poliitikaanalüüsi standarditega. Käesolev töö annab selles osas süsteemse ja metoodiliselt rangelt struktureeritud hinnangu.

Valimiteooria aluspõhimõtted on statistilises kirjanduses hästi formaliseeritud ja pakuvad teoreetilise aluse usaldusväärsete üldistuste tegemiseks, aga nende rakendamisega kaasneb sotsiaalteaduslikus kontekstis mitmeid praktilisi ja teoreetilisi piiranguid. Sihtpopulatsioonide keeruline määratlemine, raskesti ligipääsetavad populatsioonigrupid ja piiratud ressursid muudavad klassikaliste tõenäosuslike valikumeetodite rakendamise sageli keerukaks või üldse võimatuks.

Paljud sotsiaalteaduslikud uuringud seisavad silmitsi puudulike valikuraamade probleemiga, kus ametlikud registrid ei kata olulisi sihtpopulatsiooni rühmi või andmestikud on vananenud.

Sotsiaalteaduslikus uurimiskeskkonnas on sageli piiratud võimalused kallite ja ajamahukate tõenäosuslike valimimeetodite rakendamiseks, mistõttu eelistatakse praktikas mugavusvalimeid, kvootvalimeid või veebipaneelidel põhinevaid mittetõenäosuslikke valimeid. Hoolimata sellest rakendatakse mittetõenäosuslike valimite korral sageli järeldava statistika meetodeid, sealhulgas hüpoteeside testimist ja usaldusvahemike esitust. Kuna selliste valimite korral ei ole üldjuhul teada tõenäosus, millega üksused valimisse satuvad, on oluliselt keerulisem hinnata, kuivõrd valim tegelikult sihtpopulatsiooni esindab. Seetõttu võivad selliste meetodite tulemused olla piiratud usaldusväärsusega ning vajavad tõlgendamisel suuremat ettevaatlikkust.

Praktilised raskused, sealhulgas uuritavate kättesaadavus, motivatsiooni puudumine ja informeeritud nõusoleku saamine, mõjutavad samuti uuringu valimi kujunemist ja suurendavad mittevastamise viga. Uuringutest võivad süsteemselt kõrvale jääda madalama sotsiaalmajandusliku staatusega isikud, vähemusgrupid ja teised raskesti kaasatavad rühmad, mille tulemuseks on valimi kallutatatus ja sihtpopulatsiooni ebapiisav esindatus.

Kuna sellised probleemid mõjutavad otseselt järelduste usaldusväärsust, keskendub töö teoreetiline raamistik esmalt valimi esinduslikkuse küsimusele, mis on järeldava analüüsi keskne eeldus. Kui valim ei peegelda sihtpopulatsiooni olulisi omadusi, ei saa saadud tulemusi sellele üldistada, ükskõik kui täpselt oleks tehtud statistilised arvutused. See toob kaasa vajaduse hinnata, kuidas on defineeritud sihtpopulatsioon, kuivõrd hästi seda katab valikuraam, ja milliseid valikumeetodeid on kasutatud. Tõenäosuslik valikumeetod võimaldab hinnata tulemuste statistilist täpsust, mittetõenäosusliku meetodi puhul aga tuleb olla ettevaatlik, kuna puudub teadmine valimisse sattumise tõenäosuste kohta.

Märkimisväärne on kaetuse viga ehk olukord, kus osa sihtpopulatsioonist jääb valimist välja, kas valikuraami puudulikkuse või valimimeetodi kallutatuse tõttu. Kaetuse vea määratlemine eeldab valimi ulatuse ja sihtpopulatsiooni täpset kirjeldust, mille puudumine uuringuraportites muudab analüüsi tulemuste hindamise problemaatiliseks. Teine kriitiline aspekt on valikuvea ja mittevastamise vea käsitlemine. Valikuviga on tingitud juhuslikust veast, aga mittevastamise viga viitab süsteemsetele erinevustele vastajate ja mittevastajate vahel, mis võib viia kallutatud tulemuste ja ekslike järeldusteni.

Kuna andmetes esineb sageli puudulikke väärtusi, on oluline see, kuidas uuringutes käsitletakse andmete puudumist. Töö raames vaadeldakse, kas raportites kasutatakse andmetöötluse meetodeid nagu imputeerimine, ehk puuduvate andmete asendamist tõenäoliste väärtustega, ning kuivõrd sellised meetodid on dokumenteeritud ja põhjendatud. Samuti hinnatakse, kas uuringutes on

rakendatud kaalumist, mis aitab korrigeerida valimi eripärasid, et see vastaks paremini sihtpopulatsiooni jaotusele. Kaalumise läbipaistmatu või põhjendamata kasutamine võib viia andmete kallutatuseni või suurendada juhuslikku viga.

Lisaks valimiga seotud eeldustele pööratakse käesolevas töös tähelepanu statistiliste tulemuste esitamisele ja tõlgendamisele, mis on rahvusvaheliselt uuringutes sageli probleemseks osutunud. Töös käsitletakse põhjalikult p-väärtuse rolli ja selle levinumaid väärtõlgendusi, nagu näiteks väärarusaam, et madal p-väärtus kinnitab automaatselt alternatiivhüpoteesi või et see näitab kuidagi seose tugevust. Sellised väärtõlgendused on laialt levinud ja võivad uuringutulemusi tõsiselt kallutada. Hinnatakse ka usaldusvahemike kasutamist võtmetulemuste esitlemisel ja nende tõlgendamise korrektsust.

Töö keskendub kõigis nende aspektides mõistetele nagu valim, sihtpopulatsioon, valikumeetod, kaetus, esinduslikkus, statistiline ja sisuline olulisus, p-väärtus ning usaldusvahemik. Need on kontseptuaalsed tugipunktid, mille korrektne rakendamine määrab, kas järeldav analüüs on metodoloogiliselt usaldusväärne. Magistritöös on nendest mõistetest lähtuvalt loodud kodeerimisjuhend, mille abil hinnatakse süstemaatiliselt Haridus- ja Teadusministeeriumi tellitud rakendusuuringu metoodilist ja analüütilist kvaliteeti. Kodeerimisjuhend võimaldab tuvastada nii tehnilistest puudustest kui ka ekslikest tõlgendustest tulenevaid probleemkohti.

Rahvusvahelises teaduskirjanduses on järeldava statistika metoodiliste eelduste täitmise ja statistiliste järelduste adekvaatsuse probleemistikule pööratud tähelepanu mitmel tasandil. Ioannidis (2005), Amrhein jt (2019) ja Wasserstein jt (2019) on toonud esile, et ekslikud järeldused ei tulene üksnes andmete kvaliteedist, vaid sageli ka järeldava statistika mehaanilisest kasutamisest, tõlgendusvigadest ja teadusprotsessi mitmetähenduslikest ning läbipaistmatutest sammudest. Selle tulemusena on viimastel aastatel kasvanud surve suurendada statistiliste tulemuste tõlgendamise selgust ja standardiseeritust, eriti seoses p-väärtuste ja usaldusvahemike esitamisega. Samuti on uuringud, näiteks Hoekstra jt (2014) näidanud, et isegi kogenud uurijad ja õppejõud eksivad sageli nende näitajate tõlgendamisel.

Eestis ei ole seni süvitsi tegeletud nende probleemide teadusliku analüüsiga. Märkimisväärse eelneva tööna võib välja tuua autori bakalaureusetöö, millele käesolev magistritöö otseselt toetub ja mis tõi esile mitmeid metoodilisi puudusi HTM-i tellimusel koostatud uuringuraportites (Põldoja, 2023). Käesolev töö püüab seda lünka osaliselt täita, pakkudes rakendusuuringu metoodilise ja analüütilise kvaliteedi hinnangut järeldava statistika kasutamise vaates. Samuti aitab töö kaasa arutelule, kuidas statistiliste meetodite rakendamine ja tulemuste tõlgendamine

võiks olla metoodiliselt põhjendatum ja selgem. Analüüsi tulemused annavad ülevaate, millised on senised praktikad ning millistes aspektides oleks võimalik uuringute kvaliteeti parandada.

Uurimisühikuteks on käesolevas töös 30 HTM-i tellimusel aastatel 2013–2024 läbi viidud valimipõhise rakendusuuringu raportit, milles on kasutatud järelavat analüüsi.

Kogutud andmete analüüsimiseks on kasutatud kontentanalüüsi meetodit. Loodud on spetsiifiline kodeerimisjuhend, mille abil hinnatakse uuringuraportite metoodilist läbipaistvust ja analüüsi adekvaatsust. Kodeerimisjuhend katab kõiki käesolevas sissejuhatuses tutvustatud teoreetilisi ja metoodilisi aspekte, spetsiifilisemalt on need kirjeldatud kirjanduse ülevaate peatükis. Kodeerimisprotsessis on arvesse võetud kogu uuringuraport, sealhulgas lisad, tabelid ja joonised, et tagada maksimaalne informatiivsus ja anda võimalikult terviklik hinnang uuringute metoodilisele ja analüütilisele kvaliteedile. Terve kodeerimisjuhend koos hindamiskriteeriumite kirjeldustega on leitav lisast 2.

Uurimiseesmärgi saavutamiseks olen püstitanud kaks uurimisküsimust:

1. Kuivõrd on Haridus- ja Teadusministeeriumi tellimusel läbi viidud rakenduslikes uuringutes kasutatavate andmete puhul täidetud järelava analüüsi eeldused?
2. Kuivõrd on Haridus- ja Teadusministeeriumi tellimusel läbi viidud rakenduslikes uuringutes tehtud statistilised järelused selgelt ja adekvaatselt sõnastatud?

Magistritöö koosneb viiest peatükist. Esimene peatükk annab ülevaate töö teoreetilisest raamistikust ja käsitletavatest mõistetest. Teises peatükis sõnastatakse probleem ja uurimisküsimused. Kolmandas peatükis kirjeldatakse valimit, metoodikat ja kodeerimisjuhendi koostamist. Neljandas peatükis esitatakse kodeerimise tulemused. Viiendas peatükis väljendatakse uuringu järelusi ja arutletakse tulemuste tähenduse üle laiemas kontekstis, esitatakse soovitusel ja käsitletakse töö piiranguid ning edasise uurimissuuniseid.

Autor soovib tänada oma juhendajat Indrek Soidlat väärt nõuannete, kannatlikkuse, mõistva suhtumise ja inspiratsiooni eest. Samuti soovin tänada oma magistritöö retsensenti Liina-Mai Toodingut, kes andis asjatundlikku tagasisidet ja nõu.

1. KIRJANDUSE ÜLEVAADE

Antud peatüki eesmärk on anda ülevaade uurimistöös kasutuses olevatest terminitest ja eripäradest ning luua teoreetiline alus, millele hilisemates peatükkides tehtavad analüüsid ja järeldused tuginevad. Ülevaade algab töö üldise teoreetilise raamistiku selgitamisega. Järgnevalt on kirjeldatud, missugused on järeldava analüüsi eeldused ja mida tähendavad töös kasutatud terminid. Sektsiooni viimane osa kirjeldab p-väärtuste ja usaldusvahemike termineid ning levinud väärtõlgendusi, mida tehakse p-väärtuste ja usaldusvahemike rakendamisel igasugustes rakenduslike uuringute raportites.

1.1 Üldine teoreetiline raamistik

Statistiline analüüs on protsess, mille abil püütakse andmete kaudu mõista tegelikkust, ehkki andmed ise peegeldavad seda sageli vaid osaliselt ning võivad olla ebatäielikud või ebatäpsed (McElreath, 2020: 35, 545, 553). Statistiline analüüs jaguneb kaheks peamiseks haruks: kirjeldav statistika, mille eesmärk on kokkuvõtlikult esitada andmete omadusi, nagu keskmised ja hajuvus, ilma üldistusi tegemata ning järeldav statistika, mis võimaldab teha valimi põhjal järeldusi kogu populatsiooni kohta, tuginedes tõenäosuslikele meetoditele (Clark jt, 2021: 170, 318).

Sotsioloogilistes uuringutes, kus keerulisi sotsiaalseid nähtusi uuritakse sageli küsitluste, katsete või vaatlusandmete abil, võimaldavad statistilised analüüsimeetodid uurijatel tuvastada mustreid, testida hüpoteese ja hinnata põhjuslikke seoseid. Näiteks regressioonanalüüs võimaldab kvantitatiivselt uurida muutujate vahelisi seoseid, näiteks hariduse mõju sissetulekute ebavõrdsusele. Ilma statistilise analüüsita oleks igasuguste järelduste tegemine väga piiratud, sest valimi põhjal populatsiooniparameetrite hinnangute leidmine ja tulemuste usaldusväärne üldistamine oleks võimalik vaid intuiitivsete või anekdootlike näidete alusel, mis võib viia ebausaldusväärsete üldistusteni.

Järeldava statistika rakendamine eeldab mitmete metoodiliste tingimuste täitmist, et analüüsitulemused oleksid usaldusväärsed ja üldistatavad. Kui need eeldused ei ole täidetud, võib saadud statistiline tulemus olla eksitav või mitteüldistatav sihtpopulatsioonile (Warner, 2013: 5-6). Eelduste rikkumine võib ilmneda mitmel viisil, mistõttu on uuringuaruannetes oluline selgelt tervet uuringu protsessi dokumenteerida, et lugejal oleks aru saada, kuivõrd hästi järeldava analüüsi eelduseid täideti ning milliseid meetmeid kasutati võimalike probleemide leevendamiseks.

Olen võtnud järgnevaks „Järeldava analüüsi eeldused“ sektsiooniks suure hulga teksti omaenda bakalaureusetööst, et kirjeldada järeldava analüüsi eelduseid (Põldoja, 2023). Bakalaureusetöös kasutasin spetsiifiliselt küsitlusuuringute jaoks uuringu koguvea (*Total Survey Error*) raamistikku, antud raamistikku saab rakendada ka muudele andmetele peale küsitlusandmete, seega olen kohandanud bakalaureusetööst võetud teksti, et see oleks laiemalt sobilik kõigile valimipõhistele rakenduslikele uuringutele.

1.2 Järeldava analüüsi eeldused

Järeldav statistika võimaldab teha järeldusi populatsiooni kohta valimi põhjal. Selliste järelduste usaldusväärsus sõltub otseselt sellest, kuidas valim on moodustatud, kui hästi see esindab sihtpopulatsiooni ning kuidas on käsitletud andmekogumise käigus tekkinud puudujääke. Seetõttu on käesolevas töös oluline analüüsida, kas uuritavates rakendusuuringutes on valimi koostamise ja kvaliteedi aspektid piisavalt selgelt ja adekvaatselt esitatud. Järgnevalt käsitlen valimi moodustamisega seotud küsimusi, sealhulgas valimi ja sihtpopulatsiooni suhet, valikumeetodeid ning valimi suuruse määramise olulisust. Seejärel pööran tähelepanu andmekogumise protsessile ja valimi kvaliteedile, sealhulgas valimiraami määratlemisele, valikukriteeriumidele ning kontakt- ja värbamismeetodite võimalikele mõjudele. Lõpuks käsitlen valimi korrigeerimist ja puudulike andmete käsitlemist, sealhulgas valimi kaalumist ning vastamis- ja osalemismäärade rolli.

1.2.1 Kaetuse viga

Üldkogum ehk populatsioon on kogum, mille kohta soovitakse saada tulemusi, hinnanguid, kontrollida hüpoteeses, luua mudeleid ja teha järeldusi (Tiit ja Tooding, 2019: 297). Kuna üldkogumi defineerimisest lähtuvad kõik edasised valimit puudutavad sammud, siis on üldkogumi täpne defineerimine kriitilise tähtsusega. Seetõttu on ka väga oluline, et uuringu meetodika kirjeldamisel oleks üldkogum võimalikult täpselt defineeritud, sest see määrab ära, kellele on üleüldse võimalik uuringu tulemusi üldistada. Mida ebamäärasemalt või üldisemalt üldkogum defineeritakse, seda keerulisem on uuringu ja selle põhjal tehtud järelduste kvaliteeti hinnata. Sihtpopulatsioon on konkreetset määratletud suur mitmete indiviidide rühm, millest uurija võtab valimi ja millele üldistatakse valimi põhjal saadud tulemused (Neuman, 2011: 252).

Põhimõtteliselt ei ole sihtpopulatsioon erinev üldkogumist ja seega neid käsitletakse tihti sünonüümidenä. Statistikas käsitletakse sihtpopulatsiooni kui statistile ülesande jaoks täpsustatud

üldkogumit (Tiit ja Tooding, 2019: 226). Sihtpopulatsioonid on kindla suurusega, mida vähemalt teoreetiliselt saab kokku lugeda, neil võivad olla mõned ajapiirangud (eksisteerivad kindlaksmääratud aja jooksul) ja on jälgitavad, mis tähendab, et neile on juurdepääs (Groves jt, 2009). Valikuraam on populatsiooni liikmete loend (Neuman, 2011: 252). Valikuraamid oma kõige lihtsamas vormis on lihtsalt populatsioonielementide loendid. Paljude populatsioonide puhul on valikuraamid kergelt kättesaadavad, näiteks kutseorganisatsioonide liikmed, ärid konkreetses linnas või mõnes maakonnas asuvad asutused, haiglad ja koolid (Groves jt, 2009).

Statistilise mõttes on kaetuse viga kirjeldatav kahe aspektina: kui suure osa moodustab sihtpopulatsioonist see osa, mida valikuraam ei kata ja sellega, kuivõrd erineb populatsiooni kaetud osa katmata osast. Kaetuse viga puudutab valikuraami ja sihtpopulatsiooni, see eksisteerib enne valimi koostamist ja seega ei ole probleem, mis esineb alles uuringu käigus (Groves jt, 2009: 54-56). Kaetuse viga on statistiline nihe, mis ilmneb siis, kui valikuraam ja sihtpopulatsioon ei kattu täielikult, selle allikaks võib näiteks olla ebapiisav valimiraam.

Kaetuse viga tuleneb alakaetusest ja ülekaetusest. Alakaetus tekib siis, kui sihtpopulatsiooni indiviidid jäetakse ekslikult valimiraamist välja. Ülekaetus tekib siis, kui sihtpopulatsiooni indiviidid satuvad valimiraami ekslikult või rohkem, kui üks kord. Netokatteviga on ala- ja ülekaetuse erinevus (Lavrakas, 2008: 161). Väiksemate valimite puhul on tõenäolisem alakaetuse esinemine, kuna väikese valimi puhul on tunduvalt raskem kõiki sihtpopulatsiooni osasid adekvaatselt esindada (de Leeuw jt, 2008: 8). Kui mõni rühm on valimiraamis alaesindatud, siis selle rühmaga seotud sotsiaaldemograafilisi või sisulisi omadusi ei saa uuringustatistika arvutamisel arvesse võtta. See põhjustab uuringus ebatäpsust, kuna mõned sihtpopulatsiooni rühmad võivad olla valikuraamis alaesindatud ja seetõttu ka antud valikuraamist võetud valimis alaesindatud (Lavrakas, 2008: 897). Erinevad uurijad kasutavad ülekaetuse terminit erinevalt, mistõttu on oluline selgeks teha, kas antud valimiraami ülekaetuse põhjuseks on populatsiooni suhtes mittekõlblikud indiviidid valimiraamis, korduvad indiviidid valimiraamis või mõlemad (Lavrakas, 2008: 559). Ülekaetust võib lugeda väiksemaks probleemiks võrreldes alakaetusega, kuna seda on kergem tuvastada, vältida ning võimalik korrigeerida ka peale valimi võtmist (Lavrakas, 2008: 788).

1.2.2 Valikumeetodid ja valiku viga

Valim on osahulk indiviide, mille uurijad võtavad uuritavast populatsioonist ja kasutavad valimi põhjal saadud tulemusi üldistamiseks populatsioonile (Neuman, 2011: 246; Tiit ja Tooding, 2019: 281). Valimimaht on objektide arv valimis, mis määratakse uurijate poolt vastavalt antud ülesandele ja eesmärgile (Tiit ja Tooding, 2019: 282).

Valimi moodustamiseks kasutatakse erinevaid meetodeid, üldisemalt jaotatakse neid tõenäosuslikeks ja mittetõenäosuslikeks (Lagerspetz, 2017: 168). Tõenäosusliku valikumeetodi puhul on igal uuritava populatsiooni liikmel nullist erinev ja teadaolev tõenäosus valimisse sattuda. Lihtne juhuvalik on tõenäosusliku valiku erivorm, kus valik tehakse täielikult juhuslikult – kõigil populatsiooni liikmetel on sama suur tõenäosus valimisse sattuda. Tõenäosuslikud meetodid hõlmavad ka keerukamaid meetodeid, nagu kihitatud või mitmeastmeline valik, mille puhul indiviidide kaasamistõenäosused võivad olla erinevad, kuid on siiski teada ja arvesse võetavad analüüsis, näiteks kaalumise abil.

Hästi kavandatud tõenäosuslik valim tagab, et valikunihe on minimaalne (Lavrakas, 2008: 621). Mittetõenäosusliku valikumeetodi puhul ei tehta uuritavast populatsioonist valik juhuslikkuse alusel ja valimisse kaasamise tõenäosused ei ole teada. Pigem kasutatakse subjektiivseid meetodeid, et otsustada, millised indiviidid valimisse kaastakse. Erinevalt tõenäosuslikul meetodil võetud valimist on mittetõenäosuslikul meetodil võetud valimi puhul märkimisväärselt raskem saadud tulemusi laiendada uuritavale populatsioonile (Lavrakas, 2008: 523).

Tõenäosusliku valiku valimitüübid on (Lagerspetz 2017: 168-171):

- Lihtne juhuvalim – iga indiviidi võimalus valimisse kaasatud olla on ühesugune;
- Süstemaatiline juhuslik valim – iga indiviidi võimalus valimisse kaasatud olla sõltub lisaks kindlast intervallist, näiteks valitakse järjekorras iga kümnes indiviid;
- Kihtvalim ehk stratifitseeritud valim – populatsioon on jaotatud osadeks ehk kihtideks oluliste muutujate järgi – kihisisesed valikud tehakse juhuvaliku abil;
- Kobarvalim ehk klastervalim – valimi moodustamise aluseks on uurimisobjektide klastrid, mitte üksikud indiviidid. See tähendab, et kõigepealt valitakse juhuslikult klastrid ja siis küsitletakse või hinnatakse kõiki indiviide nendes klastrites;
- Mitmeastmeline valim – valimi moodustamiseks kasutatakse kombinatsiooni tõenäosuslikest valikuvõtu meetoditest. Näide sellest oleks valida juhuslikult (või kihtvalimiga) suuremad üksused ehk klastrid ja omakorda nendest üksikobjektid valida juhuslikult (või kihtvalimiga).

Kõikse valimi puhul uuritakse kõiki populatsiooni indiviide (Tiit ja Tooding, 2019: 149). Antud valikumeetodit kasutatakse juhul, kui populatsioon on piisavalt väike, et saaks uurida kõiki indiviide selles populatsioonis (Lagerspetz 2017: 168).

Mittetõenäosusliku valiku valimitüübid on (Lagerspetz, 2017: 171-175):

- Lumepallivalim – valimi moodustamiseks kasutatakse teiste vastajate abi, uuritava grupi või rühma esimeste vastajate käest palutakse nimetada teisi uuringuks sobivaid indiviide, sama protsess jätkub, kuni on olemas piisava suurusega valim;
- Mugavusvalim – valimisse võetakse need, keda on võimalik kergesti uuringusse kaasata;
- Kvootvalim – valim on mittetõenäosuslik ja esinduslikkuse parandamiseks kasutatakse teatud sotsiaal-demograafilisi tegureid ehk kvote. Kvote täidetakse kuni tekib sihtpopulatsiooni esindav valim. Valimi esinduslikkus sõltub sellest, kuivõrd järgitakse valikul lisaks kvootidele juhuslikkuse põhimõtet – kvootide kasutamine võimaldab esinduslikkust saavutada nende tunnuste suhtes, mis on kvootide aluseks. Muude tunnuste osas on võimalik esinduslikkust saavutada niivõrd, kuivõrd valik on kvootide siseselt juhuslik;
- Iseselekteeruv valim – valimi moodustavad need indiviidid, kellel endal on huvi uuringus osalemise vastu;
- Teoreetiline ehk strateegiline valim – valim moodustatakse nendest indiviididest ja gruppidest, kellel on uuritava teemaga eriomased kokkupuute viisid.

Mittetõenäosuslike valimite puhul tuleb meeles pidada nende piiranguid. Kaalumise võib esinduslikkust parandada, kuid ei pruugi seda teha ja võib ka esinduslikkust sihtpopulatsiooni suhtes teatud juhtudel halvendada. Loomulikult veapiiri arvutamine ei ole mittetõenäosusliku valimi puhul õigustatud, kuna järeltava statistika eeldused ei kehti, sest tegu ei ole tõenäosusliku valimiga (Wu, 2022). Sama loogika järgi ei ole mittetõenäosuslike valimite puhul automaatselt õigustatud usalduspiiride arvutamine ja hüpoteeside testimine statistiliste testide abil.

Paljud küsitlusuuringuid läbiviivad organisatsioonid kasutavad uuringute teostamiseks paneele. Paneel viitab uuringu valimile, milles samu indiviide küsitletakse või intervjueritakse kahel või enamal korral, lainetena. Paneelid võivad anda teavet suundumuste või muutuste kohta antud populatsioonis teatud aja jooksul. Tavaliselt saab paneeliga mõõta huvipakkuvate omaduste muutusi suurema täpsusega, kui võrreldava suurusega iseseisvate valimitega (Lavrakas, 2008: 564). Veebipaneelid jaotuvad tõenäosuslikel ja mittetõenäosuslikel valimitel põhinevateks. Mittetõenäosuslikud paneelid on märkimisväärselt levinumad, nendest enamiku puhul on tegu

ligipääsupaneelidega, kuhu indiviidid registreeruvad vabatahtlikult. Erinevalt tõenäosuslikest paneelidest ei ole sellise valimi puhul võimalik määratleda iga liikme valimisse sattumise tõenäosust, mistõttu ei saa tulemusi üldistada kogu populatsioonile (Callegaro jt, 2015: 29).

Silmas tuleks pidada paneelide kasutamise ohte, nimelt paneeliliikmed ei pruugi olla sihtpopulatsioonist valitud juhuslikult, tegu võib olla isikutega, kes on paneeliga liitunud omal initsiatiivil või kes on mingisuguste motivaatorite olemasolu tõttu nõustunud paneelis osalema. Kui uuringu valim koostatakse paneeli alusel, peab uuringu korraldaja olema teadlik paneeli koostamise meetodikast ja sellest, kuivõrd on paneel esinduslik sihtpopulatsiooni suhtes. Selline mittetõenäosuslik protsess tõstab veariski, kuna paneelis olevad indiviidid ei pruugi alati sihtpopulatsiooni nii täpselt esindada, kui võiks (de Leeuw jt, 2008: 479).

Valikuviga on juhuslik viga, mille põhjuseks on populatsiooni asemel valimi käsitlemine ehk valimi hajuvusest tingitud viga (Tiit ja Tooding, 2019: 281). Ameerika Avaliku Arvamuse Uurimise Ühing (AAPOR) leiab, et mittetõenäosusliku valimi puhul veapiiri ja teiste järeltava statistika meetodite kasutamine on eksitav (Baker jt, 2013: 97-98). Valikuviga kui juhuslikku viga saab hinnata tõenäosusliku valimi puhul, mittetõenäosusliku valimi puhul pole juhusliku vea hindamise eeldused täidetud, sest pole võimalik arvutada valimiliikmete kaasamistõenäosused.

1.2.3 Mittevastamise viga

Mittevastamise viga esineb siis, kui arvutatud vastajate statistikud ehk parameetrite hinnangud erinevad väärtustes kogu valimi andmetel arvutatud statistikutest (Groves jt, 2009: 59). Teisisõnu seda põhjustab mittevastamine, sealhulgas asjaolu, et vastajate osakogum erineb valimisse sattunud mittevastajate osakogumist (Tiit ja Tooding, 2019: 115).

Vastamismäär on osalenud valimiliikmete osakaal kogu valimist. Kui eeldatav vastamismäär on madal, võib osutada vajalikuks suurendada planeeritud valimi suurust, et tagada piisav arv vastanuid. Siiski ei pruugi suurem valim üksi tagada esinduslikkust, eriti kui vastajad erinevad süsteemselt mittevastajatest. Sellisel juhul võib kaalumise aidata parandada valimi vastavust kogu populatsioonile, kuid selle tõhusus väheneb, kui mittevastamise viga on suur. Seetõttu on oluline ennetada esinduslikkuse probleeme juba andmekogumise etapis, näiteks rakendades sihtrühmapõhist värbamist või muid meetmeid vastamismäära tõstmiseks (Bryman, 2012: 197-200).

Planeeritud valim on valimisse kuuluvate indiviidide arv, mis on ettekavandatud uuringu eesmärkide täitmiseks, saavutatud ehk tegelik valim koosneb nendest indiviididest, kes osutasid

uuringuks kõlblikuks. Planeeritud ja saavutatud valimimahud tihti erinevad üksteisest, mis on tähtis uuringu raportis välja tuua. Saavutatud valimi planeeritud valimist oluliselt väiksemaks osutumine või planeeritud valimimahu täissaamiseks lisavalimi võtmine viitavad madalale vastamismäärale, mis omakorda viitab mittevastamise veale.

Kaetus ja esinduslikkus on kaks erinevat mõistet. Kaetus näitab valikuuringute korral, kui võrd on valikuraamiga hõlmatud sihtpopulatsiooni indiviidid (Lavrakas, 2008: 159). Kaetus omakorda jaotub ülekaetuseks ja alakaetuseks (Tiit ja Tooding, 2019: 110). Esinduslikkus valimis pigem tähendab, et valimil on tugev väline valiidsus seoses sihtpopulatsiooniga. Mida kõrgem on esinduslikkus, seda julgemalt saab uuringu tulemusi sihtpopulatsioonile üldistada (Lavrakas, 2008: 720). Esinduslikkust mõjutavad kaetuse viga, valiku viga ja mittevastamise viga – kaetus on üks tegur, mis mõjutab esinduslikkust, seega kaetus ehk kaetuse vea elimineerimine on küll oluline, kuid esinduslikkust veel ei taga. Esinduslikkus ja vastamismäär on omavahel seotud – madal vastamismäär võib ohustada valimi esinduslikkust, eriti kui vastamisest hoiduvad süstemaatiliselt teatud rühmad. Ka kõrge vastamismäär ei taga automaatselt head esinduslikkust, kui valim on algselt kallutatud.

Potentsiaalsetele uuringus osalejatele motivaatorite nagu raha pakkumine on tõhus viis vastamismäära suurendada ja seeläbi võimalikku mittevastamise viga vähendada. Rahalised motivaatorid (tšekid, kinkekaardid jne) on üldiselt tõhusamad kui mitterahalised motivaatorid, isegi kui mitterahalise auhinna väärtus on suurem kui rahaline preemia (Toepoel, 2016). Teised vastamismäära tõstmise meetmed võivad olla eelkontaktid ja meeldetuletused. Võimalikku mittevastamise viga on tähtis uuringuaruannetes kirjeldada, kuna see mõjutab uuringu valiidsust ja tulemuste usaldusväärsust (Groves jt, 2009: 59).

Kvantitatiivsetes andmekogumites tuleb esile peaaegu alati puuduvaid väärtusi, mis tähendab, et mõne vastaja kohta puudub teave mõne uuritava tunnuse osas. Puuduvad andmed põhjustavad sageli olulisi probleeme statistilises analüüsis, kuna paljud analüüsimeetodid, näiteks regressioonanalüüs või keskmiste võrdlus eeldavad, et iga vaatluse kohta on olemas andmed kõigi huvipakkuvate muutujate kohta (Allison, 2004: 645-649). Andmete puudumise mehhanism võib olla (Little ja Rubin, 2019):

- Täiesti juhuslik ehk MCAR (*Missing Completely at Random*) – puudumine ei sõltu muutujatest;
- Juhuslik ehk MAR (*Missing at Random*) – puudumine sõltub vaadeldavatest muutujatest;

- Mittejuhuslik ehk MNAR (*Missing Not at Random*) – puudumine sõltub puuduvatest väärtustest endast.

Imputeerimine on meetod, millega täidetakse puuduvad andmed ühe või mitme tõenäolise väärtusega, et luua täielik andmestik. See võimaldab kasutada standardset statistilist analüüsi, mis on loodud täielike andmete jaoks, vähendades sellega keerukust võrreldes spetsiaalsete meetodite kasutamisega puudulike andmete puhul. Imputeerimise kasutamisega saab vältida olukorda, kus suur osa andmetest jäetakse analüüsist kõrvale, mis võib viia esinduslikkuse kaotamiseni ja kallutatud järeldusteni. Imputeerimise korrektne kasutamine ja dokumenteerimine on oluline uuringu tulemuste usaldusväärsuse ja meetoodilise läbipaistvuse tagamiseks (Rässler, 2004: 477-480).

Puuduvate andmete käsitlemiseks kasutatakse sageli andmete eemaldamise meetodeid, kus eelkõige eemaldatakse need indiviidid või objektid, millel on puuduv väärtus mõnes analüüsi jaoks olulises tunnuses. Kõige tavalisem meetod on arvuti abil tunnuste loetelu alusel kustutamine ehk individipõhine kustutamine. Individipõhise kustutamise korral toimub eemaldamine, kui puudub väärtus mõne analüüsitava muutuja puhul. Alternatiiv individipõhisele kustutamisele on tunnuspaariti kustutamine, kus juhtum eemaldatakse ainult siis, kui sellel puuduvad väärtused uuritava muutujapaari puhul. Tunnuspaariti kustutamise eelis on, et see võimaldab säilitada rohkem indiviide valimis kui individipõhise kustutamise puhul, kuid puuduseks on, et see võib põhjustada statistilisi ebatäpsusi või moonutusi tulemustes, kuna eri tunnuspaaride alusel tehtud analüüside puhul võivad analüüsitavad individikogumid erineda, muutes tulemuste omavahelise võrreldavuse problemaatiliseks (Lewis-Beck, 2004: 244).

1.2.4 Kohandamise viga

Kohandamise viga tekib siis, kui tehakse kohandusi, nagu kaalumine või imputeerimine, mis võib põhjustada valimi hinnangutes ebatäpsusi. Need kohandused tehakse, et parandada valimi esinduslikkust seoses kaetuse vea, valikuvea ja mittevastamise veaga. Näiteks kaalumise puhul antakse alakaetud rühmadele, nagu kindla linnapiirkonna elanikud suurem kaal, et kompenseerida valikuraami ebapiisavat kaetust. Kuigi need kohandused võivad aidata parandada esinduslikkust, võivad need ka suurendada juhuslikku viga või põhjustada süstemaatilisi vigu, tuues lõppkokkuvõttes ekslikke järeldusi (Lavrakas, 2008: 898; Groves jt, 2009: 59-60).

Netovalimit ehk individide kogumit, kes uuringu puhul uurijate jaoks piisaval määral vastasid tuleb sihtpopulatsiooni suhtes esinduslikkuse saavutamiseks kohandada, võttes arvesse

valimidisaini eripärasid ehk disainiefekti. Kui valimi koostamiseks oleks näiteks vaja ebaproportsionaalset kihtvalimit, tuleks populatsiooni parameetrite hindamisel ebavõrdsete valikutõenäosuste kompenseerimiseks välja töötada sobiv kaalumisprotseduur (Lavrakas, 2008: 898).

Kaalumine on meetod, mille korral valimi iga element korrutatakse teatava kordaja või koefitsiendiga ehk kaaluga. Kaalude valik oleneb uuringu eesmärgist ja valimidisainist. Kaalumise aitab korrigeerida nii ebavõrdseid valikutõenäosusi kui ka kompenseerida mittevastamist (Lavrakas, 2008: 957). Viimasel juhul kasutatakse empiirilistes uuringutes kaalude koostamiseks sageli võtmetunnuseid nagu sugu, vanuserühm ja elupaik (Tiit ja Tooding, 2019: 109). Tihti rakendatakse selleks järelkihistuskaale, mis kohandavad valimi koosseisu vastavusse sihtpopulatsiooniga (Groves jt, 2009: 352). Kaalumise on vastutustundlikkust nõudev protseduur, milles tehtavad valikud võivad oluliselt tulemusi mõjutada ja mis vajab täpset dokumenteerimist (Tiit ja Tooding, 2019: 109).

1.3 Statistilised järeldused

Järeldava statistika tulemuste usaldusväärsus ei sõltu ainult sobivate testide rakendamisest, vaid ka sellest, kuidas tulemusi tõlgendatakse ja esitatakse. Käesolevas töös keskendutakse sellele, kas valimi rakendusuuringutes on järeldused selgelt, adekvaatselt ja eeldusi arvestades esitatud. Eriti oluline on hinnata p-väärtuste ja usaldusvahemike kasutamist, kuna nende kasutamine on levinum.

1.3.1 P-väärtus ja NHST

Statistilised järeldused on valimiandmetel põhinevad järeldused uuritava üldkogumi kohta. Statistilised järeldused on oluliseks osaks teaduslikest uuringutest, kuna need võimaldavad uurijatel hinnata, kas leitud seosed või erinevused on seletatavad valimi hajuvusega või mitte. Selliste järelduste tegemisel on keskne roll hüpoteeside kontrollimisel, mis põhineb nullhüpoteesi (H_0) ja alternatiivhüpoteesi (H_1) vastandlikel väidetel. Nullhüpotees eeldab tavaliselt, et uuritavatel muutujatel pole ühtegi seost või erinevust, samas kui alternatiivhüpotees väidab vastupidist (Tiit ja Tooding, 2019: 192, 242).

Üks levinumaid meetodeid statistiliste järelduste tegemiseks on statistilisel olulisusel põhinev nullhüpoteesi testimine (NHST), mida kasutatakse laialdaselt sotsiaalteadustes. NHST on saanud mitmekülgset kriitikat, kuna see ei anna uurijatele sageli neid vastuseid, mida nad tegelikult otsivad. Peamised probleemid on seotud selle võimetusega pakkuda otseseid tõenäosuslikke hinnanguid hüpoteeside tõesuse kohta, selle sõltuvusega valimi suurusest ning kalduvusega ülehinnata tõendusmaterjali nullhüpoteesi vastu (Hoekstra jt, 2014). See tähendab, et tulemused, mis on statistiliselt olulised suures valimis, ei pruugi olla tegelikult sisulise tähtsusega. Samuti võivad väikesed valimid viia olukorrani, kus olulised seosed või erinevused jäävad avastamata, sest testidel puudub piisav statistiline võimsus.

Hoekstra jt (2014) uuringus esitati 70 akadeemilisele psühholoogile üks statistiliselt oluline p-väärtus ning paluti neil hinnata, kas kuus erinevat tõlgendust sellest on õiged või valed. Kõik kuus väidet olid tegelikult valed, kuid keskmiselt nõustusid osalejad 2,5 valega, näidates, et isegi eksperdid tõlgendavad p-väärtusi sageli väärtalt. Sarnased tulemused saadi kordusuuringutes, kus selgus, et ka statistikat õpetavad professorid teevad samu vigu. See viitab sügavamale probleemile: statistiliste testide formaalne tundmine ei pruugi tähendada nende korrektset rakendamist. Vale tõlgendamine ei ole ainult individuaalne probleem, vaid see mõjutab ka teaduskirjanduse kvaliteeti, kuna ekslikud tõlgendused võivad kanduda edasi uutesse uuringutesse ja antud uuringute põhjal tehtud otsustesse.

Olulisuse tõenäosus ehk p-väärtus näitab, kui suur on tõenäosus saada vaatlustulemus, mis on vähemalt sama äärmuslik kui tegelikult saadud tulemus, eeldusel et nullhüpotees on tõene (Tiit ja Tooding, 2019: 193). P-väärtuseid tuleb tõlgendada ettevaatlikult: madal p-väärtus ei tõesta automaatselt alternatiivhüpoteesi, vaid näitab, et vähemalt sama ekstreemse tulemuse saamine on nullhüpoteesi kehtimisel ebatõenäoline. Olulisuse tõenäosus on täpne, kui on täidetud kõik eeldused, nii konkreetse meetodi kui ka järeltava analüüsi eeldused üldiselt – kahjuks on see ideaaljuhtum, mida praktikas tihti ette ei tule, sest reeglina mingi viga andmetes esineb. Järeldav analüüs arvestab ainult valikuveaga, mitte muude vealiikidega. Tulemuse usaldusväärsus sõltub oluliselt kasutatava meetodi eelduste täitmisest ning andmete kvaliteedist, mistõttu p-väärtust tuleks alati analüüsida koos teiste statistiliste näitajate ja kontekstiga.

Olulisuse nivoo (α) on eelnevalt määratud piirväärtus, mille alusel otsustatakse, kas p-väärtus on piisavalt väike nullhüpoteesi kummutamiseks. Tavapäraselt kasutatakse α väärtust 0.05, mis tähendab, et kui p-väärtus on alla selle, peetakse tulemust statistiliselt oluliseks ja nullhüpotees kummutatakse (Tiit ja Tooding, 2019: 193). Siin ei ole absoluutset reeglit – sõltuvalt uurimisvaldkonnast ja analüüsi kontekstist võib olulisuse nivoo olla ka rangem 0.01, 0.001 või leebem 0.10, sõltuvalt sellest, kui suur vea risk on uurija jaoks aktsepteeritav. Olulisuse nivoo ja p-väärtus on omavahel seotud, kuid erinevad mõisted: olulisuse nivoo on eelnevalt määratud kriteerium, p-väärtus on konkreetse analüüsi tulemus, mida võrreldakse selle kriteeriumiga.

Kuna olulisuse nivoo määramine mõjutab seda, kui sageli tehakse valesid järeldusi, on oluline mõista ka I ja II tüüpi vigu. I tüüpi viga tekib siis, kui nullhüpotees lükatakse tagasi, kuigi see tegelikult kehtib, ehk tegemist on valepositiivse tulemusega. II tüüpi viga esineb siis, kui nullhüpoteesi ei lükata tagasi, kuigi see on tegelikult väär, ehk tegemist on valenegatiivse tulemusega (Rotello jt, 2025: 83).

1.3.2 P-väärtusega seotud levinud väärtõlgendused

P-väärtuste põhjal tehtavad otsused nähtuste olemasolu kohta on sageli problemaatilised, kuna selline lähenemine ignoreerib valepositiivse tulemuse riski ehk olukorda, kus statistiliselt oluline tulemus tuleneb ainult valimi hajuvusest ega peegelda tegelikku seost. Valepositiivse tulemuse riski hindamine võimaldab realistlikumalt hinnata, kas otsus võib olla vale. Valepositiivse riski suurus sõltub mitmest tegurist: olulisuse nivoost (α), testi võimsusest, saadud p-väärtusest ning eelnevast tõenäosusest, et uuritav seos on päriselt olemas. Näiteks kui enne uuringut hinnatakse seose olemasolu tõenäosuseks 50%, siis $p < 0.05$ korral võib valepositiivse riski suurus olla kuni

26% (olenevalt testi võimsusest), mis on oluliselt kõrgem, kui tavaliselt eeldatakse (Colquhoun, 2017). See tähendab, et kõige laiemalt levinum olulisuse nivoo $p < 0.05$ ei pruugi olla piisavalt usaldusväärne, et eristada tegelikke seoseid juhuslikest leidudest ning seda, et p-väärtus üksi ei ole piisav tõend seose olemasolu kohta.

Pernet (2017) toob välja mõned levinumad eksiarvamused p-väärtuste mõistmisel, mis võivad viia valede või eksitavate järeldusteni statistilises analüüsis:

1. „P- väärtus näitab seose tugevust“ – P-väärtus ei anna infot uuritava seose tugevuse ega usaldusväärsuse kohta, kuna see näitab vaid tõenäosust saada vaadeldud või äärmisem tulemus eeldusel, et nullhüpotees kehtib. Seose tugevust ei saa p-väärtuse põhjal määrata, sest see sõltub nii seose tugevusest populatsioonis kui ka valimi suurusel. Greenland jt (2016) rõhutavad sarnaselt, et üks levinum väärtõlgendus p-väärtuste puhul on, et statistiliselt olulisel tulemusel, näiteks $p < 0.05$ on automaatselt ka sisuline tähtsus või et see peegeldab kindlat erinevust.
2. „1 - p näitab tulemuse kordumise tõenäosust“ – Sageli arvatakse, et väike p-väärtus tähendab suurt tõenäosust saada järgmises uuringus samasugune tulemus, kuid see ei ole korrektne. Kuna p-väärtus ei anna infot seose kohta, ei ole see ka replikatsiooni tõenäosuse mõõdupuu. Lisaks sõltub p-väärtus otseselt valimi suurusel ja ainult piisavalt suure valimi ning võimsuse korral saab tulemusi usaldusväärselt tõlgendada. Greenland jt (2016) toovad samuti välja, et p-väärtust on väärtõlgendatud kui taasloodavat mõõdikut, kuigi isegi identsete tingimuste korral võivad arvutatud p-väärtused erineda.
3. „Madala p-väärtuse olemasolu tähendab, et alternatiivhüpotees on tõene“ – Väike p-väärtus näitab ainult seda, et nullhüpotees ei sobitu hästi vaadeldud andmetega, kuid see ei toeta ühtegi konkreetset alternatiivhüpoteesi. Nullhüpoteesi hülgamine võib tuleneda ka teistest teguritest. Greenland jt (2016) märgivad sarnase olemusega väärtõlgendust, madala p-väärtuse, näiteks $p \leq 0.05$, automaatselt tõlgendamise kui tõendina selle kohta, et nullhüpotees on väär või see tuleks tagasi lükata – see on vale, kuna väike p-väärtus ei tõesta, et nullhüpotees on vale, vaid viitab sellele, et saadud andmed on ootamatud eeldusel, et kõik statistilise testi eeldused on täidetud, madal p-väärtus võib tuleneda ka juhuslikust veast, mitte tegelikust seosest. $P \leq 0.05$ tähendab ainult seda, et kui ainus tulemusi mõjutav tegur oleks juhuslikkus, siis nii suur või suurem erinevus ilmneks vähem kui 5% juhtudest.

4. „P-väärtus on nullhüpoteesi (H_0) tõesuse tõenäosus“ – Levinud eksiarvamus on, et p-väärtus näitab nullhüpoteesi kehtivuse tõenäosust. Tegelikult väljendab p-väärtus tõenäosust saada vähemalt sama ekstreemne vaatlus tingimusel, et nullhüpotees on tõene. Greenland jt (2016) selgitavad, et p-väärtusega seonduvad levinud väärtõlgendused hõlmavad ka arvamust, et $p > 0.05$ tähendaks vajadust nullhüpoteesi vastu võtta. Tegelikult ei näita suur p-väärtus nullhüpoteesi õigsust, see võib lihtsalt tähendada, et andmetest ei piisa olulise erinevuse tuvastamiseks, näiteks valimi suuruse tõttu. Statistiliselt mitteoluline tulemus $p > 0.05$ ei ole samaväärne nullhüpoteesi tõestusega, vaid näitab vaid seda, et vaadeldud andmed ei ole piisavalt veenvalt vastuolus nullhüpoteesiga.

Mõned teised p-väärtuste väärtõlgendused, mida Greenland jt (2016) välja toovad ja mida peaks silmas pidama:

1. „ $P = 0.05$ ja $p \leq 0.05$ tähendavad sama asja“ – Need ei ole samatähenduslikud. Väide $p = 0.05$ viitab konkreetsele väärtusele, samas $p \leq 0.05$ hõlmab ka palju väiksemaid p-väärtusi, näiteks 0.0001, mis võivad viidata palju tugevamale vastuolule nullhüpoteesiga. Selline erineva tähendusega p-väärtuste samasse kategooriasse liitmine muudab tulemuse tõlgenduse ebaselgeks ja eksitavaks.
2. „P-väärtusi on õige esitada võrratustega, näiteks $p < 0.02$ või $p > 0.05$ “ – Ei, sest see raskendab tulemuste täpset tõlgendamist. Täpne p-väärtus annab selgema ülevaate, võrratustega esitamine võib kallutada tulemuste tähendust. Erandina võib väga väikeste p-väärtuste, näiteks $p < 0.001$ puhul kasutada võrratust, kuna sel juhul pole praktilist vahet.
3. „Statistiline olulisus on uuritava nähtuse omadus“ – Ei ole, sest statistiline olulisus ei ole nähtuse sisemine omadus, vaid statistilise testi tulemuste põhjal tehtav järeldus, mis põhineb subjektiivselt valitud olulisuse nivool, tavaliselt 0.05. Levinud eksiarvamus on, et test „tuvastab“ statistilise olulisuse, kuid tegelikult on see vaid p-väärtuse kategooriline tõlgendus valitud olulisuse alusel. Uuritav seos kas eksisteerib või mitte, sõltumata sellest, kas p-väärtus langeb alla 0.05.

Sageli jaotatakse tulemused meelevaldselt kahte kategooriasse: „statistiliselt olulised“, näiteks $p < 0.05$ ja „statistiliselt mitteolulised“, näiteks $p \geq 0.05$. Selline binaarne lähenemine loob eksliku mulje, et statistiline testimine võimaldab teha selgepiirilise otsuse selle kohta, kas seos või erinevus eksisteerib või mitte. Tegelikuses ei tähenda $p < 0.05$, et seos või erinevus on olemas, ega $p \geq 0.05$, et seost või erinevust pole. Selline lähenemine ignoreerib tõenäosuslikku loomust, millele statistilised testid tegelikult tuginevad, ning ei arvesta valimi suuruse mõju tulemuste tõlgendamisele (Colquhoun, 2017). Statistilise olulisuse kontseptsioon on tihti valesti tõlgendatud.

Näiteks võib statistiliselt oluline tulemus olla praktikas tähtsusetu, kui seos või erinevus on väike. Samuti ei tähenda statistiliselt mitteoluline tulemus automaatselt, et seost või erinevust ei eksisteeri – see võib lihtsalt olla tingitud piisava andmekogumi puudumisest. Bernardi jt (2017) rõhutavad, et oluline on eristada statistilist olulisust sisulisest olulisusest, mis viitab tulemuste praktilisele tähtsusele reaalses kontekstis.

Kokkuvõttes on p-väärtus statistiline mõõdik, mis näitab, kui tõenäoline on saada vähemalt sama äärmuslik vaatlustulemus eeldusel, et nullhüpotees (H_0) on tõene. See ei mõõda seose või erinevuse tugevust ega otsuse usaldusväärsust, ei anna infot nullhüpoteesi tõenäosuse kohta ning ei näita, kas nullhüpotees või alternatiivhüpotees on tõene. P-väärtus ei ole taasloodav mõõdik, isegi identsete arvutamise tingimuste korral. P-väärtus ei ole nullhüpoteesi tõesuse tõenäosus ning suur või väike p-väärtus ei kinnita ega lükka automaatselt ümber järeldusi. Madal p-väärtus viitab ainult sellele, et vaadeldud andmed on ootamatud, kui nullhüpotees ja kõik eeldused kehtivad, kuid see ei tõesta, et nullhüpotees on väär või alternatiivhüpotees tõene. Samuti võib madal p-väärtus tuleneda nii juhuslikest kui ka süstemaatilistest vigadest või valikulisest raporteerimisest, mitte tegelikust seosest.

1.3.3 Usaldusvahemikud

Usaldusvahemik on tundmatu populatsiooniparameetri, näiteks keskmise või standardhälbe vahemikhinnang, mis eeldatavasti katab parameetri tegeliku väärtuse. Usaldusvahemik on valimi põhjal määratav piirkond, mis katab parameetri tegeliku väärtuse etteantud läve ületava tõenäosusega – usalduspiirkonna määramisel kasutatavat läve nimetatakse usaldusnivooks, selle väärtus on tavapäraselt 0.95 või 0.99. Usalduspiirid on usaldusvahemikku määravad arvud, kahepoolse usaldusvahemiku korral alumine ja ülemine usalduspiir või ühepoolse usaldusvahemiku korral vastavalt alumine või ülemine usalduspiir (Tiit ja Toeding, 2019: 276). Hu jt (2010: 211-215) kirjeldavad usaldusvahemikku järgnevalt: see on tundmatu populatsiooniparameetri vahemikhinnang, mis konstrueeritakse juhuvalimi põhjal ja on alati seotud kindla usaldusnivooga. See usaldusnivoo näitab protseduuri usaldusväärsust – tõenäosust, et korduval valimi võtmisel konstrueeritud usaldusvahemikud sisaldaksid parameetri tegelikku väärtust. Usaldusnivoo esitatakse nii murdarvuna (0.90; 0.95; 0.99) kui protsentidena (90%; 95%; 99%), kusjuures mõlemad vormid on levinud. Näiteks 95% usaldustase tähendab, et kui korrata valimivõttu ja usaldusvahemike konstrueerimist lõpmatu arv kordi, sisaldaks umbes 95% nendest

vahemikest populatsiooniparameetri tegelikku väärtust. Kõrgem usaldusnivoo toob tavaliselt kaasa laiemad usaldusvahemikud.

Usaldusvahemik on arvuline vahemik, mis on konstrueeritud parameetri hinnangu ümber. Selline vahemik ei näita otseselt parameetri omadust, vaid viitab usalduspiiride arvutamise meetodi omadustele. Nagu p-väärtuste puhul, ei võimalda usaldusvahemikud teha tõenäosuslikke väiteid parameetrite või hüpoteeside kohta (Hoekstra jt, 2014). Usaldusvahemikke on soovitatud p-väärtuste alternatiivina, kuna need võimaldavad hinnata statistilist olulisust ja annavad samal ajal seose tugevuse hinnangu. Eeldades, et usaldusvahemiku sümmeetria ja laius on õiged, annavad need ka mõningase ettekujutuse tõenäosusest, et sarnane väärtus võib esineda tulevastes uuringutes. Sama valimi suurusega kordusuuringute puhul tagavad 95% usaldusvahemikud umbes 83% tõenäosuse replikatsiooni õnnestumiseks. Kui valimi suurused uuringute vahel erinevad, ei taga usaldusvahemikud a priori katvust (Pernet, 2017).

Pernet (2017) rõhutab, et kõige levinum viga on tõlgendada usaldusvahemikku kui tõenäosust, et mingi parameeter nagu populatsiooni keskmine jääb sellesse vahemikku X%. Greenland jt (2016) nõustuvad sellega, tuues välja üheks levinumaks usaldusvahemike väärtõlgenduseks „konkreetne 95% usaldusvahemik sisaldab 95% tõenäosusega tegelikku väärtust“. Õige tõlgendus on, et korduvate mõõtmiste korral, kasutades sama valimi suurust ja võttes andmeid samast populatsioonist, sisaldab X% juhtudest saadud usaldusvahemik tegelikku parameetri väärtust (Pernet, 2017). Usaldusvahemikud on statistiliste järelduste tegemisel olulised, kuna need ei anna ainult binaarset jah/ei vastust nullhüpoteesi tagasilükkamise kohta, vaid pakuvad ka lisainfot seose tugevuse ja selle võimaliku varieeruvuse kohta. Erinevalt p-väärtusest, mis näitab vaid seda, kas tulemus on statistiliselt oluline, aitab usaldusvahemik hinnata, kui täpne ja millise võimaliku ulatusega see tulemus on. Kitsas usaldusvahemik viitab sellele, et hinnang on täpsem, samas kui lai usaldusvahemik näitab suuremat määramatust (Greenland jt, 2016).

Greenland jt (2016) toovad veel välja järgnevad levinud usaldusvahemike väärtõlgendused:

1. „Parameetri väärtus väljaspool 95% usaldusvahemikku on välistatud või ümber lükatud“ – See ei vasta tõele. Nagu p-väärtuste puhul, põhineb ka usaldusvahemiku arvutamine mitmetel eeldustel. Kui need eeldused ei pea paika, võivad saadud tulemused olla eksitavad. Seega ei ole seose tugevuse või erinevuse väljajäämine usaldusvahemikust iseenesest piisav tõend selle ümberlükkamiseks, see sõltub nii eeldustest kui ka kasutatud 95% kriteeriumist.

2. „Kui kaks usaldusvahemikku kattuvad, siis see tähendab, et erinevus ei ole statistiliselt oluline“ – See on vale. Kahe 95% usaldusvahemiku osaline kattumine ei välista statistiliselt olulist erinevust. Usaldusvahemikud on loodud näitamaks, kus kumbki tegelik keskmine võiks asuda, sõltumata teise keskmise väärtusest. Seetõttu ei piisa ainuüksi usaldusvahemike kattumise hindamisest, et järeldada erinevuse olulisust, oluline on ka keskmiste vahe ja selle varieeruvus.
3. „95% usaldusvahemik ennustab, et 95% tulevastest uuringute hinnangutest langeb sellesse vahemikku“ – See on vale mitmel moel. 95% viitab sellele, kui sageli paljude uuringute käigus saadud uued usaldusvahemikud sisaldaksid tegelikku seose tugevust või erinevust, mitte sellele, kui sageli üks konkreetne vahemik sisaldaks tulevasi hinnanguid. Isegi ideaalsetes tingimustes on tõenäosus, et järgmise uuringu punktihindang langeb praegusesse usaldusvahemikku tavaliselt palju väiksem kui 95%. Näiteks kui kahes sõltumatus uuringus saadakse normaaljaotusega punktihindangud, millel on sama standardviga, siis esimese uuringu 95% usaldusvahemiku tõenäosus sisaldada teise uuringu punktihindangut on vaid 83%.
4. „Kui üks 95% usaldusvahemik sisaldab nullväärtust ja teine mitte, siis see, mis nullväärtust ei sisalda on täpsem“ – Ei vasta tõele. Parameetri hinnangu täpsust mõõdetakse usaldusvahemiku laiusena, mitte sellega, kas see sisaldab nullväärtust. Kitsam usaldusvahemik tähendab väiksemat määramatust, olenemata sellest, kas nullväärtus jääb vahemikku või mitte.

Usaldusvahemike õige tõlgendamine on uurimustes oluline, aga paljud uuringute teostajad ei oska neid seda teha, mis ohustab statistiliste järelduste usaldusväärsust (Hoekstra jt, 2014). See on eriti problemaatiline, kuna usaldusvahemikke peetakse tihti paremaks alternatiiviks p-väärtustele, aga nende vääralt tõlgendamine võib viia samasuguste ekslike järeldusteni. Kui uurijad usaldavad ekslikult, et usaldusvahemik näitab tõenäosust parameetri tegeliku väärtuse kohta, võib see kallutada uuringutulemuste tõlgendamist ning luua ebausaldusväärseid järeldusi. Hoekstra jt (2014) sõnul võivad probleemi juured peituda selles, et NHST-l põhinevat usaldusvahemikku tõlgendatakse intuiitiivselt nagu Bayeslikku statistikat, kuigi need on filosoofiliselt ja matemaatiliselt erinevad käsitlused.

2. PROBLEEMISEADE

Minu bakalaureusetöö „Aastatel 2015-2022 Haridus-ja Teadusministeeriumi tellimisel läbiviidud küsitlusuuringute aruannete metoodikakirjelduste analüüs“ peamiseks järelduseks osutus see, et poliitikat kujundavatel ja seega väga tähtsatel töödel olid tõsised puudujäägid raportite uuringumetoodika nendes aspektides, millest sõltub, kas järeldavale analüüsile seatud valimiga seotud eeldused on täidetud. Tollel ajal käsitlesin neid aspekte kõigest metoodikakirjelduste kontekstis ja põhimõtteliselt kontrollisin, kas antud aspekte oli üldsegi raportites esile toodud. Mõned nende aspektide olemasolu uurimise negatiivsemad tulemused olid: adekvaatne valikuraami kirjeldus oli vaid 2/3 valimisse kuulunud uuringuraportites (30 raportit); ainult pooltes oli kirjeldatud, kuivõrd hästi valim sihtpopulatsiooni esindab; vaid 1/3 uuringutes oli selgelt sõnastatud, kas valim oli võetud kasutades tõenäosuspõhiseid või mittetõenäosuslikke meetodeid; vähem kui 1/3 uuringutes oli välja toodud planeeritud ja saavutatud valimite mahud ning nende eripärasid võrreldud (Põldoja, 2023).

Tuvastatud puudujäägid valimiga seotud metoodikakirjeldustes viitavad vajadusele süveneda põhjalikumalt sellesse, kuivõrd rakenduslikes uuringutes täidetakse järeldava statistika kasutamiseks vajalikke eeldusi. Kuna järeldav analüüs on usaldusväärsete teaduslike ja rakenduslike järelduste aluseks, siis ei piisa sobivate testide rakendamisest, vältimatult on oluline ka see, et andmete kogumine ja valimi moodustamine oleksid läbipaistvalt kirjeldatud ning vastaksid analüüsi usaldusväärstust tagavatele tingimustele. Kui eeldused nagu valimi esinduslikkus ei ole täidetud võib analüüsi tulemus olla eksitav või isegi väär, sõltumata kasutatud statistilisest meetodist. Samuti on oluline hinnata, kuivõrd selgelt ja sisukalt on uuringuraportites sõnastatud statistilised järeldused, ebaselged või kontekstitud järeldused võivad viia väärade tõlgendusteni, mis omakorda võivad mõjutada poliitika kujundamist. Seetõttu on mitte üksnes vajalik määratleda, kas eeldused on formaalselt olemas, vaid hinnata ka nende sisu kvaliteeti ning seda, kas järeldav analüüs on esitatud viisil, mis võimaldab lugejal aru saada, mil määral saadud tulemused on üldistatavad ja usaldusväärsed.

Järeldava statistika mõistmine ja korrektne rakendamine on uurimistöö kvaliteedi seisukohalt keskse tähtsusega, aga mitmed uuringud on näidanud, et isegi eksperdid võivad selle aspektidega eksida. Olemasoleva praktilise näitena sellest, kuivõrd levinud on selliste järeldava analüüsi aspektide väärarusaamad on Hoekstra jt (2014) artiklis kirjeldatud autorite tehtud uuring, kus esitati psühholoogia valdkonda kuuluvatele 120 uurijale ja 442 üliõpilasele kuus konkreetset väidet, mis hõlmasid usaldusvahemike erinevaid tõlgendusi. Kuigi kõik kuus väidet olid valed,

kinnitasid nii uurijad kui ka üliõpilased keskmiselt rohkem kui kolm väidet, mis viitab usaldusvahemike tõsisele valesti mõistmisele.

Ministeeriumite poolt tellitud rakenduslike uuringute kvaliteet on oluline, kuna nende tulemused ja nendel põhinevad järeldused mõjutavad poliitiliste otsuste tegemist ja ühiskondlike suundade kujundamist. Ebaadekvaatselt kirjeldatud analüüsi eeldused või statistilised järeldused võivad viia vigaste järeldusteni, mis omakorda mõjutavad ebaõigesti poliitikaid, ressursijaotust ja arenguplaane. Kuna statistilise analüüsi usaldusväärsus sõltub eelduste täitmisest, on kriitilise tähtsusega hinnata, kas järeldavad analüüsid on korrektselt teostatud ja selgelt väljendatud.

Käesoleva uurimistöö eesmärk on hinnata valitud Haridus- ja Teadusministeeriumi tellitud rakendusuuringute, milles on kasutatud järeldavat analüüsi, meetodilist kvaliteeti, keskendudes sellele, kuivõrd on nendes täidetud järeldava statistika kasutamiseks vajalikud eeldused ning kas statistilised järeldused on esitatud selgelt ja sisukalt, võimaldades teha usaldusväärseid ja põhjendatud üldistusi.

Uurimiseesmärgi saavutamiseks olen püstitanud kaks uurimisküsimust:

1. Kuivõrd on Haridus- ja Teadusministeeriumi tellimusel läbi viidud rakenduslikes uuringutes kasutatavate andmete puhul täidetud järeldava analüüsi eeldused?
2. Kuivõrd on Haridus- ja Teadusministeeriumi tellimusel läbi viidud rakenduslikes uuringutes tehtud statistilised järeldused selgelt ja adekvaatselt sõnastatud?

3. METOODIKA

3.1 Valim

Käesoleva uurimistöö valimi moodustasid Haridus- ja Teadusministeeriumi tellimusel ajavahemikus 2013–2024 koostatud uuringud, mis olid valimipõhised ja milles rakendati järeldavat analüüsi. Tööst jäid välja kõik uuringud, mida viis läbi HTM ise või selle allasutused, kuna nende puhul ei ole raportid alati täielikud. Samuti jäeti välja uuringud, mis üksnes kirjeldasid varasemate uuringute tulemusi – välja arvatud juhul, kui varasemate uuringute andmeid kasutati uue järeldava analüüsi läbiviimiseks.

Oluline on rõhutada, et selle töö analüütiliseks allikaks olid üksnes uuringute HTM-i kodulehel avaldatud raportid. Kui mõnes uuringus viidati täiendavatele materjalidele või lisainformatsioonile, mis ei olnud kodulehelt kättesaadav, neid andmeid käesoleva töö raames ei kasutatud. See tähendab, et järeldused ja hinnangud tuginevad üksnes nendele andmetele ja metoodika kirjeldusele, mis olid avalikus raportis esitatud.

Tuleb arvestada, et raportites kirjeldamata asjaolud ei pruugi viidata sellele, et vastavaid aspekte uurimistöös ei käsitletud – küll aga tähendab see, et metoodilise läbipaistvuse ja retsenseerimise seisukohalt jäävad need aspektid puudulikuks. Seetõttu keskendub käesolev töö eelkõige sellele, kuivõrd uuringuraportid, mitte uuringute terviklik sisu, võimaldavad hinnata järeldava analüüsi eelduste täitmist ja statistiliste järelduste adekvaatsust. Seepärast on põhjendatud keskenduda üksnes avalikule raportile. See on lõpptulemus, mille põhjal otsuseid tehakse. Kui olulised metoodilised aspektid jäävad raportis kajastamata, ei ole uuringu tulemusi võimalik usaldusväärselt hinnata ega kriitiliselt analüüsida.

Rakendusuuringu puhul kehtib läbipaistvuse ja reprodutseeritavuse põhimõte, mille järgi peab raport sisaldama kogu asjakohase teabe metoodika ja analüüsi kohta, ilma vajaduseta otsida lisamaterjale või pöörduda autorite poole (Lang ja Altman, 2019). Probleem ei seisne seega niivõrd selles, mida uuringus tegelikult tehti, vaid selles, et raporti põhjal ei ole seda võimalik tuvastada ega kontrollida, mistõttu väheneb uuringu praktiline väärtus ja usaldusväärsus poliitiliste otsuste tegemisel.

Samuti on raportite põhjalikkus oluline tavakodanikule. Kuna uuringud on rahastatud maksumaksja rahaga ja mõjutavad ühiskondlikult olulisi otsuseid, peaks igaühel olema võimalus hinnata, kui põhjendatud ja usaldusväärsed need järeldused on. Kui oluline teave jääb avalikust raportist välja, piirab see kodanike võimalust hinnata riigi poliitika põhjendatust, õonestades

usaldust mitte üksnes uuringute, vaid ka ministeeriumi otsustusprotsesside vastu. Läbipaistvus on seetõttu vajalik mitte ainult teaduslikel ega halduslikel, vaid ka demokraatlikel kaalutlustel.

Valimi moodustamiseks kontrolliti kõigi HTM-i (2025) kodulehel kättesaadavate uuringute vastavust seatud kriteeriumidele. HTM-i andmebaasis oli analüüsi hetkel üle 200 uuringu või analüüsi, millest pärast kriteeriumite rakendamist jäi valimisse 30 raportit. Valimi koosseisust on toodud detailsem ülevaade lisas 1.

Valimi uuringuraportite mahud ja teemad varieerusid märkimisväärselt, moodustades üsna heterogeense kogumi. Raporti pikkus võis ulatuda mõnekümnest leheküljest kuni mitmesaja leheküljeni. Selline varieeruvus tähendab, et ka rakendatud metoodikate detailsus ja analüütiline sügavus olid väga erinevad. Mõnes raportis oli metoodiline osa äärmiselt põhjalik, teistes aga napp ja üldsõnaline.

Valimisse kuulunud uuringutes käsitleti haridus- ja noortevaldkonna probleemistikku väga erinevatest vaatenurkadest, keskendudes nii õppekorralduse ja hariduspoliitika tõhususele kui ka õppijate ja töötajate kogemustele. Uuringutes analüüsiti hariduskorralduse arendamist ja muutunud õpikäsituse rakendamist, distantsõppe mõjusid eri haridustasemetel ning haridusteemaliste strateegiate ja läbivate teemade, sealhulgas jätkusuutliku arengu loomimist koolides. Tööjõu ja haridusvaldkonna seoste käsitlemisel uuriti õpetajate ja akadeemiliste töötajate töötingimusi, töötasustamist ja karjääri võimalusi ning vaadeldi ka õppepuhkuste ja õppetootuste süsteemide toimimist. Noorte ja õppurite vaates keskenduti õpingute katkestamise põhjustele, haridustee jätkamise ja katkestamise teguritele, noorte rahulolule noorsootööga ning kutse- ja kõrghariduse lõpetajate tööalasele toimetulekule. Tähelepanu pöörati ka teemadele nagu keelehoiakud ning muukeelsete ja uusimmigrantidest õpilaste toimetulekule Eesti haridussüsteemis. Lisaks käsitleti ettevõtlus- ja üldpädevuste arendamist erinevatel haridustasemetel, koduse õpikoormuse mõju kooliõpilastele ning haridusvalikute ja tööturu ootuste vahelisi seoseid.

Käesolev uurimistöö keskendub just HTM-i tellitud uuringuraportitele eelkõige seetõttu, et nende uuringute tulemusi kasutatakse riiklikul tasandil haridus-, teadus- ja noortepoliitika kujundamisel, mistõttu on äärmiselt oluline nende metoodiline usaldusväärsus ja järelduste adekvaatsus. Lisaks on minul autorina olemas eelnev analüütiline kogemus HTM-i uuringute läbivaatamisel, mis võimaldab käsitlust järjepidevalt süvendada. Kuigi antud töö valim piirdub ühe konkreetse tellija uuringutega, on alust eeldada, et mitmed teatavad mustrid, näiteks potentsiaalsed puudujäägid järeldava analüüsi eelduste käsitluses või statistiliste järelduste esituses võivad esineda ka teiste

tellijate rakendusuuringutes. Kuna ka teiste riigiasutuste rakendusuuringud tellitakse sageli samade turuosaliste näiteks suuremate uuringufirmade või teadusasutuste poolt ja tehakse sarnases institutsionaalses raamistikus, võiks töö tulemusi pidada mingil määral indikatiivseteks ka laiemas avaliku sektori rakendusuuringute kontekstis.

3.2 Meetod

Käesolevas töös rakendasin samu meetodeid, mida kasutasin oma bakalaureusetöös. Seetõttu põhineb see alapeatükk minu bakalaureusetöö meetodite kirjeldusest (Põldoja, 2023). Selle uurimistöö analüüsiosa läbiviimiseks kasutasin kontentanalüüsi meetodit. Kontentanalüüs on uurimismeetod, mille abil süstemaatiliselt kodeeritakse ja kategoriseeritakse uuritav tekstiline sisu, et tuvastada korduvaid mustreid, tähendusi või struktuure ning võimaldada nende kvantitatiivset või kvalitatiivset analüüsi (Neuman, 2011: 49). Antud töö kontekstis on sobivaim kirjeldada kontentanalüüsi kui meetodit, mille abil saab teksti omadusi mõõta arvuliselt, mis võimaldab andmeid omavahel kvantitatiivselt analüüsida (Lagerspetz, 2017: 190).

Kontentanalüüsi peamiseks töövahenditeks on kodeerimisjuhend ja kodeerimisleht (Lagerspetz, 2017: 191). Kodeerimisjuhend on juhiste kogum, mis sisaldab kõiki võimalikke kategooriaid iga kodeeritava dimensiooni kohta. See hõlmab kõigi muutujate loendit, iga muutuja alla kuuluvate kategooriate kirjeldust, igale kategooriale vastavaid numbreid ehk koodi ning juhiseid selle kohta, kuidas erinevad aspektid omavahel seotud on ja mida tuleks arvestada konkreetse koodi määramisel (Bryman, 2012: 299). Tavaline kodeerimisleht on tabel, kuhu kantakse iga muutuja väärtused – igale analüüsiüksusele vastab tabelis üks rida ning igale muutujale on oma veerg (Lagerspetz, 2017: 192).

Kodeerimisjuhendi puhul on oluline tagada kodeerijate vaheline kooskõla ehk intersubjektiivsus sõltumata sellest, kas tööd teeb ainult üks inimene või mitu (Bryman, 2012: 299). Intersubjektiivsus tähendab, et analüüsi tulemused peaksid olema põhimõtteliselt samad hoolimata sellest, kes kodeerimist teostab. Seetõttu peavad muutujad ja väärtused olema defineeritud viisil, mis ei jäta ruumi mitmeti tõlgendamiseks. Intersubjektiivsuse tagamine aitab vältida kahte peamist ohtu: esiteks süstemaatilist viga, mis võib tuleneda kodeerijate erinevast arusaamast muutujate sisust ja teiseks kodeerimise juhuslikkust (Lagerspetz, 2017: 193).

Kontentanalüüsi peamised eelised on selle läbipaistvus, sõltumatus otsesest kontaktist uuritavatega ning metoodiline paindlikkus. Meetod võimaldab uurida olemasolevaid tekstimaterjale struktureeritult ja ühtsete kriteeriumide alusel, muutes analüüsi korratavaks ja

võrreldavaks. Meetodi piirangutena tuleb arvestada, et analüüsi kvaliteet sõltub allikmaterjalide sisust ja kvaliteedist, kodeerimisjuhend võib paratamatult jätta ruumi subjektiivsele tõlgendusele ning meetod ei sobi hästi uurimisküsimustele, mis otsivad vastust „miks“-tüüpi küsimustele (Bryman, 2012: 304-307).

Analüüsiprotsessis käsitleti kõiki uuringuraporteid terviklikult, sealhulgas nende põhiosa, lisa, lisad, tabeleid ja jooniseid. Lähtuti põhimõttest, et oluline meetodiline või analüütiline teave võib paikneda igasuguses raporti tekstiosas, aga ka visuaalsetes esitustes ja lisamaterjalides, mistõttu ei piirdunud analüüs ainult raporti põhitekstiga.

Kodeerimisprotsess jagunes kaheks etapiks. Esimeses etapis koostati meetodilisele kirjandusele ja varasemale bakalaureusetööle tuginedes kodeerimisjuhend, et katta süstemaatiliselt kõik olulised valimi kujunemise ja statistiliste järelduste adekvaatsusega seotud aspektid. Valimiga seotud küsimused formuleeriti konkreetsemalt ja standardiseeritult, sest need tuginevad hästi defineeritud ja kergemini kontrollitavatele meetodilistele kriteeriumitele (nt valikumeetod, vastamismäär, kaalumise), mille esitamine on rakendusuuringutes tavapärane ja sageli ka meetodika peatükkides üheselt leitud. Statistilisi järeldusi puudutavate küsimuste koostamine oli keerulisem ja mõneti otsivam protsess, kuna sellised teemad on rakendusuuringutes esitatud väga erineva põhjalikkuse ja selgusega. Samuti on statistiliste mõistete korrektne tõlgendamine tihti seotud nüansside ja konteksti sobivuse hindamisega, mida ei saa alati lihtsalt standardreeglite abil fikseerida.

Teises etapis viidi läbi raportite detailne analüüsimine vastavalt koostatud juhendile. Esmalt tutvuti raportite sisukorraga, et tuvastada potentsiaalselt asjakohased peatükid, seejärel sirvisin raportid algusest lõpuni läbi. Kui esimene ülevaatus oli tehtud, algas vastuste otsimine konkreetsetele kodeerimisjuhendi küsimustele. Iga küsimuse puhul kujunes töö käigus välja unikaalne vastuse leidmise strateegia, mis hõlmas muu hulgas raportite elektroonilises vormis märksõnade sihipäraselt otsimist otsingufunktsioonide abil sobivate tekstiosade kiireks leidmiseks.

Eriti mahukate ja keerukate statistiliste järelduste hindamisel ei piirdunud ainult aspektide olemasolu fikseerimisega, vaid pöörati tähelepanu ka sellele, kuidas neid esitati ja tõlgendati. Sageli tähendas see raporti erinevate osade vahel edasi-tagasi liikumist, et kontrollida, kas järeldused olid kooskõlas kasutatud meetodikaga ja kas tõlgendused olid sisuliselt põhjendatud. Kontentanalüüs ei taandunud seega ainult andmete kirjanemisele, vaid hõlmas ka raportite argumentatsiooni ja meetodilise läbipaistvuse kriitilist hindamist. Selline põhjalikum lähenemine võimaldas anda ka keerulisemate ja mahukamate küsimuste puhul sisuliselt põhjendatud ja usaldusväärseid hinnanguid uurimuste meetodilise ja analüütilise kvaliteedi kohta.

3.3 Kodeerimisjuhendi koostamine

Käesoleva uurimistöö teostamiseks koostasin kodeerimisjuhendi, mille eesmärk on võimaldada süsteemset ja reprodutseeritavat hinnangut uuringuraportites esitatud järeltava analüüsi eelduste täitmisele ja statistilistele järeldestele. Kodeerimisjuhend põhineb neljal peamil alikal, millest igaüks panustas erinevatesse aspektidesse juhendi ülesehituses: Lang ja Altman (2015) poolt loodud „*Statistical Analyses and Methods in the Published Literature*“ (SAMPL) eeskiri, Ameerika Psühholoogide Assotsiatsiooni (APA) „JARS-Quant“ (2024) juhend, minu enda 2023 aasta bakalaureusetööst „Aastatel 2015-2022 Haridus- ja Teadusministeeriumi tellimisel läbiviidud küsitlusuuringute aruannete metoodikakirjelduste analüüs“ pärinev „AAPOR eeskirjal põhinev kodeerimisjuhend“ ning Bernardi jt (2017) uuringu küsimustik.

SAMPL juhised, mille on koostanud Thomas A. Lang ja Douglas G. Altman, on loodud selleks, et parandada biomeditsiiniliste ajakirjade artiklites statistilise aruandluse kvaliteeti. Need pakuvad praktilist raamistikku, mis aitab autoritel, toimetajatel ja retsensentidel tagada, et teadustöodes esitatav statistiline teave oleks täpne, selge ja läbipaistev. Juhiste peamine eesmärk on vältida statistiliste meetodite ja tulemuste puudulikkust või eksitavat esitust, mis võib viia väärade järeldesteni ning seeläbi kahjustada teadusuuringute usaldusväärsust. SAMPL sisaldab soovitusi statistiliste meetodite kirjeldamiseks, valimi omaduste esitamiseks, tulemuste selgeks ja piisavalt üksikasjalikuks raporteerimiseks ning usaldusvahemike ja p-väärtuste korrektseks kasutamiseks. Eesmärk on tagada, et lugejal oleks võimalik hinnata analüüsi adekvaatsust ja tulemuste tõlgendamise põhjendatust. Lang ja Altman (2015) rõhutavad statistiliste meetodite ja tulemuste raporteerimisel üht juhtpõhimõtet: statistilisi meetodeid tuleb kirjeldada piisava detailsusega, et teadlik lugeja saaks tulemusi ise kontrollida – eeldusel, et tal on juurdepääs algandmetele. See tähendab, et kõik leiud tuleb võimalusel kvantifitseerida ja esitada koos vastavate statistiliste näitajatega.

JARS–Quant (2024), mille on välja töötanud APA, on põhjalik juhend kvantitatiivsete uuringute raporteerimiseks. See määratleb standardid kogu uurimisprotsessi kajastamiseks, alates valimi kirjeldamisest ja andmete kogumisest kuni analüüsistrateegiate ja tulemuste tõlgendamiseni. Erilist tähelepanu pööratakse statistiliste meetodite täpsele esitamisele, seoste tugevuse ja usaldusvahemike selgele näitamisele ning puuduvate andmete käsitlemisele. JARS–Quant juhised on rakendatavad kõigis valdkondades, kus kasutatakse kvantitatiivseid uurimismeetodeid.

Minu enda loodud AAPOR eeskirjal põhinevast kodeerimisjuhendist (Põldoja, 2023) võtsin kasutusele need osad, mille abil hinnata, kui selgelt ja täpselt on dokumenteeritud valimi

moodustamise ning andmetöötluse aspektid. Kuna järeltava analüüsi eelduseks on esinduslik valim ja metoodiliselt läbipaistev lähenemine, on oluline kontrollida, kas uuringuaruannetes on piisavalt detailselt kirjeldatud valimi suurus, planeeritud ja saavutatud valimite mahud, nende eripärade võrdlus ning võimalike erinevuste põhjused. Samuti hindan, kas valimi võtmise meetod on selgelt määratletud, kas on kasutatud tõenäosuslikke või mittetõenäosuslikke valikumeetodeid ning kuivõrd on käsitletud valimi juhuslikkuse ja esinduslikkuse kirjeldamist ning põhjendamist. Kuigi mitmed järeltuste eeldusi käsitlevad küsimused kodeerimisjuhendis lähtuvad varasemast bakalaureusetööst, kujunes nende lõplik valik välja käesoleva uurimistöö eesmärkide, uurimisküsimuste ning eelnevalt viidatud SAMPL- ja JARS-Quant-standardite põhjal.

Bernardi jt (2017) uuring lisas kodeerimisjuhendile olulise mõõtme, pakkudes struktuuri, mille abil hinnata, kuivõrd selgelt ja sisukalt on uuringuaruannetes sõnastatud statistilised järeldused. Uuringus kasutati küsimustikku, mis keskendus sellele, kas teadusartiklites esitatud järeldused on loogiliselt kooskõlas kasutatud analüüsimeetoditega ning kas need arvestavad piisavalt valimi omadusi ja tulemuste üldistatavuse piiranguid. Kuna käesoleva uurimistöö eesmärk on hinnata ka järelduste selgust ja adekvaatsust, sobib Bernardi jt (2017) küsimustik hästi täiendama eelmainitud juhiseid ning aitab operatsionaliseerida just neid komponente, mis puudutavad järelduste sisulist kvaliteeti. Selle alusel loodud hindamiskriteeriumid integreeriti käesoleva töö kodeerimisjuhendisse. Gelmani & Sterni (2006) artikkel toetas antud küsimuste formuleerimist.

Koostatud kodeerimisjuhend koosneb kolmest osast. Küsimused 1–5 käsitlevad iga analüüsiüksusena vaadeldava uuringu raporti põhiantmeid, nagu pealkiri, avaldamise aasta, läbiviija(d) ja uuringu faili viide. Küsimused 6–19 keskenduvad järeldava analüüsi eelduseks olevatele metoodilistele aspektidele, sh valimi kujunemine, valikumeetodite ja sihtpopulatsiooni kirjeldus, vastamismäärad, kaalumise ning andmetöötlusprotsessid. Küsimused 20–30 on suunatud konkreetsete statistiliste tulemuste ja järelduste esituse ning tõlgendamise hindamisele – sealhulgas kontrollitakse p-väärtuste, usaldusvahemike ja koefitsientide esituse täpsust ning järelduste loogilist vastavust analüüsi tulemustele. Terve kodeerimisjuhend koos hindamiskriteeriumite kirjeldustega on leitav lisast 2.

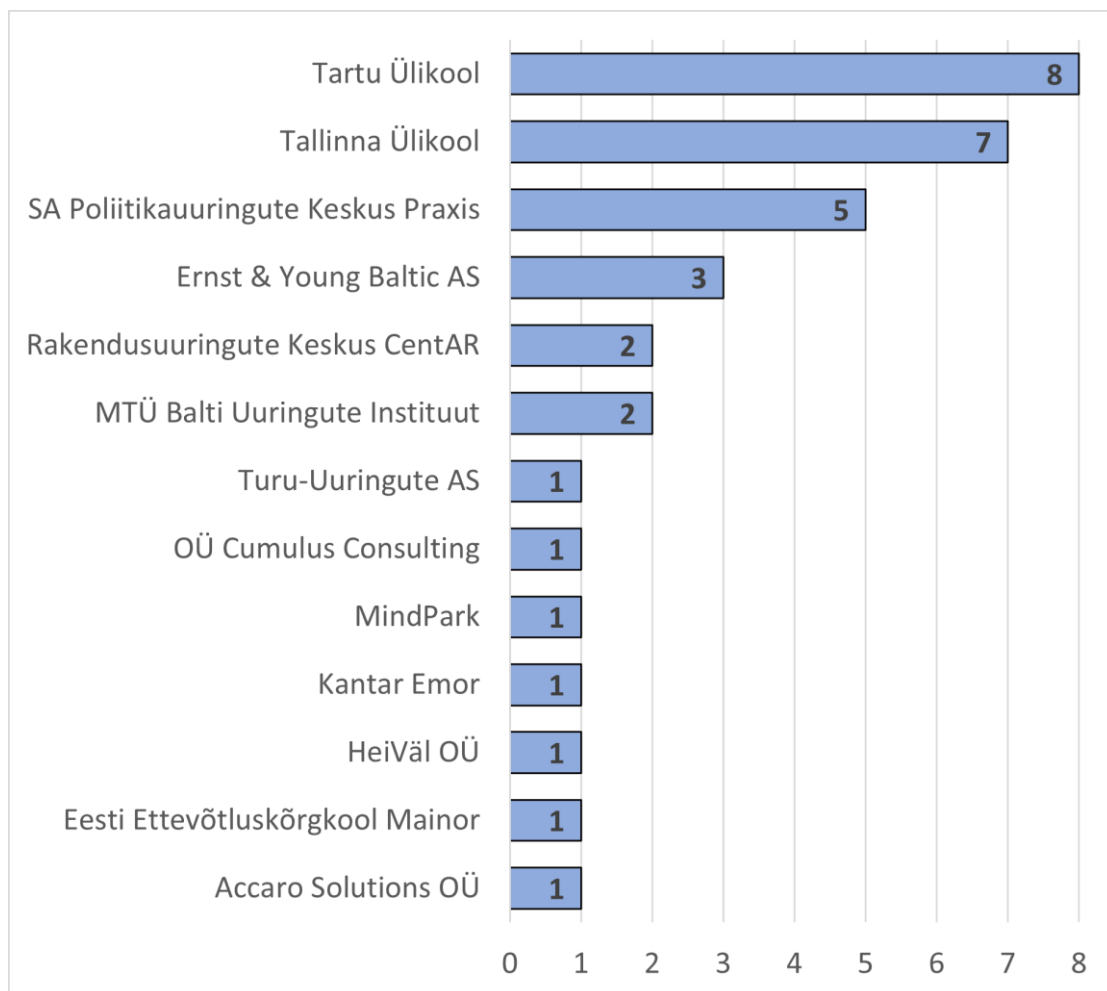
Kodeerimisjuhend on üles ehitatud viisil, mis võimaldab anda süsteemse hinnangu nii järeldava analüüsi eelduste kajastamisele kui ka statistiliste järelduste esituse selgusele ja põhjendatusele. Küsimused on formuleeritud struktureeritud kujul, kus igale hinnatavale aspektile vastab võimalikult üheselt tõlgendatav vastusevariant koos hindamiskriteeriumiga. Samuti on võimalik eristada juhtumeid, kus konkreetne aspekt ei ole käsitletava uuringu puhul asjakohane. Selline

ülesehitus võimaldab hinnata mitte ainult tehnilisi näitajaid, vaid ka seda, kuivõrd esitatud andmed ja tulemused loovad aluse usaldusväärseteks järeldusteks.

4. TULEMUSED

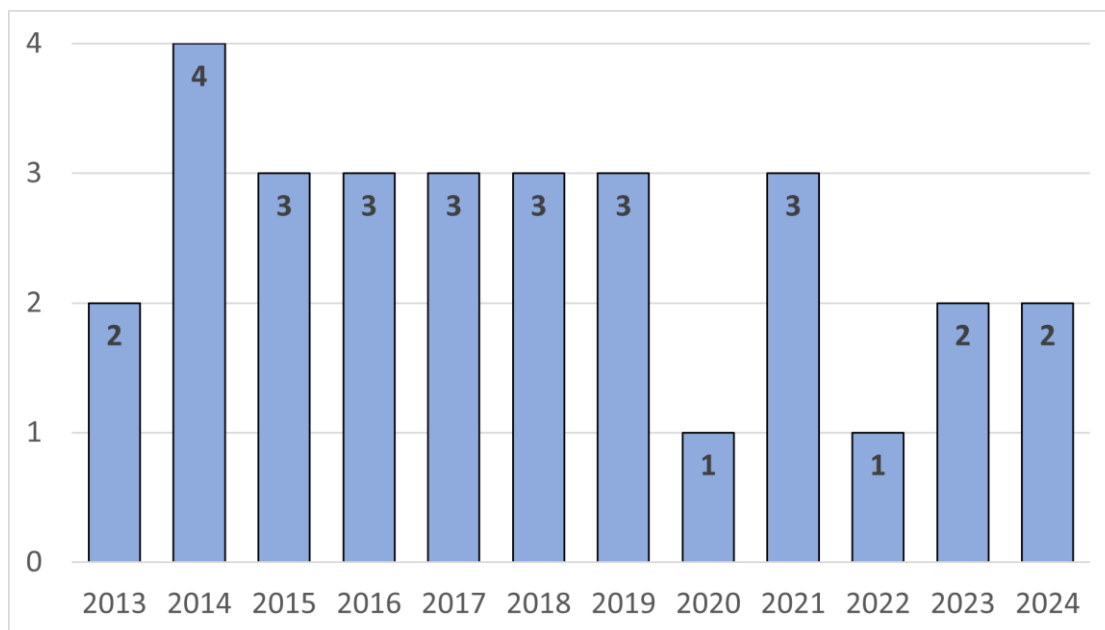
4.1 Analüüsiüksused

Valimi moodustasid 30 Haridus- ja Teadusministeeriumi tellimusel aastatel 2013–2024 koostatud uuringute raportit, mis olid valimipõhised ja sisaldasid järeltavat analüüsi. Joonisel 1 on kujutatud nende uuringuaruannete koostamisel osalenud asutuste jaotust. Kokku oli töodel 34 läbiviijat, kuna mõned uuringud viidi läbi mitme asutuse koostööna. Kokku oli kaasatud 13 erinevat asutust, kellest kõige sagedamini esinesid Tartu Ülikool (8 uuringut), Tallinna Ülikool (7 uuringut) ja SA Poliitikauuringute Keskus Praxis (5 uuringut). Tartu Ülikooli koosseisu on siin kaasatud ka Sotsiaalteaduslike rakendusuuringu keskuse ehk RAKE ja Tartu Ülikooli Haridusteaduste instituut. Terve valim on leitav lisas 1.



Joonis 1. Aruandeid läbiviinud asutuste sagedusjaotused

Joonisel 2 on kujutatud valimisse kuuluvate uuringuaruannete jaotus nende koostamise aastate lõikes. On näha, et aastatel enne 2020 ilmus HTM-i tellimusel rohkem sobivaid uuringuaruandeid, samas kui alates 2020. aastast on nende arv märgatavalt väiksem.

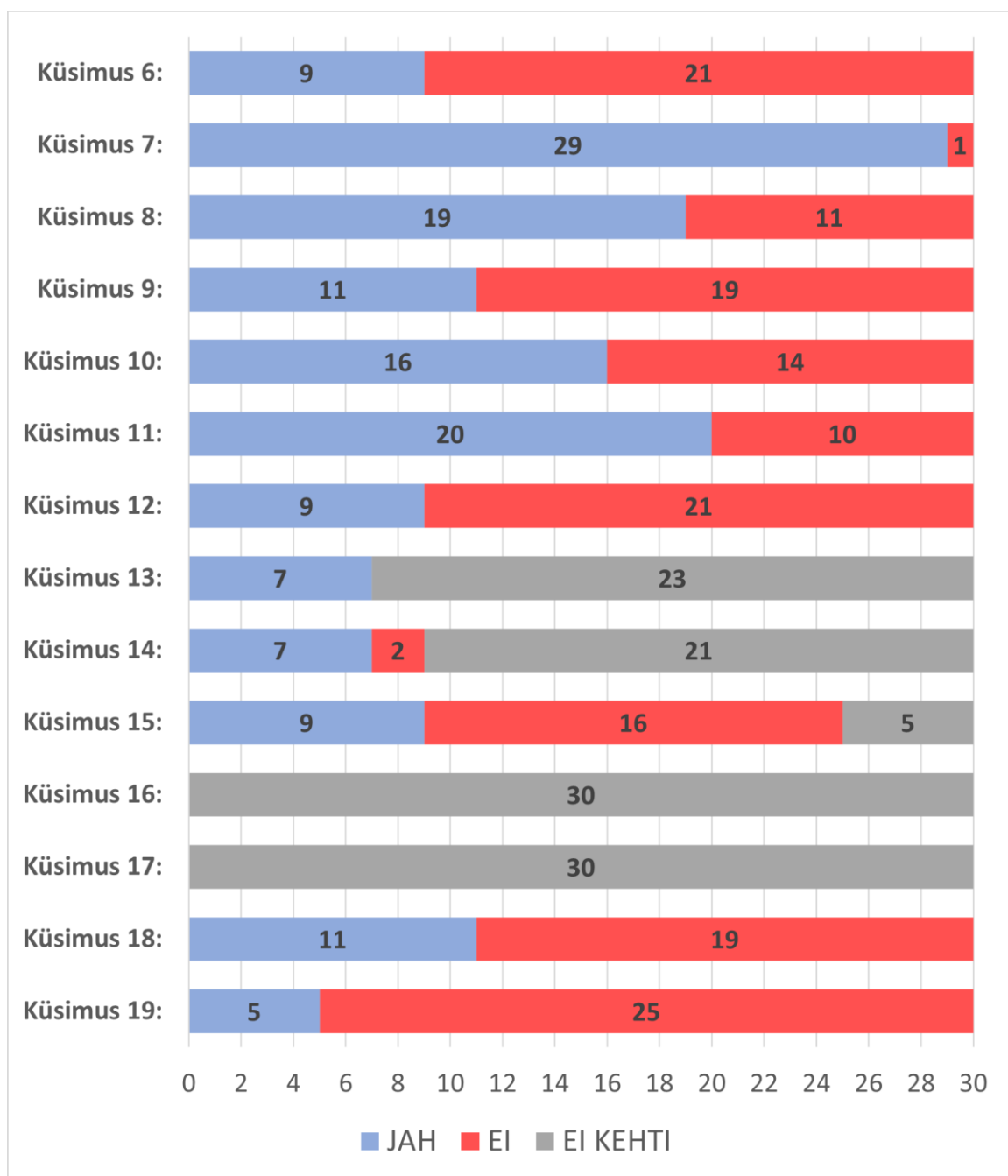


Joonis 2. Aruannete koostamise aastate sagedusjaotused

4.2 Järeldava analüüsi eeldused

Joonisel 3 esitatud sagedusjaotused näitavad kodeeritud vastuseid kodeerimisjuhendi küsimustele 6–19, mis käsitlesid järeldava analüüsi eelduste täitmisega seotud metodoloogilisi aspekte. Terve kodeerimisjuhend koos hindamiskriteeriumite kirjeldustega on leitav lisast 2. Need küsimused keskendusid eeskätt valimi kujundamise ja andmete kogumise protsesside läbipaistvusele ja adekvaatsusele, mis on aluseks tehtavate järelduste kehtivusele ja tõlgendatavusele. Tõenäosuslike ja mittetõenäosuslike valikumeetodite eristamine (küsimus 6), valimi suuruse ja valimiraami kirjeldamine (küsimused 7 ja 8) ning planeeritud ja saavutatud valimi võrdlemine (küsimus 9) annavad ülevaate valimi struktuurist ja uuritava populatsiooni kaetusest. Järgnevalt käsitleti valiku- ja värbamismeetodeid, sealhulgas valikukriteeriume ja võimalikku kallutatust (küsimus 10) ning sihtpopulatsiooni määratlust ja valimi esinduslikkust (küsimused 11 ja 12). Kaalude kasutamise loogika ja andmetöötlusmeetmete nagu imputeerimine või andmete eemaldamise (küsimused 13 ja 14) ning osalusmäära suurendamise strateegiate (küsimus 15) vaatlemine võimaldasid hinnata andmete kvaliteeti ja usaldusväärsust. Käsitleti ka kvootide või paneelide kasutamist (küsimused 16 ja 17), vastamis- ja osalemismäärade esitust (küsimus 18) ning uurimuse metodoloogiliste piirangute sõnastamist (küsimus 19).

Need küsimused võimaldasid hinnata, kuivõrd läbipaistvalt ja metoodiliselt põhjendatult on kujundatud uuringu valim ning läbi viidud andmete kogumine, mis omakorda mõjutab järeltava analüüsi usaldusväärsust. Kuna need aspektid on aluseks üldistuste tegemisel, on nende dokumenteerimise ja põhjendamise kvaliteet oluline kogu uurimistulemuste tõlgendatavuse seisukohalt. Järgnevalt tuuakse detailsemalt välja kodeerimistulemused iga küsimuse lõikes. Tulemused on esitatud üldistatud kujul, ilma konkreetsete raportite või asutuste viiteid välja toomata, kuna töö eesmärk ei ole hinnata üksiku uuringu või asutuse töö kvaliteeti, vaid analüüsida rakendusuuringute metodoloogilist praktikat laiemalt.



Joonis 3. Järeltava analüüsi eelduste täitmist puudutavate küsimuste sagedusjaotused

4.2.1 Valimimeetodi sõnastus

Küsimus 6 kodeerimisjuhendis oli: „*Kas on selgelt sõnastatud, kas valim on võetud kasutades tõenäosuspõhiseid või mittetõenäosuslikke meetodeid?*“ Selle küsimuse eesmärk oli hinnata, kas uurimisaruanDES on üheselt mõistetavalt esitatud, millist valikumeetodit kasutati, ja kas tegemist oli tõenäosusliku või mittetõenäosusliku valikuga.

Hinnangu „Jah“ said need 9 raportit, kus kasutatud meetod oli kas otsesõnu nimetatud, näiteks „koostasime lihtsa juhuvalimi“ või „tegemist on mittetõenäosusliku kvootvalimiga“ või kus esitatud kirjeldus või kontekst võimaldas üheselt kinnitada kasutatud valimimeetodi liiki. Nendes töödes ei jäänud tõsist kahtlust, millist tüüpi valimimeetodiga oli tegemist.

Ülejäänud 21 raportit hinnati „Ei“, kuid mitte alati seetõttu, et meetodi kirjeldus oleks täielikult puudunud. Probleemiks oli enamasti see, et info oli ebaselge, puudulik või liiga üldine. Esines väiteid nagu „võeti valim X inimesest“ või „valim esindab Eesti õpetajaid“, mis ei võimalda ilma täiendava selgituseta järeldada, kas tegemist oli tõenäosusliku valikuga. Hindamisel rakendati siin teadlikult konservatiivset lähenemist ja ainult kaudsetest vihjetest ei piisanud „Jah“ hinnangu andmiseks.

Oluline on märkida, et kõik „Ei“ hinnangud ei olnud sisult võrdsed – mõnes töös puudus meetodi kirjeldus täielikult, teistes võis valik olla isegi aimatav, kuid mitte üheselt sõnastatud. Kui hinnangukriteeriumit oleks rakendatud leebemalt ja lubatud teha rohkem oletusi põhjalikuma kontekstitõlgenduse alusel, oleks hinnangute jaotus ilmselt olnud ligikaudu 20 „Jah“ ja 10 „Ei“. Siiski otsustasin kasutada rangemat hinnangukriteeriumi, lähtudes põhimõttest, et valimimeetodi tuvastamine peab põhinema selgel ja ühemõttelisel teabel, mitte tõlgenduslikel oletustel.

4.2.2 Valimi suurus ja valimiraam

Küsimus 7 kodeerimisjuhendis oli: „*Kas on välja toodud valimimahud, sh valimiraamiti?*“ Selle küsimuse eesmärk oli kontrollida, kas uurimisaruanDES on selgelt esitatud netovalimi suurus, see tähendab, kui palju inimesi uuringus tegelikult osales. Tulemused olid väga head: 29 raportit hinnati „Jah“, vaid 1 raport hinnati „Ei“ vääriliseks.

See üks erandjuhtum oli jätku-uuring, kus autorid viitasid varasemale raportile, milles olevat valimi üksikasjad. Kuigi see võiks mõnes mõttes olla loogiline praktika, ei vastanud see käesoleva töö hindamis põhimõtetele, mille järgi iga raport pidi iseseisvalt sisaldama kogu vajaliku

metoodilise teabe. Samasuguses olukorras leidis valimis ka teine raport, kus vaatamata viitele varasemale uuringule, oli valim selgelt uuesti lahti seletatud.

Nii kõrge „Jah“ hinnangute osakaal tuleneb osaliselt ka sellest, et valimi suuruse esitamine on enamike rakenduslike uuringute puhul elementaarne ja isegi kõige pealiskaudsemad aruanded kipuvad vähemalt selle info esitama. Valimimahu avaldamine on üldjuhul ka tellija ja lugeja ootuse osa, mis teeb sellest ühe vähestest metoodilistest elementidest, mis on uuringuraportites standardina esindatud.

Küsimus 8 kodeerimisjuhendis oli: *„Kas on kirjeldatud valimiraam või sellele vastav andmekogum?“* Siin hinnati, kas aruandes on välja toodud, millise loendi või registri alusel valim moodustati, näiteks kas kasutati rahvastikuregistrit, koolide nimekirja, töötajate andmebaasi vms ning kas aruandest selgub, kui hästi see valikuraam hõlmas kogu sihtpopulatsiooni. Samuti hinnati, kas kirjeldati kaetuse ulatust, näiteks keda valikuraam ei hõlmanud.

Selle kriteeriumi puhul olid hinnangud jaotunud oluliselt ühtlasemalt: 19 raportit hinnati „Jah“ ja 11 raportit hinnati „Ei“, need negatiivsed hinnangud anti eelkõige siis, kui valimiraamist ei olnud juttu üldse või kui kirjeldus oli liiga napilt esitatud, näiteks mainiti küll, et valim võeti „X sihtgrupist“, kuid ei täpsustatud, millise loendi alusel need inimesed valiti või mis oli raamiks, mille alusel üldse kaasamist tehti. Samuti esines olukordi, kus viidati mingile üldisele andmestikule või registrile, näiteks „kasutati Äriregistri andmeid“, kuid töö kontekstis ei olnud selge, mida see täpselt tähendas või kui esinduslik see raam sihtpopulatsiooni suhtes oli.

„Jah“ hinnangute puhul oli varieeruvust selles, kui põhjalikult oli valikuraam kirjeldatud. Mõnes üksikus töös olid valimiraami detailid väga läbipaistvalt esitatud, näiteks lisade sektsioonis välja toodud täielik koolide loetelu, millest valim moodustati. Need olid ideaalsed näited, kuid nõuda sellist detailsust kõikidelt töölt oleks ebamõistlik, eriti juhtudel, kus raam on väga suur või sisaldab isikustatud andmeid. Seetõttu aktsepteeriti „Jah“ all ka tööd, kus raam oli küll üldisemalt kirjeldatud, kuid siiski piisaval määral, et lugeja saaks aru, kuidas valim moodustati ja milliseid piire see raam seadis.

4.2.3 Planeeritud ja saavutatud valim

Küsimus 9 kodeerimisjuhendis oli: *„Kas on välja toodud planeeritud ja saavutatud valimite mahud, nende eripärade võrdlus ja erinevuste põhjused ning kuidas nendega seotud probleemidega tegeleti?“* Planeeritud valimi mahuna peetakse siin silmas vastajate arvu, mille

kokkusaamist uuringu kavandamise faasis eesmärgiks seati. 11 raportit hinnati selle kriteeriumi puhul „Jah“, 19 raportit hinnati „Ei“.

See küsimus kuulus nende hulka, mille puhul hindamiskriteeriumit rakendati teadlikult leebemalt, eelkõige „eripärade võrdluse“ komponendi osas. See tähendab, et „Jah“ hinnati juba juhul, kui raportis oli selgelt välja toodud nii planeeritud kui ka saavutatud valimi suurus. Kui see info oli olemas, siis eripärade võrdlus loeti vaikimisi täidetuks.

Hinnanguks „Jah“ saanud raportites oli selgelt eristatud vähemalt kaks valimimahtu ning enamasti lisatud ka lühike selgitus, miks erinevus tekkis, näiteks uuringu toimumise ajast või mingisugustest tehnilistest piirangutest. Mõnes raportis oli lisatud ka info selle kohta, kas ja kuidas neid kõrvalekaldeid üritati kompenseerida, näiteks kasutades täiendavat värbamist, kaalumist või muul viisil proovides vähendada erinevust planeeritud- ja saavutatud valimi vahel.

Kõik „Ei“ hinnangud tulenesid sellest, et raportis puudus teave planeeritud valimi suuruse kohta. Sellisel juhul ei ole võimalik hinnata, kas andmete kogumise käigus tekkis märkimisväärsed kõrvalekaldeid või milline oli mittevastamise ulatus, mis omakorda piirab arusaama uuringu esinduslikkusest ja võimalike vigade allikatest.

4.2.4 Osalejatega kontakteerumine ja valikukriteeriumid

Küsimus 10 kodeerimisjuhendis oli: *„Kas on selgelt kirjeldatud meetodid, millega osalejatega kontakteeruti, millega nad välja valiti või värvati, sh kirjeldatud kõik valikukriteeriumid ja võimalik üle-esindatus valimis?“* Kriteeriumi alusel hinnati „Jah“ 16 raportit ning „Ei“ 14 raportit.

Hinnanguks anti „Jah“ juhtudel, kus kontakteerumine oli arusaadavalt kirjeldatud, näiteks oli esitatud, kes osalejatega kontakti võttis, mis kanalit kasutati (näiteks koolide kaudu, avalik üleskutse, otsekontakt jms), ning kirjeldatud valikukriteeriumid. Hinnang „Jah“ anti ka mõnel juhul, kui üle-esindatust ei käsitletud, kuid ülejäänud info oli selgelt esitatud ja aruande põhjal oli võimalik aru saada, kuidas valim kujunes.

„Ei“ hinnati raportid, kus vastavad kirjeldused puudusid täielikult või piirdusid üksikute üldiste väidetega. Tüüpilised „Ei“ hinnanguga vastused hõlmasid olukordi, kus öeldi vaid, et „võeti ühendust e-posti teel“ või „saadeti kutse osalemiseks“, ilma selgituseta, kes valiti, kuidas nad valiti või miks just need isikud osalesid. Mõnes aruandes oli mainitud osalejate arvu ja üldist profiili, kuid kontakteerumise protsessi ei olnud kirjeldatud.

4.2.5 Sihtpopulatsiooni määratlus ja valimi esinduslikkus

Küsimus 10 kodeerimisjuhendis oli: „*Kas on täielikult ja täpselt määratletud populatsioon, sh asukoha, vanuse, muude sotsiaalsete ja demograafiliste omaduste, aja lõikes (kui see on uuringu spetsiifikat arvestades relevantne)?*“ Selle küsimuse puhul keskenduti eelkõige selgusele, kas on ühemõtteliselt aru saada, millisele rühmale tulemusi üldistada taheti. 20 uuringut hinnati „Jah“ ja 10 uuringut „Ei“.

„Jah“ hinnangu puhul piisas tavaliselt selgest ja konkreetsest sõnastusest sihtpopulatsiooni kohta, näiteks: „uuringu sihtrühmaks olid kõik üldhariduskoolide 7.–9. klassi õpilased Eestis 2021. aastal“. Selline sõnastus sisaldab nii geograafilist piirkonda kui ka demograafilisi tunnuseid ning võimaldab üheselt mõista, kellele tulemusi taheti üldistada.

„Ei“ hinnangud jagunesid kaheks. Mõned raportid ei sisaldanud üldse sihtpopulatsiooni määratlust, ehk uuringu konteksti sügavalt teadmata ei oleks lugejal olnud võimalik aru saada, kellele uuring suunatud oli. Levinum oli teine juhtumitüüp, kus sihtpopulatsioon oli määratletud liiga üldiselt või umbmääraselt, nt „õpilased“ või „õpetajad“, ilma et oleks selgitatud, millist gruppi õpilasi või õpetajaid tegelikult uuring vaadelda tahtis. Siin jäi mulje, et autorid ise mõistsid, et nende uuring ei hõlmanud kogu nimetatud rühma, kuid metoodiline sõnastus jäi umbmääraseks ja detailidest puudulikuks.

Küsimus 10 kodeerimisjuhendis oli: „*Kas on välja toodud, kuivõrd hästi valim sihtpopulatsiooni esindab ja selle järeld?*“ 9 raportit hinnati „Jah“ ja 21 raportit „Ei“.

Valdava osa „Ei“ hinnangute puhul esinduslikkust ei käsitletud üldse, ei olnud mainitud, kas valim üldse esindab sihtpopulatsiooni ega püütud hinnata selle mõju tulemuste üldistatavusele. Mõnes üksikus töös oli mainitud esinduslikkust, kuid seda tehti pealiskaudselt ja ilma selgitusteta, mis tugineksid kvantitatiivsetele või loogilistele argumentidele, kasutati terminoloogiat nagu „valim on esinduslik“ ilma tõenduseta.

„Jah“ hinnangud tuginesid selgetele ja sisukatele esitluse viisidele. Näiteks oli kasutatud tabelit, kus võrreldi valimi ja kogu sihtpopulatsiooni jaotust mõne võtmetunnuse lõikes (nt sugu, piirkond, vanus), või oli tehtud kvalitatiivne analüüs, mis tõi välja, milliste rühmade osas esines võimalik üle- või alaesindatus.

4.2.6 Kaalumine ja andmete imputeerimine

Küsimus 13 kodeerimisjuhendis oli: „*Kas on kirjeldatud, kuidas kaalud arvutati, sh välja toodud kasutatud tunnused ja kaalumisel kasutatud parameetrite allikad?*“ 7 uuringut hinnati „Jah“, 23 uuringut hinnati „Ei kehti“ kategooriasse. Ühtegi „Ei“ hinnangut ei antud.

Selle tulemuse puhul tuleb arvestada, et enamik töid ei käsitlenud kaalumist üldse, ei olnud mainitud, kas kaale kasutati ning puudus igasugune viide nende võimaliku rakendamise kohta. Selliste juhtumite puhul hinnati „Ei kehti“, sest kaalumise olemasolu ei olnud võimalik tuvastada, mitte seetõttu, et kirjeldus olnuks puudulik. Samas ei saa välistada, et mõnes või isegi enamikus nendest uuringutest andmeid tegelikult kaaluti, kuid seda ei olnud aruandes adekvaatselt dokumenteeritud. Sellisel juhul oleks tegemist tõsise puudusega, sest puuduliku dokumentatsiooni tõttu vastaksid need uuringud tegelikult hinnangule „Ei“. Kõik 7 „Jah“ hinnanguga tööd sisaldasid kaalude käsitlemise juures piisavalt informatsiooni. Tavaliselt oli selgitatud, milliste tunnuste (nt sugu, vanus, piirkond) alusel kaalud arvutati ja viidatud vastavatele allikatele.

Küsimus 14 kodeerimisjuhendis oli: „*Kas on käsitletud andmete imputeerimist, andmestikust eemaldamist või asendamist?*“ Siin hinnati, kas aruandes oli toodud, kuidas käsitleti puudulikke andmeid, kas andmeid imputeeriti, eemaldati või muul viisil töödeldi. 7 raportit hinnati „Jah“, 2 raportit „Ei“ ning 21 raportit paigutati „Ei kehti“ kategooriasse. Sarnaselt kaalumisega, ei olnud enamikes töödes andmete käsitlemist üldse mainitud, mistõttu ei saanud antud kriteeriumi puhul neid positiivselt ega negatiivselt hinnata.

Kahe „Ei“ hinnangu saanud töö puhul oli andmete töötlemist siiski mainitud, viidati andmete puhastamisele või ebakvaliteetsete ankeetide eemaldamisele, kuid ei selgitatud, millised andmed ja kuidas täpselt eemaldati või töödeldi. Ilma selle infota ei saa hinnata, kas töötlus võis mõjutada valimi esinduslikkust või tulemuste tõlgendust. „Jah“ hinnangu saanud töödes oli vastav info selgelt välja toodud. Oli antud selgelt märku, et andmeid on enne analüüsi eeltöödeldud ning kirjeldatud, missuguseid andmeid spetsiifiliselt ja miks.

4.2.7 Osalusmäära suurendamise strateegiad

Küsimus 15 kodeerimisjuhendis oli: „*Kas on ära toodud üksikasjad koostöö saavutamiseks kasutatud strateegiate kohta (nt eelkontaktid, kirjad, skriptid, rahalised motivaatorid ja nende kättetoimetamise viis, kontaktid keeldujate ümberveenmiseks)?*“ 9 uuringut hinnati „Jah“, 16 uuringut „Ei“ ning 5 „Ei kehti“.

„Jah“ hinnanguga töödes oli esitatud teave rahaliste või muude motivaatorite kasutamise kohta (nt kinkekaardid, loosimised) või oli väga selgelt ja detailselt kirjeldatud osalejatega kontakti loomise protsessi. Tüüpiliselt sisaldasid need aruanded teavet kasutatud kanalite, kontaktide ajastuse ning kontaktmaterjalide sisu kohta, mõnel juhul olid need dokumendid ka lisades täistekstidena esitatud, mis võimaldasid saada selge ülevaate.

„Ei“ hinnangu puhul ei olnud probleemiks strateegiate kirjelduste puudumine, vaid pigem see, et nende kirjeldus jäi liiga üldsõnaliseks. Enamasti piirdus info lakoonilise märkusega, et küsitlus saadeti „e-kirja teel“ või „võeti ühendust koolidega“, ilma et oleks täpsustatud, mida kontakt sisaldas, kas kasutati meeldetuletusi, millised olid motiveerivad elemendid või kuidas korraldati ümberveenmist mitteosalejate puhul. See minimaalne detailsus oli peamiseks põhjuseks, miks paljud tööd ei vastanud kriteeriumile.

„Ei kehti“ hinnati raportid, kus ei esitatud üldse teavet osalejatega kontakti või värbamisstrateegiate kohta ja ei olnud põhjust uskuda, et neid strateegiaid rakendati.

Oluline aspekt selle küsimuse puhul oli ka andmekogumise viis, kuna suurem osa valimis olevatest töödest rakendasid ühe või põhimeetodina veebiküsitlust, oleks olnud ootuspärane, et kontaktikirjade sisu ja motivatsioonimeetodid on aruandes standardina esitatud. Tegelikult osutus see pigem haruldaseks. Mitmed aruanded piirdusid formaalse mainimisega, et kontakt loodi, kuid ei andnud lugejale võimalust mõista, milline roll oli suhtlusel vastamismäära mõjutamisel.

4.2.8 Kvoodid ja paneelid

Küsimus 16 kodeerimisjuhendis oli: „*Kas on kirjeldatud kvootide kasutamine?*“ ning küsimus 17 oli „*Kas paneelide puhul on kirjeldatud paneeliliikmete värbamine ja paneeli haldamine?*“. Käesoleva uurimistöö valimis ei leitud ühtegi uurimisaruannet, kus oleks selgelt kasutatud kvote või paneele. Seetõttu hinnati kõik tööd mõlema küsimuse puhul „Ei kehti“.

See tulemus ei viita metoodilise aruandluse puudujääkidele, vaid peegeldab pigem valimi spetsiifikat: tegemist oli uuringutega, kus valimivõtt toimus teistsuguste mehhanismide kaudu. Kvootide ja paneelide metoodiline läbipaistvus on siiski oluline teema laiemas kontekstis, eriti tänapäevaste veebipõhiste uuringute kasvavas mahus. Kui sarnane analüüs tehtaks paneel-uuringute või turu-uuringutele keskenduvate raportite alusel, oleks need küsimused ilmselgelt rakendatavad. Käesolevas valimis jäid need küsimused pigem teoreetilisteks täiendiks, mille kaudu oli võimalik tagada kodeerimisraamistiku täielikkus.

4.2.9 Vastamis- ja osalemismäärad

Küsimus 18 kodeerimisjuhendis oli: „*Kas tõenäosusliku valimi puhul on välja toodud vastamismäär, selle definitsioon ja arvutuskäik? Kas mittetõenäosusliku valimi puhul on välja toodud osalemismäär?*“

11 uuringut hinnati „Jah“, kuna nendes oli vastamismäär või osalemismäär selgelt esitatud. Leidus ka väiksemamahulisi uuringuid, kus oli küll esitatud ainult saavutatud valim, kuid arvestades valimi lihtsust ja läbipaistvust, oli võimalik määr otse välja lugeda. Sellistel juhtudel hinnati, et täiendava määratlemise puudumine ei vähendanud info kasulikkust lugejale.

Ülejäänud 19 töö puhul, mis hinnati „Ei“, oli olukord ühesem, määrasid lihtsalt ei esitatud. Raportid sisaldasid sageli infot ainult selle kohta, kui palju inimesi tegelikult vastas, kuid puudus igasugune viide sellele, kui paljudele inimestele kutse saadeti või mitu inimest kuulusid sihtpopulatsiooni, kelle hulgast osalejad valiti. Sellistes olukordades ei olnud võimalik hinnata mitteosalemise ulatust.

4.2.10 Uuringudisaini piirangud

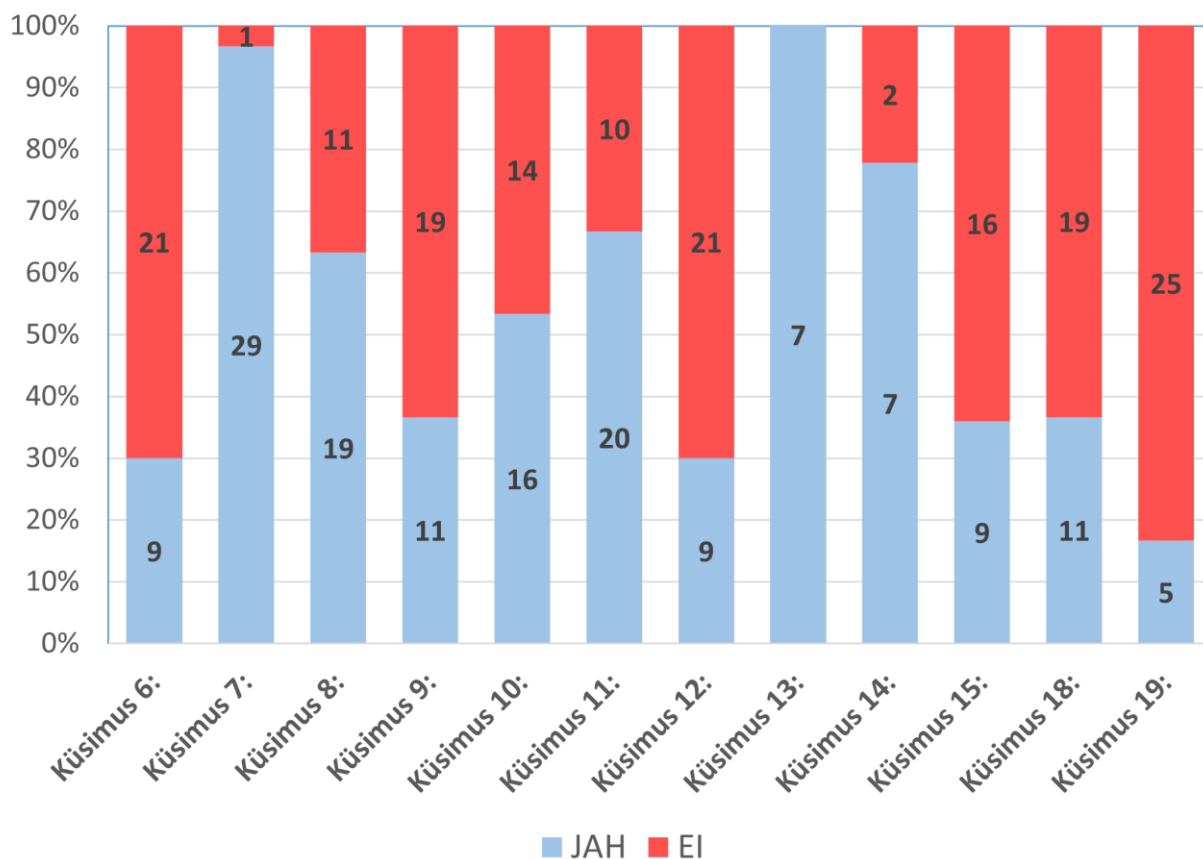
Küsimus 19 kodeerimisjuhendis oli: „*Kas on välja toodud avaldus, milles kirjeldatakse kitsendusi ja piiranguid uuringudisainis ja andmete kogumises?*“ Selle küsimuse kaudu hinnati, kas uuringuaruanded sisaldasid sisulist refleksiooni metodoloogiliste piirangute üle, kas oli esitatud vähemalt üks põhjalik avaldus, milles käsitleti uuringudisaini või andmete kogumisega seotud piiranguid.

Selle kriteeriumi täitsid ainult 5 uuringut. Need tööd hinnati „Jah“, kuna nendes leidis selge ja arusaadav käsitus aspektidest, mis piirasid antud uuringu usaldusväärsust või üldistatavust. See ei tähendanud, et raportis oleks pidanud olema eraldi pealkirjastatud „piirangute“ alapeatükk, piisav oli ka, kui piirangute arutelu leidis hajutatult metoodika osas või analüüsi kirjelduses.

Ülejäänud 25 tööd ei sisaldanud hinnatavat käsitlust uuringudisaini piirangutest ja seega hinnati „Ei“. See oli üks silmatorkavamaid tulemusi terves kodeerimise protsessis, sest isegi töödes, kus metoodika kirjeldused olid üldiselt tugevalt ja detailselt kirjeldatud, puudus sageli igasugune refleksioon piirangute üle.

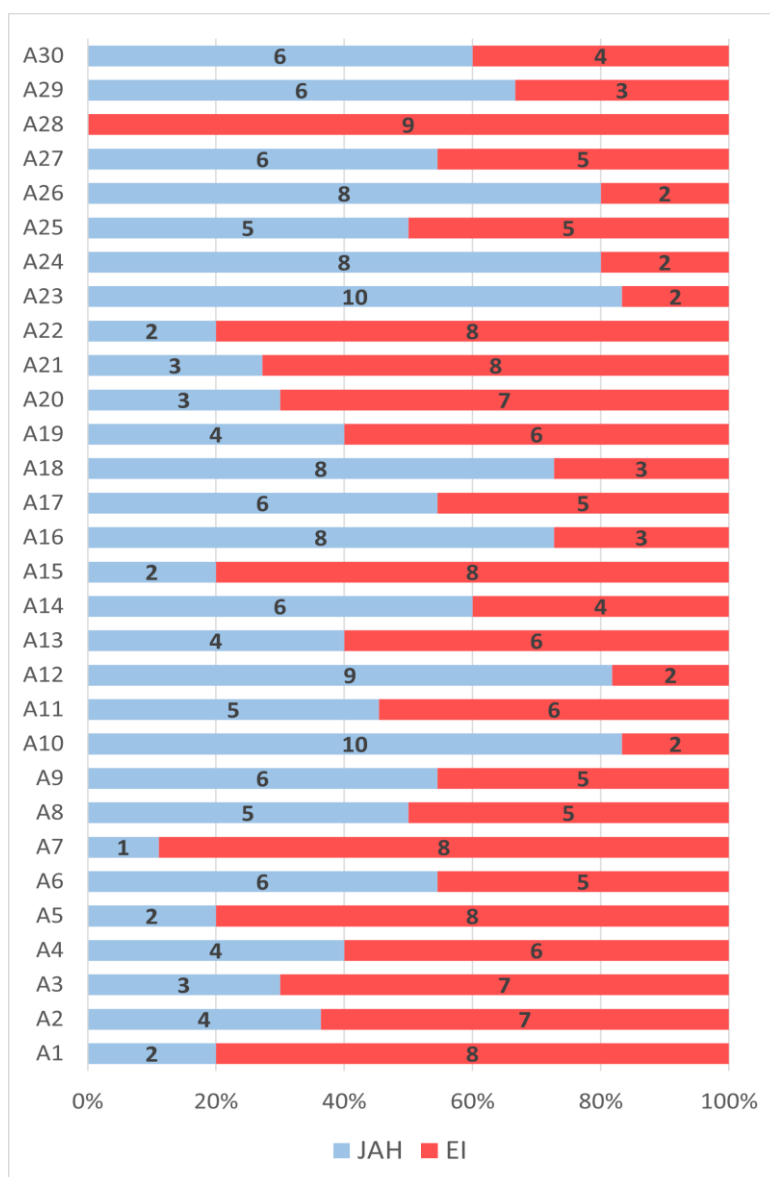
4.2.11 Järeldava analüüsi eelduste kokkuvõte

Joonisel 4 on esitatud järeldava analüüsi eelduste küsimused ning nende täitmise osakaalud. Arvutamisel võeti arvesse ainult need raportite küsimused, mis hinnati "Jah" või "Ei". Küsimused, millele määrati hinnang "Ei kehti" spetsiifiliste raportite hindamisel jäeti analüüsist välja. Seetõttu jäid arvutustest täielikult välja kvootide ja paneelide käsitlemist puudutavad küsimused. Antud joonisel tähistab 100% „Jah“ olukorda, kus uuring täitis kõik antud küsimuse hinnatavad kriteeriumid. Tulemuste põhjal ilmnas, et täidetud kriteeriumite osakaal varieerus märkimisväärselt. Kokku 6 küsimust 12-st hinnati enam kui 50% ulatuses "Jah". Siiski ei tähenda see käesoleva uurimistöö kontekstis, et olukorda võiks pidada positiivseks. Näiteks küsimuse 13 puhul, kus kõigile hinnatud töödele anti hinnang "Jah", võiks tehniliselt arvestada, et kaalumise protseduuride mittekirjeldamine, sõltumata kaalude tegelikust kasutamisest, võiks olla käsitletav veana ning seega "Ei" hinnanguna.



Joonis 4. Järeldava analüüsi eelduste täitmise küsimuste vastuste jaotus („Jah“, „Ei“) 100% skaalal

Joonisel 5 on esitatud järeltava analüüsi eeldusi käsitlevatele küsimustele antud „Jah“ ja „Ei“ hinnangute osakaalud iga valimisse kuuluva raporti kohta. Täpsemal vaatlusel selgub, et 30 uuringuraporti peale kogunes kokku 152 „Jah“ ja 159 „Ei“ hinnangut, mis tähendab, et „Jah“ vastused moodustasid ligikaudu 49% kõigist 311 vastusest. Raportite vahel esines märkimisväärset varieeruvust, osa uuringuraportideid täitis suure osa järeltava analüüsi eelduste kriteeriumidest, samas kui teistes jäid neist enamik täitmata. Kõrge „Jah“ hinnangute osakaaluga raportid eristusid eelkõige põhjaliku ja selgelt struktureeritud metoodikakirjeldusega, mis oli esitatud eraldi metoodikaosas. Vastupidiselt sellele oli raportites, kus metoodikaosa oli napisõnaline või puudulik, otsitavat infot muudes osades märksa raskem leida. Kuigi mõnel juhul esines vajalikke andmeid ka mujal kui metoodikaosas, oli selliseid juhte vähem.

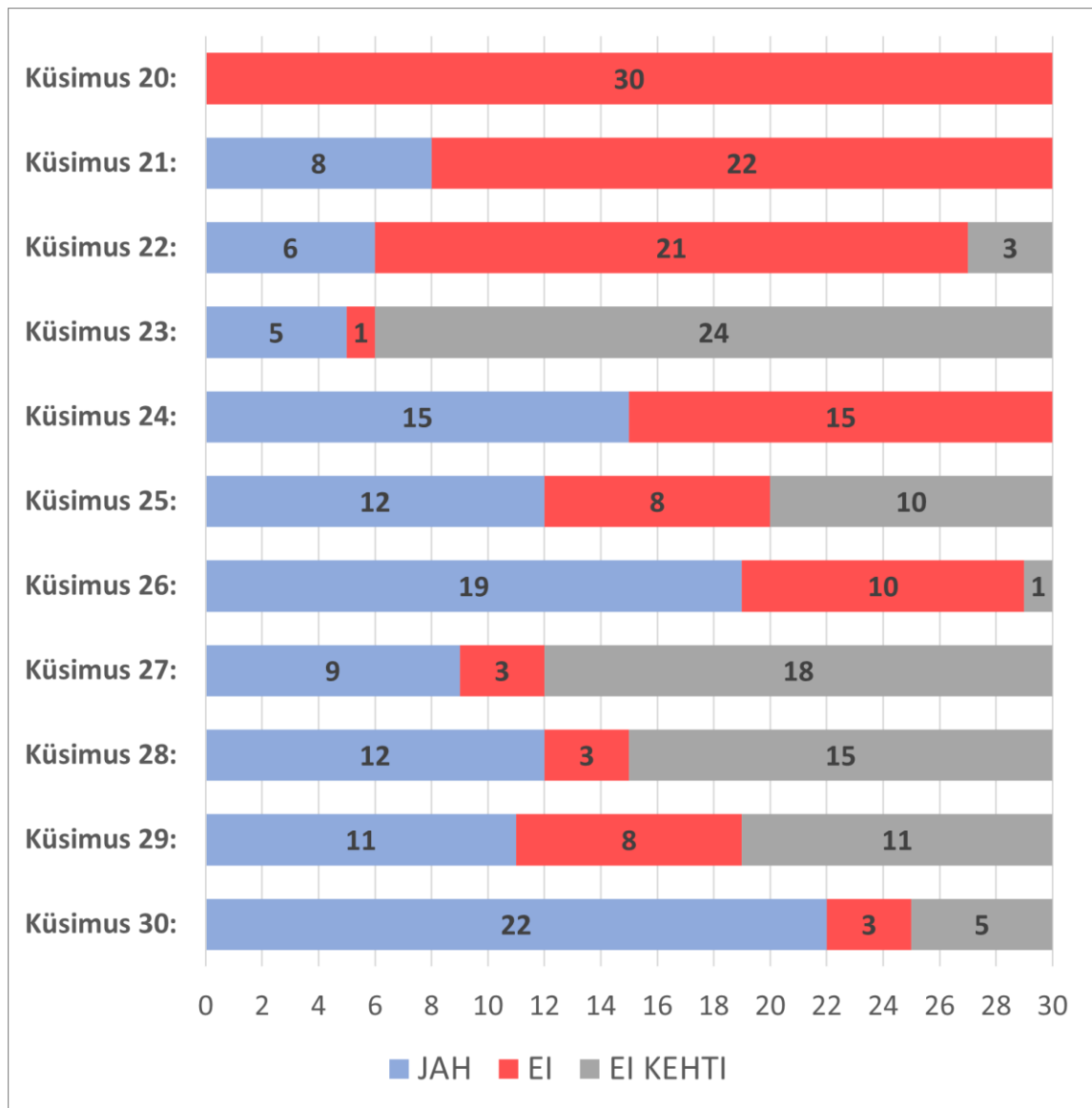


Joonis 5. Järeltava analüüsi eelduste täitmise küsimuste vastuste jaotus iga valimi raporti puhul („Jah“/„Ei“) 100% skaalal

4.3 Järeldav analüüs

Joonisel 6 esitatud sagedusjaotused kajastavad vastuseid kodeerimisjuhendi küsimustele 20–30, mis keskendusid analüüsitulemuste esitamise ja tõlgendamise täpsusele ning statistiliste järelduste põhjendatusele. Terve kodeerimisjuhend koos hindamiskriteeriumite kirjeldustega on leitav lisast 2. Statistiliste näitajate esitusviisi hinnati küsimuses statistiliste tulemuste täpsuse kohta (küsimus 20), millele lisandus hinnang igasugusele p-väärtuste väärtõlgendamise vältimisele (küsimus 21). Usaldusvahemike olemasolu ja kasutamise otstarbekust vaadeldi eraldi võtmetulemuste kontekstis (küsimus 22) ning nende tõlgenduste asjakohasust hinnatud eraldi küsimuses (küsimus 23). Statistiliste järelduste kooskõla analüüsi põhjal leitud tõenditega käsitleti küsimuses 24. Järgnevalt vaadeldi, kas statistiliselt mitteolulisi tulemusi tõlgendati vääralt tõendina seose või erinevuse puudumise kohta (küsimus 25) ning kas välditi eksitavat terminikasutust statistilise ja sisulise olulisuse osas (küsimus 26). Mudelipõhise analüüsi puhul hinnati, kas muutujate kaasamine ja väljajätmine oli sisuliselt põhjendatud (küsimus 27) ning kas koefitsiendid olid esitatud viisil, mis võimaldas hinnata seose tugevust ja tähendust (küsimus 28). Lõpuks hinnati, kas tulemuste sisuline olulisus oli arutelu fookuses (küsimus 29) ning kas oli välditud hinnangute statistilise olulisuse otsest võrdlemist ilma sisulise tähenduseta (küsimus 30).

Need küsimused võimaldasid hinnata, kuivõrd läbipaistvalt ja metoodiliselt põhjendatult olid esitatud ja tõlgendatud uuringute statistilised tulemused ning nende põhjal tehtud järeldused. Kuna analüüsi tulemuste esitamise ja tõlgendamise viis mõjutavad oluliselt järelduste usaldusväärsust ja rakendatavust, oli nende aspektide selgus ja täpsus oluline kogu uurimistulemuste tõlgendatavuse seisukohalt. Järgnevalt tuuakse detailsemalt välja kodeerimistulemused iga küsimuse lõikes. Tulemused on esitatud üldistatud kujul, ilma konkreetsete raportite või asutuste viiteid välja toomata, kuna töö eesmärk ei ole hinnata üksiku uuringu või asutuse töö kvaliteeti, vaid analüüsida rakendusuuringu metodoloogilist praktikat laiemalt.



Joonis 6. Järeldava analüüsi selgust ja adekvaatsust puudutavate küsimuste sagedusjaotused

4.3.1 Statistiliste tulemuste esitamise täpsus

Küsimus 20 kodeerimisjuhendis oli: „Kas statistilised tulemused on esitatud piisava täpsusega (nt täpsed p -väärtused, seose tugevus ja usaldusvahemikud)?“

Kõik valimisse kuulunud 30 raportit hinnati selle küsimuse puhul „Ei“. Selline tulemus ei viita tingimata sellele, et kõikide raportite statistilised tulemused oleksid olnud metoodiliselt ebakorrektsed, pigem oli hindamiskriteerium väga rangelt seatud. Hinnang „Jah“ oleks nõudnud, et kõik p -väärtused oleksid esitatud täpsete arvudena, näiteks $p = 0.035$, mitte üldistusena nagu $p < 0.05$. Kahjuks ei olnud sellist täpsust üheski raportis täidetud.

Kõige sagedasem põhjus negatiivse hinnangu saamiseks oli p-väärtuste liigne üldistamine, kus tulemusi kirjeldati näiteks kujul $p < 0.05$ ilma konkreetset väärtust välja toomata. Mõnes raportis üldistati veel tugevamalt, näiteks leidis paar uuringuraportit, kus tulemusi esitati vaid märkusena, et need on „statistiliselt olulised“, ilma et oleks olulisuse nivood täpsustatud. Leidis paar raportit, kus üldistamine oli tehtud kõigest $p < 0.01$ kujul, aga isegi neid tulemusi piisavalt täpselt lugedes, et anda „Jah“ hinnang, siis käesoleva küsimuse hinnangute sagedusjaotus ei muutuks märkimisväärselt.

Teiste statistiliste näitajate nagu seose tugevuse või usaldusvahemike puhul jäi täpsuse hindamine teisejärguliseks, sest p-väärtuste üldistamine oli piisav, et anda raportile automaatselt „Ei“ hinnang. Vaatamata sellele oli üldmulje siiski, et muude statistiliste näitajate esitus oli valimis enamasti rahuldaval tasemel.

4.3.2 P-väärtuste väärtõlgendamise vältimine

Küsimus 21 kodeerimisjuhendis oli: „Kas p-väärtused on esitatud ja tõlgendatud viisil, mis väldivad väärtõlgendusi?“ Selle küsimuse eesmärk oli tuvastada, kas uuringuaruanded käsitlesid p-väärtusi viisil, mis vastab nende statistilisele sisule, mitte ei anna neile eksitavat või liialdatud tõlgendust. Eelkõige hinnati, kas p-väärtust on esitatud kui „tõendit“, et mingi seos või erinevus eksisteerib või vastupidi.

Hinnangu „Jah“ said ainult 8 raportit. Nendes töödes ei esinenud alati selgesõnalist arutelu p-väärtuse tähenduse üle, kuid nende tulemuste esitusest oli selge, et autorid ei langetanud otsuseid üksnes statistilise olulisuse põhjal. Näiteks heaks praktikaks oli järelduste tegemisel statistilise olulisuse kombineerimine seose tugevusega või argumenteeritud sisulise kaalutlusega.

22 tööd hinnati „Ei“. Siin ei olnud peamine probleem otsesed väärtõlgendused, vaid see, et väärtõlgendusi oli peaaegu võimatu tuvastada, sest tõlgendust lihtsalt ei olnud. Valdavas osas valimi aruannetes kasutati p-väärtust järelduste peamise mõõdupuuna, piirdudes statistilise olulisuse sedastamisega ilma, et oleks võetud arvesse seose tugevust või erinevuse ulatust ja arutlemata, kas tulemused on ka sisuliselt märkimisväärsed. Seda näitas ka tüüpiline vormistus, nagu „X tulemus oli statistiliselt oluline ($p < 0.05$)“. Selle küsimuse alusel aruannete kodeerimine osutus üheks keerulisemaks kodeerimisraamistikus. Kuigi kirjanduse ülevaates on dokumenteeritud tüüpilised p-väärtuse väärtõlgendused, siis käesoleva valimi raportites ei väljendunud need väärtõlgendused nii selgelt. Enamasti ei kirjutatud midagi, mis oleks otseselt

„vale“, vaid probleem oli pigem selles, mis jääb ütle mata: pole refleksiooni selle üle, mida p-väärtus mõõdab, ega arutelu selle piirangutest.

4.3.3 Usaldusvahemike esitamine võtmetulemustes

Küsimus 22 kodeerimisjuhendis oli: „*Kas võtmetulemuste puhul on esitatud usaldusvahemikud?*“ Ainult 6 tööd hinnati „Jah“, mis tähendab, et nendes oli vähemalt mõne põhitulemuse kõrval esitatud usaldusvahemik, tavaliselt regressioonikoefitsiendi puhul.

21 raportit hinnati „Ei“, sest nendes ei kasutatud usaldusvahemikke üldse olukordades, kus need oleksid võinud olla kasulikud, et anda lisainfot tulemuste täpsuse ja valimi hajuvusest tuleneva määramatuse kohta. Siin tuleb sisse keerukas hindamiskomponent – nimelt eeldas antud küsimus hindaja poolset kaalumist: kas oleks olnud mõistlik, et siin oleks esitatud usaldusvahemik? Selle küsimuse puhul on seetõttu märkimisväärselt palju subjektiivsust, kas hinnata midagi „Ei“ või „Ei kehti“. Käesoleva kodeerimise käigus valdavalt järeldati, et usaldusvahemikest oleks olnud käesolevates raportites piisavalt kasu, et hinnata see kriteerium „Ei“.

Kolme töö puhul hinnati „Ei kehti“, kuna nende struktuur või analüütiline lähenemine oli selline, kus usaldusvahemikud ei olekski andnud mingisugust sisulist tähendust, näiteks väga lihtsad kirjeldavad uuringud väga piiratud statistiliste võrdlustega.

4.3.4 Usaldusvahemike väärtõlgendamise vältimine

Küsimus 23 kodeerimisjuhendis oli: „*Kas usaldusvahemikud on esitatud ja tõlgendatud viisil, mis kajastab määramatust tulemustes ja väldib eksitavaid üldistusi?*“ Selle küsimuse eesmärk oli hinnata, kas usaldusvahemike tõlgendused uuringuaruannetes peegeldasid nende tegelikku statistilist tähendust ehk kas vahemikku käsitleti korrektselt kui hinnangu määramatuse indikatsiooni, mitte kui tõenäosusväidet või kindlat piiri, kuhu tegelik väärtus kindlasti kuulub.

Käesolevas valimis hinnati 5 tööd „Jah“, 1 töö „Ei“, ja 24 tööd „Ei kehti“. Nii palju „Ei kehti“ hinnanguid oli selle pärast, et usaldusvahemikke esitati vaid kuues töös, mistõttu olid enamik raporteid sellised, kus lihtsalt polnud võimalik hinnata, kas tõlgendus oli korrektne või mitte.

Nendes üksikutes töödes, kus usaldusvahemikud olid esitatud ja hinnanguks kujunes „Jah“, ei antud enamasti põhjalikku ega süstemaatilist selgitust usaldusvahemike tähenduse kohta. Pigem piirduti tõlgendustega, mis vältisid otseselt eksitavat keelekasutust, kuid ei selgitanud statistilise näitaja tähendust väga sisuliselt.

Ainus „Ei“ hinnangu saanud töö sisaldas spetsiifilist eksitavat väidet: „*Usaldusintervall näitab vahemikku, mis sisaldab 95% tõenäosusega populatsiooni keskmist*“. See on küll laialt levinud sõnakasutus, kuid on tegelikult üpris ekslik, sest pärast usaldusvahemiku arvutamist antud andmete põhjal ei saa enam rääkida tõenäosusest, tõenäosuse mõiste kehtib vaid protseduuri kordamise kontekstis, mitte ühe konkreetse vahemiku puhul.

4.3.5 Statistiliste järelduste ja analüüsitulemuste kooskõla

Küsimus 24 kodeerimisjuhendis oli: „*Kas statistilised järeldused on kooskõlas analüüsi põhjal leitud tõenditega?*“ Selle küsimuse kaudu hinnati, kas uuringuaruannetes esitatud statistilised järeldused põhinesid tehtud analüüsidel või kalduti üleliigselt üldistama, toetudes näiteks ainult statistilisele olulisusele või p-väärtustele ilma sisulise kontekstita. Valimis jagunesid tulemused täpselt pooleks – 15 tööd hinnati „Jah“ ja 15 tööd „Ei“.

„Jah“ hinnang ei eeldanud tingimata keerukat metodoloogilist refleksiooni, piisav oli, kui töö tulemuste tõlgendamisel ei kasutatud statistilist olulisust ainsa argumendina. Mitmes sellises raportis võeti arvesse ka seose tugevus või usaldusvahemikke (töodes, kus need olid olemas) ja tõlgendused jäid tagasihoidlikuks ning mõõdukaks.

Samas hinnati pooled tööd „Ei“, sest nendes ilmnes selgelt see, kui tugevalt on p-väärtus jäänud domineerima järelduste kirjeldamises. Levinud muster oli näiteks: „...tulemus on statistiliselt oluline ($p < 0.05$)... X mõjutab Y“, ilma täiendava aruteluta seose tugevuse, tähenduse või määramatuse kohta. Selliste juhtumite puhul jäi mulje, et järeldust ei tuletatud mitmetahulisest tõlgendamisest, vaid üksnes statistilise olulisuse olemasolust.

Eraldi probleem on see, et paljud aruanded ei käsitlenud üldse seda, kuidas nende järeldused võiksid analüüsi piirangutega suhestuda. Näiteks kui valim oli väike või seos nõrk, aga tulemus „statistiliselt oluline“, siis paljudes raportites ei tulnud esile kaalutlust, mida see järelduste jaoks tähendab.

4.3.6 Statistiliselt mitteoluliste tulemuste tõlgendamine

Küsimus 25 kodeerimisjuhendis oli: „*Kas välditakse statistiliselt mitteoluliste tulemuste tõlgendamist tõendina seose või erinevuse puudumise kohta?*“ Selle küsimuse keskmes oli hindamine, kas uuringuaruannetes käsitleti statistiliselt mitteolulisi tulemusi ettevaatlikult ja metoodiliselt vastutustundlikult, see tähendab, et kas välditi ekslikku järeldust, et „tulemus ei

olnud statistiliselt oluline = seos puudub“. Antud kriteeriumi puhul jagunesid tulemused kolme gruppi: 12 tööd hinnati „Jah“, 8 tööd „Ei“, ning 10 tööd „Ei kehti“.

Hinnang „Jah“ anti eelkõige töödele, kus statistiliselt mitteolulisi tulemusi kas ei tõlgendatud üle või nendest räägiti tagasihoidlikult. Mõnel juhul lihtsalt nimetati tulemus statistiliselt mitteoluliseks ilma sellele sisulise tähenduse omistamiseta, mida hinnati pigem positiivselt, kuna see vähendas eksliku seose loomise ohtu. Väga harvadel juhtudel leidis märkus, et mitteolulisus võib olla tingitud näiteks valimi suurusest või testimise vähesest võimsusest, mida võib pidada meetoodiliselt heaks praktikaks.

Hinnang „Ei“ anti raportitele, kus mitteolulise tulemuse järel esitati otsekohene sisuline järeldus, justkui see kinnitaks mingit nähtuse puudumist. Sellised väited ignoreerisid võimalust, et test ei olnud piisavalt tundlik või seda, et seos või erinevus võis siiski olemas olla. Seos või erinevus võis olla lihtsalt väike või väikese valimi tõttu ei õnnestunud seost statistiliselt tuvastada. Tegemist oli eksitava loogikaga, kus p-väärtuse alusel vaikumisi eeldati, et justkui statistiliselt mitteoluline tulemus võimaldaks nullhüpoteesi kinnitada.

Huvitava nüansina hinnati kolmandik raporteid „Ei kehti“, sest statistiliselt mitteolulistest tulemustest ei räägitud üldse. Nende tööde puhul oli tähelepanu ainult statistiliselt olulistel tulemustel ning statistiliselt mitteolulisi kas ei mainitud või need jäeti teadlikult arutelust välja, näiteks märkides: „Aruandes on esitatud ainult statistiliselt olulised erinevused ($p < 0.05$)“. See ei tähenda automaatselt head praktikat, vaid pigem viitab sellele, et statistiliselt mitteoluliste tulemuste roll uuringute tõlgenduses on sageli marginaliseeritud. Lisaks peitub siin ka oht eksijärelduste tekkeks, sest ainult statistiliselt oluliste tulemuste esitamine võib kallutada üldpilti. Näiteks juhul, kui populatsioonis tegelikult seoseid ei eksisteeri ja juhuvalimi põhjal tehakse palju teste, võidakse üksikutel juhtudel jõuda statistiliselt oluliste tulemusteni hoolimata seoste puudumisest populatsioonis. Kui sellisel juhul pannakse aruandesse ainult statistiliselt olulised tulemused ja statistiliselt mitteolulised jäetakse välja, luuakse ekslik mulje, et leitud seosed või erinevused on tegelikult olemas ja usaldusväärsed, kuigi tegelikult on tegemist esimest liiki vea ilminguga.

4.3.7 Statistilise ja sisulise olulisuse eristamine

Küsimus 26 kodeerimisjuhendis oli: „Kas välditakse terminite „oluline“ ja „mitteoluline“ mitmetähenduslikul või eksitaval viisil kasutamist?“ Selle küsimuse eesmärk oli hinnata, kas uurimisaruannetes on kasutatud terminoloogiat viisil, mis teeb selgeks, kas jutt käib statistilisest

või sisulisest olulisusest. Selle küsimuse puhul hinnati 19 tööd „Jah“, 10 tööd „Ei“, ning 1 töö „Ei kehti“. Selle küsimuse puhul avaldus eriti selgelt hindaja rolli subjektiivsus, sest terminite mõistmise selgus sõltub paljuski sellest, kuidas lugeja ise neid sõnu tõlgendab.

„Jah“ ei tähendanud tingimata seda, et igas raportis oleks igal kasutuskorral olnud selge deklaratsioon stiilis „statistiliselt oluline ($p < 0.05$)“, vaid hinnati positiivselt siis, kui terminite kasutus oli loogiline ja üheselt mõistetav, näiteks kui oli selgelt viidatud kas p-väärtusele või sisulisele tähendusele või kui kontekst välistas arusaamatuse.

„Ei“ hinnati need tööd, kus sõna „oluline“ esines tulemuste kontekstis, aga polnud selge, kas peeti silmas statistilist või sisulist olulisust. Mõned raportid viitasid lihtsalt „olulisele erinevusele“ või „mitteolulisele tulemusele“ ilma ühegi täpsustava märkuseta, millest võib tavalisel lugejal kergesti tekkida mulje, et „oluline“ tähendab „tugev seos“ või „tähtis järeldus“. Osades raportites jäeti mulje, et statistiline olulisus võrdubki sisulise tähendusega, mis on ilmselgelt ekslik.

Üks töö hinnati „Ei kehti“, sest terminite „oluline“ või „mitteoluline“ kasutus puudus täiesti. Samas oli see töö nende hulgas, kus ükski järeldava analüüsi osa ei saanud hinnangut „Jah“, mistõttu võib pidada seda raportit erandlikuks.

4.3.8 Mudeli muutujate kaasamise ja väljajätmise põhjendatus

Küsimus 27 kodeerimisjuhendis oli: „*Kas välditakse mudelist muutujate väljajätmise põhjendamist ainult nende hinnangute statistilise olulisuse põhjal?*“ Kogu valimi lõikes hinnati 9 tööd „Jah“, 3 tööd „Ei“, ning 18 tööd „Ei kehti“. Viimane kategooria sisaldas töid, kus mingeid järeldavaid mudeleid ei kasutatud või kus mudelistruktuure polnud piisavalt esitatud, et seda küsimust üldse hinnata.

Tööd, mis hinnati „Jah“, ei olnud tingimata väga põhjalikud oma mudelivaliku kirjeldustes, kuid nendes ei jäänud muljet, et muutujaid oleks kaasatud või eemaldatud ainult p-väärtuste põhjal. Nendes töödes põhimõtteliselt viidati otseselt sellele, et muutujaid valiti varasema uurimuse alusel või arutleti nende sisulise asjakohasuse üle. Mõnikord võidi statistilist mitteolulisust mainida, aga juhul kui seda ei kasutatud ainsa põhjendusena muutuja väljajätmiseks, siis selline käsitus oli kooskõlas selle küsimuse eesmärgiga.

„Ei“ hinnangu saanud juhtumid olid seevastu huvitavad oma sõnastuse tõttu. Näiteks ühes töös oli selgelt välja toodud, et muutuja jäeti mudelist välja, kuna see ei osutunud „statistiliselt oluliseks“. Probleem ei seisnenud mitte selles, et see oleks vale otsus, vaid selles, et ainus antud põhjendus

oligi p-väärtuse puudumine, mis ignoreerib võimalust, et muutuja võib olla sisuliselt oluline isegi siis, kui selle mõju ei ole piisavalt täpselt mõõdetav. Kahe teise „Ei“ hinnangu saanud töö puhul ei olnud muutujate valiku protsess piisavalt lahti seletatud ja statistiline olulisus oli jäetud ainukese indikatsioonina sellest, mis võis juhtida sellist muutujate valikut.

4.3.9 Koefitsientide esitamise selgus

Küsimus 28 kodeerimisjuhendis oli: „*Kas koefitsiendid on esitatud otstarbekal ja arusaadaval kujul, mis hõlbustab seose tugevuse mõistmist?*“ Selle küsimuse kaudu hinnati, kas uuringuaruannetes esitatud regressioonikoefitsiendid ja muud seose tugevust kirjeldavad näitajad on esitatud nii, et nende põhjal oleks lihtsam hinnata seoste tugevust ja tähendust, eriti juhtudel, kus kasutati keerukamaid mõõdikuid, näiteks logit-suhet või teisendatud skaalaga tunnuse regressioonikordajat. Hindamine keskendus sellele, kas tulemuste tõlgendamine oli toetatud arusaadava esitlusviisiga, näiteks tabelite, jooniste või lisaselgitustega. 12 tööd hinnati „Jah“, 3 tööd „Ei“ ja 15 tööd „Ei kehti“. Viimane hinnang viitas juhtudele, kus koefitsiente ei olnud üldse esitatud, näiteks ei kasutatud uuringutes regressioonimudeleid või muid kvantitatiivseid mudeleid, mille tulemusel oleks olnud vaja mingit koefitsientide tõlgendust teha.

„Jah“ hinnangu saanud tööd jagunesid kaheks. Esiteks need, kus koefitsiendid olid toodud tabelitena või graafiliselt ning olid esitatud loogilises ja arusaadavas kontekstis. Ja teiseks need, kus keerukamate koefitsientide puhul oli lisatud täiendavaid selgitusi, näiteks astendatud vormis, mis muutsid tulemuste praktilise tähenduse arusaadavaks.

Hinnang „Ei“ anti raportitele, kus koefitsiendid küll esinesid, kuid olid esitatud kas ilma selgituseta või viisil, mis ei võimaldanud hinnata nende tähendust. Ühes töös oleks võinud eeldada, et mudeli tulemused (sh koefitsiendid) on esitatud, kuid need olid välja jäetud ilma igasuguse põhjendusega. Kahes teises esines olukord, kus keerukad mudelikoefitsiendid olid toodud küll tabelitena, kuid ilma ühegi täiendava kontekstuaalse kommentaarita. See küsimus tõi hästi välja, et koefitsientide esitusel peaks pöörama tähelepanu mitte ainult tehnilisele täpsusele, vaid ka kommunikatiivsele selgusele.

4.3.10 Sisuline ja praktiline olulisus

Küsimus 29 kodeerimisjuhendis oli: „*Kas arutletakse tulemuste nagu teststatistikute, seosekordajate ja koefitsientide sisulise olulisuse üle, mitte ainult statistilise olulisuse üle?*“ 11

tööd hinnati „Jah“, 8 tööd „Ei“ ja leidis koguni 11 tööd, mis hinnati „Ei kehti“, sest nendes ei olnud selliseid statistilisi tulemusi.

Hinnang „Jah“ anti raportitele, kus leidis vähemalt mingil kujul arutelu tulemuste tähenduslikkuse üle – näiteks, millist sisulist või praktilist tähtsust võiksid leitud seosed, erinevused või mõjud omada uuritava nähtuse või konteksti seisukohalt. Selline arutelu ei pidanud olema väga põhjalik, kuid juba näiteks mõte, et „kuigi statistiline olulisus puudus, võib mõju olla sisuliselt märkimisväärne“, andis tunnistust sisulisest kaalutlusest.

„Ei“ hinnangu saanud tööd olid need, kus kogu tulemuste tõlgendus taandus statistilisele olulisusele. Nende raportite puhul oli selgelt kogu tulemuste esitamise ja arutelu kirjutamine ehitatud ainult p-väärtuste ümber. Sisulisemad küsimused või arutlused jäid pigem tahaplaanile.

4.3.11 Statistiliste olulisuste võrdlemise vältimine

Küsimus 30 kodeerimisjuhendis oli: „*Kas välditakse statistikute võrdlemist pelgalt nende statistilise olulisuse või olulisuse tõenäosuste põhjal?*“ See küsimus suunas tähelepanu ühele üsna spetsiifilisele, aga levinud eksitusele statistilises järeldamises, kus võrreldakse eri näitajaid lihtsalt selle põhjal, kas ühe p-väärtus on teisest väiksem, jättes kõrvale tegeliku erinevuse suuruse või seose konteksti. Käesolevas valimis hinnati 22 tööd „Jah“, 3 tööd „Ei“ ja 5 tööd „Ei kehti“. „Ei kehti“ raportid olid natukene erandlikud, kuna nendes oli tekstisiselt statistiliste järelduste kirjeldamine puudulik, aga ikka oli tegemist töödega, mis sobisid valimisse.

Suur osa töödest pääses „Jah“ hinnanguga üsna lihtsalt, mitte selle tõttu, et oleks eraldi rõhutatud mõistlikku statistiliste erinevuste käsitlemist, vaid sest nad lihtsalt ei võrrelnud statistiku omavahel või tegid seda hoolikalt ja sisuliselt.

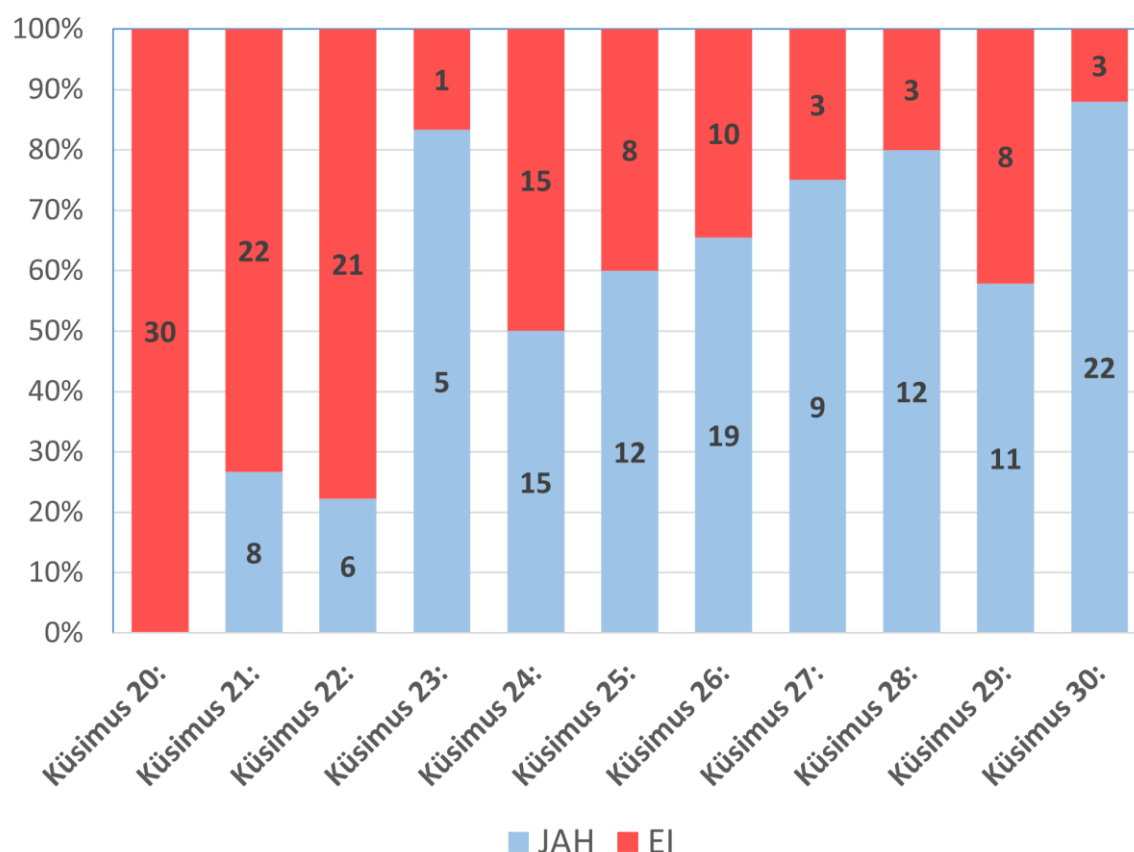
Samas hinnati kolm tööd „Ei“, sest nendes leidis otsene vihje sellele, et statistilise olulisuse aste, näiteks üks tulemus on „statistiliselt oluline“, teine „statistiliselt mitteoluline“, tähendab kohe statistilist ja sisulist erinevust kahe tulemuse vahel. Näiteks esitati tulemused kõrvuti ja nendest järeldati midagi erinevuse kohta ilma, et oleks hinnatud sisulist konteksti.

Kogu selle küsimuse hindamisel oli oluliseks metodoloogiliseks raskuseks see, et võrdluste sisu leidmine eeldas üsna tähelepanelikku lugemist ja konteksti mõistmist. Põhimõtteliselt ei olnud töödes tegu otseste väidetega nagu „üks seos on teisest tugevam, sest esimese puhul on olulisuse tõenäosus madalam kui teisel“. Selle asemel võis võrreldavate tulemuste tähtsust eristada viisi

järgi, kuidas neid kirjeldati, näiteks milliseid sõnu kasutati tulemuste rõhutamiseks, milliseid järeldusi neist tehti või millise tähelepanu mõni tulemus teksti sees sai võrreldes teistega.

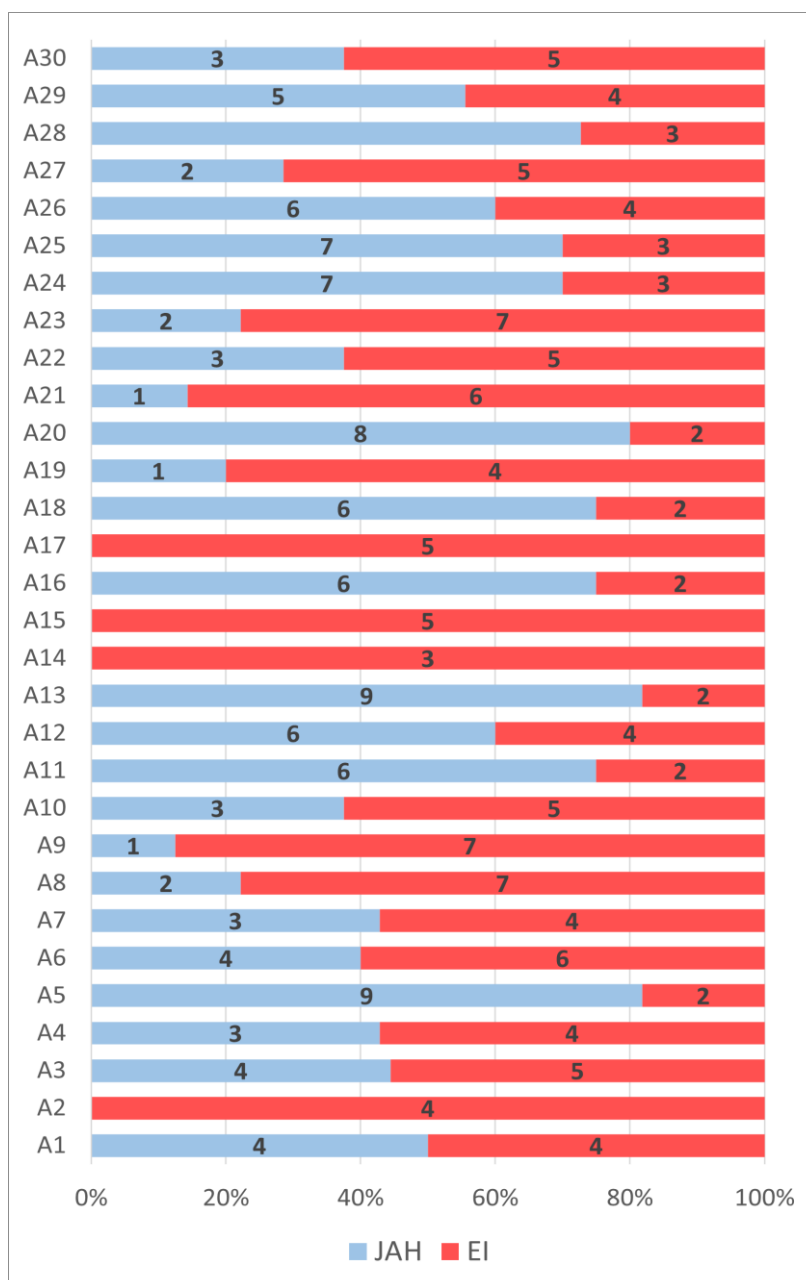
4.3.12 Järeldava analüüsi kokkuvõte

Joonisel 7 on esitatud järeldava analüüsi selguse ja adekvaatsuse küsimused ning nende täitmise osakaalud. Arvutamisel võeti arvesse ainult need raportite küsimuste hinnangud, kus anti hindeks "Jah" või "Ei". Küsimused, millele määrati hinnang "Ei kehti" spetsiifiliste raportite hindamisel jäeti analüüsist välja. Antud joonisel tähistaks 100% „Jah“ olukorda, kus uuring täitis kõik antud küsimuse hinnatavad kriteeriumid. Järeldava analüüsi küsimuste olukord on antud kriteeriumite põhjal parem, kui eelduste küsimuste oma. Kokku 8 küsimust 11-st hinnati 50% või enam ulatuses „Jah“. Nagu järeldava analüüsi eelduste kokkuvõttes välja toodi, siis see tegelikult ei tähenda head olukorda, kuna nagu läbi eelnevate seksioonide kirjeldati, „Ei kehti“ tulemused on tihti pigem negatiivse varjundiga.



Joonis 7. Järeldava analüüsi selguse ja adekvaatsuse küsimuste vastuste jaotus („Jah“/„Ei“) 100% skaalal

Joonisel 8 on esitatud järeltava analüüsi selguse ja adekvaatsuse küsimustele antud „Jah“ ja „Ei“ hinnangute osakaalud iga valimisse kuuluva raporti kohta. Täpsemal vaatlusel selgub, et 30 uuringuraporti peale kogunes järeltava analüüsi selgust ja adekvaatsust puudutavatele küsimustele kokku 119 „Jah“ ja 124 „Ei“ hinnangut. Seega moodustasid „Jah“ vastused ligikaudu 49% kõigist 243 vastusest. Mõned raportid vastasid enamikule kriteeriumitest, samas kui mõnes teises ei olnud täidetud ükski. Uuringute lõikes oli keeruline tuvastada selgeid tunnusjooni, mille põhjal oleks saanud töid sisuliselt grupeerida ja eristada või põhjendada, miks mõnes raportis oli rohkem kriteeriume täidetud kui teistes.



Joonis 8. Järeltava analüüsi selguse ja adekvaatsuse küsimuste vastuste jaotus iga valimi raporti puhul („Jah“/„Ei“) 100% skaalal

5. ARUTELU JA JÄRELDUSED

Käesolev peatükk koondab uurimistöö peamised järeldused ja pakub nendele sisulist arutelu. Eesmärk on mitte ainult vastata uurimistöös püstitatud kahele uurimisküsimusele, vaid ka analüüsida leide laiemas metoodilises ja teoreetilises kontekstis, käsitledes nende praktilist tähendust.

Uurimistöös püstitatud uurimisküsimused:

1. Kuivõrd on Haridus- ja Teadusministeeriumi tellimusel läbi viidud rakenduslikes uuringutes kasutatavate andmete puhul täidetud järeltava analüüsi eeldused?
2. Kuivõrd on Haridus- ja Teadusministeeriumi tellimusel läbi viidud rakenduslikes uuringutes tehtud statistilised järeldused selgelt ja adekvaatselt sõnastatud?

Need järeldused ja nende arutelu kujunevad välja kontentanalüüsi tulemustel, mille käigus hinnati 30 valimipõhise rakendusliku uuringu metoodikakirjeldust ja statistiliste järelduste esitlust vastavalt eelnevalt koostatud kaheosalise kodeerimisjuhendi kriteeriumitele. Töö analüütiline fookus oli suunatud sellele, kas ja kuidas on järeltava statistika eeldused nendes aruannetes kirjeldatud, eriti valimi moodustamise ja kvaliteedi, näiteks valikumeetodi, valimiraami, vastamismäära, kaalumise ja andmete järeltöötlemise aspektidest lähtudes. Samuti hinnati, millisel määral vastavad esitatud statistilised tulemused tõlgendamisstandarditele, seal hulgas, kuidas kasutatakse p-väärtusi ja usaldusvahemikke ning kuidas on sõnastatud järeldused.

Peatükk ei käsitle ainult tehnilist vastavust formaalsetele nõuetele, vaid keskendub sellele, kuivõrd võimaldab esitatud teave hinnata analüüsi usaldusväärsust ning kui selgelt on uuringu tulemused arusaadaval ja veenval viisil lugejale esitatud. Järeldused on tuletatud nii tulemuste peatükis välja toodud kodeerimise tulemuste kui ka kirjanduse ülevaates loodud teoreetilise raamistiku põhjal, mille kohaselt tuleb järeltava statistika rakendamisel järgida teatud eeltingimusi, et tagada tulemuste üldistatavus ning vältida väärtõlgendusi.

Seega moodustab käesolev peatükk töö sisulise tuuma, kus leidude põhjal sõnastatakse vastused uurimisküsimustele ning rõhutatakse peamisi metoodilisi probleeme ja tugevusi, mis iseloomustavad valitud Haridus- ja Teadusministeeriumi tellimusel valminud rakendusuuringute aruandeid aastatel 2013–2024.

5.1 Järeldava analüüsi eelduste hinnangud

Kontentanalüüsi tulemusena ilmnes, et valdav osa valitud Haridus- ja Teadusministeeriumi tellimusel koostatud rakenduslike uuringute aruannetest, milles kasutatakse järeldava statistika meetodeid, ei vasta täiel määral järeldava statistika rakendamiseks vajalikele metoodilistele eeldustele. Ehkki statistilisi meetodeid kasutatakse kõigis analüüsitud töödes, jääb sageli puudulikuks nende meetodite rakendatavuse eelduseks olevate tingimuste kirjeldamine. Tulemused viitavad selgelt, et analüüsides ei võeta alati arvesse, kuidas valimi kujundamine, valikumeetod ja vastamisprotsessid mõjutavad järelduste üldistatavust. See võib seada kahtluse alla osade järelduste kehtivuse sihtpopulatsiooni suhtes.

Esiteks tuleb esile tõsta süstemaatilist puudulikkust valikumeetodite kirjeldamisel. Enam kui kahes kolmandikus juhtudest ei olnud üheselt tuvastatav, kas ja millisel määral oli valimimeetod tõenäosuslik või mittetõenäosuslik. Kuna statistiliste üldistuste metodoloogiline põhjendus sõltub valimi moodustamise viisist ja osalejate kaasamise protsessi läbipaistvusest, tähendab see, et osade aruannete puhul puudub usaldusväärne alus tehtud järelduste kehtivuse hindamiseks. Probleem ei seisne üksnes ebapiisavas dokumenteerimises, vaid järeldava analüüsi aluseks oleva usaldusväärse tagamises – ilma teabeta valimi koostamise protsessi kohta ei ole võimalik hinnata, kuivõrd on üldistused metodoloogiliselt õigustatud. Valimis on rohkem töid, mis kasutavad mittetõenäosuslikke meetodeid, kui tõenäosuslikke, mistõttu on igasugused puudulikud kirjeldused veel problemaatilisemad – Clark jt (2021: 168) toovad välja, et kui valimi moodustamisel kasutatakse mittetõenäosuslikku valikumeetodit, on tõenäoline, et valiku kujundab inimlik hinnang, mille tulemusel suureneb osade populatsiooniliikmete valimisse sattumise tõenäosus.

Samuti näitab valimiraamide ebapiisav kirjeldus, et uuringu läbiviijad ei ole alati raportis reflekteerinud selle üle, kuivõrd nende valik hõlmab sihtpopulatsiooni tervikuna. Clark jt (2021: 168) kirjeldavad, et valikuraami puudulikkuse või ebatäpsuse puhul ei saa valim esindada sihtpopulatsiooni hoolimata sellest, kas tegemist on tõenäosusliku või mittetõenäosusliku valikumeetodiga. Kui valikuraami ei esitata või ei analüüsita selle kaetust, ei saa lugeja teada, kas osa sihtpopulatsioonist jääb kaasamata või uuringutulemused peegeldavad ainult osasid segmente. Selline alakaetus võib olla süsteemne, mõjutades otseselt tulemuste usaldusväärsust. Veelgi enam, kuna valikuraamile antav hinnang nõuab detailsust, ei piisa ainult selle nimetamisest. Selle puudujääk tähendab, et ka kaetuse vea hindamine ei ole võimalik.

Valimi esinduslikkuse kirjeldamine, samuti planeeritud ja saavutatud valimi võrdlemine esines harva. Sellest järeldub, et uuringute läbiviijad kas ei ole alati pööranud piisavalt tähelepanu valimi kvaliteedi hindamisele pärast andmete kogumist või sellele on tähelepanu pööratud, kuid seda ei ole kirjeldatud aruandes. Valimi esinduslikkus on otseselt seotud võimalusega järeldusi üldistada, kui seda ei hinnata ega esitata, ei saa eeldada, et järeldused on sihtpopulatsiooni suhtes kehtivad – nagu Traugott (2004: 967) välja toob, terve valimi disainimise põhimõte on valida sihtpopulatsiooni liikmeid viisil, mis kogumikuna sihtpopulatsiooni esindab. Lisaks sellele muudab planeeritud ja tegeliku valimimahu erisuse mitteavaldamine raskemaks hinnata, kas valimi koostamisel või uuringu läbiviimisel on esinenud kõrvalekaldeid, mis võivad mõjutada tulemusi. Enamjaolt ei ole teada, kuivõrd on tegelikult valimis olevate uuringute läbiviijad enne töö alustamist oma valimeid planeerinud – Warner (2013: 327, 333) toob välja, et valimi planeerimine on eriti kasulik statistiliste järelduste vaatepunktist, kuna uurija saab planeeritud valimi suuruse põhjal hinnata, kuidas on valimi planeerimisel arvestatud statistiliste testide võimsusega.

Teiseks olid märgatavalt alaesindatud kaalumise ja puuduvate ning puudulike andmete käsitlemisega seotud aspektid. Komplekssete valimidisainidega uuringutes on tavaline, et valimi üksustel on erinevad valikutõenäosused, vastamismäärad varieeruvad oluliste alarühmade lõikes ning sihtrühmade jaotused ei lange kokku populatsiooni teadaolevate omadustega, sellistes olukordades kasutatakse sageli kaale, et neid erisusi tasakaalustada (Groves jt, 2009: 348). Eraldi probleemiks on see, et enamikes töödes ei selgitatud üldse, kas selliseid protseduure rakendati või mitte – tegemist on olukorraga, kus kaalumise või andmete töötlemine võib oluliselt muuta analüüsi tulemusi, seega sellise info puudumine on tõsine läbipaistvuse probleem.

Kolmandaks olid osalejatega kontakteerumise ja valikukriteeriumite esitusega seotud aspektid vähesel määral läbipaistvusega, mis raskendab oluliselt hinnangute andmist võimalikele valiku- ja mittevastamise vigadele. Kontaktistrateegiate ning värbamismeetodite puhul piirdusid paljud aruanded üldsõnalise kirjeldamisega või ei sisaldanud seda teavet üldse. Eriti problemaatiline on olukord, kus ei ole teada, kas ja milliseid meetmeid kasutati osalemismäära suurendamiseks. Kuna sellised strateegiad võivad mõjutada vastajate profiili ning seega ka valimi esinduslikkust on nende esitamata jätmine sisuliselt oluline analüütiline puudujääk. Ühe näitena, nii Bryman (2012: 236) kui ka Gallagher (2004: 740) rõhutavad, et parim meetod tõsta osalemismäära, on pakkuda vastajatele raha, isegi väikesest kogusest piisab, et oluliselt määrata tõsta. Igasugustel erinevatel strateegiatel võivad olla erinevad mõjud erinevate gruppide vastamismääradele. Kahjuks

käesolevas uurimistöös vaadeldud rakendusuuringute puhul ei ole alati võimalik leida, kas neid strateegiaid lihtsalt ei rakendatud või need jäid raportis kirjeldamata.

Valimite esinduslikkuse puudujäägid ei ole alati paratamatud, aga sellistel juhtudel eeldatakse uurijatelt reflektiivset käsitlust, see tähendab, et aruandes tuleb selgelt välja tuua, kelle suhtes tulemusi saab üldistada ja kelle suhtes mitte. Kui valimi koostis erineb süsteemselt sihtrühmast, näiteks on mõni rühm ala- või üleesindatud, tuleks seda käsitleda ja võimalusel korrigeerida. Analüüsitud uuringutes esines harva olukordi, kus autorid üldse arutlesid tulemuste üldistatavuse või selle piiride üle. Seda võib tõlgendada kui vähest teadlikkust valimi omaduste mõjust järelduste usaldusväärsusele, soovimatust tegeleda võimalike metoodiliste piirangutega või eeldust, et analüüsitulemused on võimalike esinduslikkust mõjutavate vigade suhtes piisavalt robustsed, aga selle kontrollimata jätmisel on põhimõtteliselt tegu põhjendamata eeldusega.

Neljas kriitiline aspekt puudutab osalus- ja vastamismäärade esitamist. Statistilise järeldamise eelduseks on mitte ainult see, et valim oleks võetud juhuslikkuse alusel, vaid ka see, et andmed oleksid piisavalt täielikud ja saavutatud valim säilitaks esinduslikkuse. Ka juhuslikkuse alusel võetud valim võib mittevastamise vea tõttu kaotada oma esinduslikkuse. Bryman (2012: 199-201) rõhutab, et vastamismäära esitamine on oluline selleks, et hinnata andmekogumi kvaliteeti ja uuringutulemuste üldistatavust, ilma selle teabeta ei ole võimalik isegi hinnanguliselt otsustada, kuivõrd uuringutulemused võivad peegeldada kogu sihtpopulatsiooni, mitte vaid neid, kes uuringus osalesid. Kuna mittevastamise viga on üks peamisi süsteemsete nihete allikaid küsitlusuuringutes, on selle kohta info mitteesitamine tõsine metoodiline möödalask – isegi kui andmekogum oli probleemideta, tuleks seda raportis näidata (AAPOR, 2023).

Lõpuks tuleb välja tuua, et kuigi paljud uuringud olid osaliselt metodoloogiliselt läbipaistvamad, näiteks andmete kogumise meetodite või valimi suuruse esitamise osas – valimi netosuurus oli esitatud 29 uurimisraportis 30-st, ei piisa üksikutest korrektselt täidetud aspektidest järeldava statistika rakendamise usaldusväärsuse tagamiseks. Järeldava analüüsi eelduste täitmine on süsteemne protsess, mis algab sihtpopulatsiooni määratlemisest ja lõpeb andmetöötluse läbipaistvusega. Kui mõni nendest lülidest puudub või on esitatud puudulikult, katkeb ahel, mille kaudu oleks võimalik teha õigustatud üldistusi või lugejal üldistuste õigsuses veenduda.

Analüüsitud uuringuaruannetes kvoote- ega paneeluuringute metoodikat ei kasutatud või ei tulnud metoodikakirjeldustest välja, et oleks kasutatud, mistõttu ei tõusnud see teema tulemuste analüüsis esile. Kuigi kvootide ja paneelide kasutamine on sotsiaaluuringutes levinud praktika, eriti mittetöenäosuslike valimite puhul, ei ole see antud valimi uuringute puhul rakendust leidnud.

Kokkuvõttes võib öelda, et järeltava analüüsi eelduste mittejärgimine või nende täitmise dokumenteerimata jätmine on HTM-i tellimusel koostatud rakenduslike uuringute puhul sagedane ja laiaulatuslik probleem. See seab küsitluse alla osades nendes töödes tehtud üldistuste kehtivuse ning viitab vajadusele kehtestada ühtsemad ja konkreetsemad standardid, mida peaksid uuringu läbiviijad valimi käsitlemisel ja tulemuste esitamisel järgima.

5.2 Järeltava analüüsi hinnangud

Statistiliste järeltulemuste esitamine ja tõlgendamine on rakendusuuringute keskne komponent, mitte ainult tehnilise analüüsi osa, vaid ka alus selleks, et leitud oleks mõju poliitikakujundamisele, sekkumiste disainile ja avaliku sektori otsustusprotsessidele. Käesolevas uurimistöös analüüsitud HTM-i tellimusel valminud uuringuaruannete põhjal tuleb siiski järeltulemuste esitamise kohta öelda, et statistiliste järeltulemuste käsitus on nendes töödes sageli metodoloogiliselt ebajärjekindel, sisuliselt pealiskaudne ja analüütilises mõttes ebapiisav. Seetõttu tuleb küsida, kui võrd võrreldavad uurimisraportite järeltulemused tegelikult usaldusväärsetele ja õigesti tõlgendatud statistilistele tulemustele.

Üldjoontes iseloomustab käesoleva töö valimis olevaid uuringuaruandeid ennekõike tulemuste pealiskaudne esitus, statistilisi analüüse kasutatakse, aga nende tulemuste sisuline tõlgendamine ei ole alati piisavalt põhjalik. Arvulised väärtused, näiteks p-väärtused, esitatakse mehaaniliselt ja üldistatud vormis, näiteks $p < 0.05$, ilma aruteluta selle üle, mida need sisuliselt tähendavad, kuidas need suhestuvad uurimisküsimustega ning milliseid järeltulemusi nende alusel teha saab. Üldistatud vormis esitus on omaette probleem, sest nagu rõhutavad Wasserstein jt (2019), tegelikult pole vahet $p = 0.049$ ja $p = 0.051$ väärtustel, samuti on suur vahe $p = 0.049$ väärtusel ja $p = 0.001$ väärtusel, kuigi mõlemad on $p < 0.05$. Samuti rõhutavad Greenland jt (2016), et isegi väga väike p-väärtus ei tähenda tingimata, et seos oleks sisuliselt oluline, kuna p-väärtus ei näita nähtuse olemasolu tõenäosust, vaid ainult seda, kui äärmuslik on leitud tulemus eeldusel, et nullhüpotees kehtib. Colquhoun (2017) lisab, et tüüpilise rakendusuuringu kontekstis võib isegi väike p-väärtus anda petlikult kindlustunde tulemuse paikapidavusest, sest sageli alahinnatakse uuringu tegelikku valepositiivse tulemuse riski. Statistika on raportites olemas, aga sageli ilma selge loogilise raamita – see tähendab, et järeltulemused ei ole analüütiliselt põhjendatud, vaid pigem oletuslikult seotud esitatud arvuliste tulemustega.

Kõige sagedasem probleem seondub p-väärtuste käsitlemisega. Enamikes töödes on p-väärtus ainukene väärtus, mille alusel tehakse statistiline järeltulemus seose või erinevuse olemasolu kohta. Tulemused peegeldavad üldist rahvusvahelist kriitikat statistiliste tulemuste väärti tõlgendamise

kohta. Rahvusvahelises teaduskirjanduses on järjepidevalt rõhutatud, et p-väärtuste liigne tähtsustamine ja sisuliste tõlgenduste puudumine suurendavad tulemuste väärtõlgendamise riski ning vähendavad uuringute usaldusväärsust (Greenland jt, 2016; Amrhein jt, 2019; Gelman & Stern, 2006). Kui seda väärtust kasutatakse ilma kontekstita või muud tüüpi toetava teabeta on oht, et tulemusi tõlgendatakse valesti, eeskätt nähes statistilist olulisust otseselt sisulise olulisusena. Selline eksitav esitusviis on analüüsitud aruannetes sagedane. Kuigi enamasti puudus võimalus hinnata, kas see viga viis konkreetsel juhul alusetute järeldusteni, on juba see probleem märkimisväärne, et aruannetes tugineti lihtsustatud otsustusviisile, kus statistiline olulisus samastati sisulise olulisusega.

Statistilise olulisuse mõiste on tänapäeval teadusmetoodika seisukohalt problemaatiline. Nagu rõhutavad Wasserstein jt (2019), on statistilise olulisuse kasutamine tihti viinud ekslike ja ülemääraste järeldusteni. Traditsiooniline p-väärtusel põhinev statistilise olulisuse käsitlus, näiteks tulemuste kategoriseerimine "oluliseks" või "ebaoluliseks" vastavalt sellele, kas $p < 0.05$ või mitte, on olemuselt meelevaldne ja soodustab põhjendamatu üldistusi, sealhulgas väidet, et statistiliselt mitteoluline tulemus tähendab mõju puudumist. Samuti Amrhein jt (2019) toovad välja, et tulemuste mehaaniline kategoriseerimine „statistiliselt oluliseks“ või „mitteoluliseks“ ainult p-väärtuse alusel võib viia tõsiselt eksitavate järeldusteni ning takistada sisulist analüüsi tulemuste praktilise olulisuse kohta. Kuigi statistilise olulisuse mõiste täielikku hülgamist on korduvalt soovitatud, kasutatakse seda praktikas jätkuvalt. Sellisel juhul on oluline, et autorid eristaksid statistilist olulisust ja sisulist olulisust selgelt. Kui sellist eristust ei tehta, näiteks kui statistiliselt olulise tulemuse olemasolu põhjal tehakse otseseid ja ulatuslikke järeldusi praktilise mõju kohta või statistiliselt mitteolulise tulemuse põhjal järeldatakse ekslikult seose või erinevuse puudumist, kujutab see endast olulist metoodilist puudujääki. See võib omakorda viia ebakorreksete sisuliste järeldusteni, millel võib olla tõsine mõju, eriti rakendusuuringute puhul, mille tulemusi kasutatakse poliitikakujunduses või praktilistes otsustusprotsessides.

Usaldusvahemike kasutamine, mis võiks pakkuda sisulisemat vaadet statistilistele hinnangutele oli väga haruldane. Mõnes üksikus töös esines usaldusvahemikke, ühes nendest esines problemaatiline juhtum, kus usaldusvahemikku tõlgendati ekslikult kui tõenäosusvahemikku. Selline tõlgendus ei vasta sagedusliku statistika lähtealustele ja viitab järeldava statistika tähendusliku kirjeldamise puudulikkusele. Hoekstra jt (2014) näitavad, et isegi kogenud uurijad kipuvad usaldusvahemikke tõlgendama valesti, näiteks käsitledes neid ekslikult tõenäosuslikena ehk arvatakse, et saaks leitud usaldusvahemiku õigsusele omistada usaldusnivoo väärtuse alusel mingi tõenäosuse, näiteks 95%, mis kajastus ka käesoleva töö tulemustes. Pernet (2016) rõhutab

omakorda, et kuigi usaldusvahemikud võivad pakkuda väärtuslikumat informatsiooni kui p-väärtused, ei realiseeru see potentsiaal, kui nende kasutamise ja tõlgendamise metoodilised alused jäävad ebaselgeks. Käesoleva valimi puhul osutus probleemiks pigem see, et usaldusvahemikke üldiselt ei esitatud, mistõttu oli nende tõlgendamise puudujääke piiratud võimalus hinnata.

Järjepidevus analüüsitulemuste ja järelduste vahel oli tihti puudulik. Analüüsi tulemustele ei viidatud järelduste sõnastuses piisava selgusega, mistõttu võis jääda lugejale ebaselgeks, kuidas esitatud järeldused olid seotud analüüsi tulemustega. Samuti esines juhtumeid, kus statistiliselt mitteoluliste tulemuste põhjal jõuti järeldusele, et seos või erinevus puudub, kuigi metoodiliselt tuleks sellisel juhul arutada ka võimalikke seletavaid faktoreid nagu madal statistiline võimsus. Lisaks esines juhtumeid, kus statistiliselt mitteolulised tulemused jäeti üldse raporteerimata, keskendudes aruannetes üksnes statistiliselt olulistele leidudele. Selline selektiivne esitus võib jätta lugejale eksliku mulje, et uuritud nähtustes valitsesid valdavalt statistiliselt olulised erinevused või seosed. Tegemist on probleemse praktikaga, mis võib põhjustada järelduste kallutatust. Statistiliselt mitteoluliste tulemuste raporteerimine, isegi juhul kui nendega ei jõuta kindlate järeldusteni, suurendaks aruannete läbipaistvust ja võimaldaks lugejal teha iseseisvalt tasakaalustatud hinnanguid tulemuste kohta.

Üheks korduvaks probleemiks osutus seose tugevuse ja praktilise olulisuse käsitlemise puudumine. Ehkki statistilise olulisuse rõhutamine on valdav, ei toimunud valimi raportites piisavalt arutelu selle üle, kas leitud seosed või erinevused on ka tähenduslikud uurimiskonteksti seisukohalt. Isegi väga madal p-väärtus võib tuleneda suurest valimist ja peegeldada minimaalseid erinevusi, millel ei pruugi olla mingit praktilist tähendust. Sellistel juhtudel on eriti tähtis tuua välja kasvõi indikatiivsed hinnangud seose tugevusele. Wasserstein ja Lazar (2016) toovad välja, et statistiline olulisus ei ole samaväärne teadusliku olulisusega, väiksemad p-väärtused ei pruugi viidata tugevamale või tähtsamale seosele ning suuremad p-väärtused ei pruugi viidata tähtsuse puudumisele või isegi seose puudumisele – iga seos, ükskõik kui tühine, võib anda väikese p-väärtuse.

Mudelipõhise analüüsi korral, näiteks regressioonimudelite kasutamisel, jäi paljudes töödes selgusetuks, millised kaalutlused juhtisid muutujate kaasamist mudelisse ning milline oli nende sisuline roll. See vastab üldisele rahvusvahelises kirjanduse kirjeldatud probleemile, näiteks Greenland jt (2016) kirjutavad, et eelduste mõistmise ja hindamise raskust suurendab asjaolu, et statistilise mudeliga seotud informatsiooni esitatakse tavaliselt äärmiselt kokkusurutult ja tihedalt või isegi abstraktselt, kui üldse esitatakse. Ka tulemuste kvantitatiivne esitus oli piiratud,

regressioonikordajaid küll esitati, kuid nende sisuline tähendus jäi vahel lahtiseks, mis muudab tulemused mitteeksperdist lugejale raskesti mõistetavaks.

Kokkuvõttes on statistiliste järelduste esitamine analüüsitud rakendusuuringutes küll vormiliselt olemas, kuid sisuliselt sageli ebapiisav. Kuigi järeldavat statistikat kasutatakse, piirdub selle rakendamine enamasti p-väärtuste esitamisega, ulatumata seose tugevuse põhjaliku tõlgendamiseni. Tähenduslikke järeldusi tehakse, aga nende tegemise loogika ei vasta sellele, mida kasutatud järeldava statistika tööriistad, näiteks p-väärtus, võimaldavad. Statistiline analüüs näib mõnikord olevat rakendatud mehaaniliselt, ilma et oleks põhjalikult läbi mõeldud, milliseid hüpoteese tegelikult testitakse, mida p-väärtus antud kontekstis lubab järeldada ning kuidas see sobitub uuringu disainiga. Näiteks jääb tihti avamata, mida nullhüpoteesi tagasilükkamine sisuliselt tähendab ja kas see võimaldab teha mõtestatud järeldusi uuritava nähtuse kohta. Statistilised järeldused on esitatud viisil, mis ei ava nende metoodilist alust ega võimalda hinnata tulemuste kehtivust ja usaldusväärsust, mistõttu jääb lugejale, olgu ta poliitikakujundaja või laiema avalikkuse esindaja, ebaselgeks, mida tulemused tegelikult tähendavad.

Sellised metoodilised puudused ei ole ainult akadeemiline probleem, vaid neil võivad olla otsesed negatiivsed mõjud poliitikakujundamisele ja avaliku sektori otsustusprotsessidele. Ioannidis (2005) toob välja, et suur osa avaldatud uuringutulemustest võib olla tegelikult väär just seetõttu, et otsuseid tehakse sageli üksnes statistilise olulisuse põhjal. Bernardi jt (2017) märgivad, et selline metoodiline pinnapealsus ei võimalda otsustajatel adekvaatselt mõista uurimistulemuste reaalselt tähendust, vähendades uuringute praktilist väärtust ja piirates nende mõju sotsiaalsetele muutustele või poliitikate kujundamisele. Seega tuleb rakendusuuringutes pöörata senisest enam tähelepanu tulemuste metoodiliselt korrektsele esitusele ja sisulisele tõlgendamisele, mis oleks mõistetav mitte ainult ekspertidele, vaid ka laiema avalikkuse ja poliitikakujundajate jaoks.

Nullhüpoteesi testimise (NHST) metoodika puudused on sotsiaalteadustes pälvinud tähelepanu juba aastakümneid. Hoolimata selle laialdasest levikust, on NHST-d kritiseeritud nii kontseptuaalsete kui ka praktiliste probleemide tõttu, mis mõjutavad otseselt järelduste usaldusväärsust. NHST-d on näiteks peetud ebasobivaks vastamaks küsimustele, mis uurijaid tegelikult huvitavad, kritiseeritud selle kalduvuse tõttu ülehinnata tõendeid nullhüpoteesi vastu ning osutatud raskustele selle tulemuste õigetel tõlgendamistel, mis on põhjustanud sageli ekslikke järeldusi (Hoekstra jt, 2014).

P-väärtustel põhinev statistiline analüüs on hoolimata laialdasest kriitikast säilinud teaduspraktikas, kuna puudub lihtsam ja laialdaselt aktsepteeritud alternatiiv selle täielikuks

asendamiseks. Samas pole ka eriti realistlik oodata, et seda suudetaks edaspidi ühtviisi ja järjepidevalt korrektselt rakendada, arvestades, et hüpoteeside testimise probleemidest on räägitud juba ligi sada aastat. Viimastel aastatel on küll tõstatatud uusi algatusi ja üleskutseid, näiteks statistilise olulisuse kontseptsiooni hülgamine (Amrhein jt, 2019), kuid nende edukus on kaheldav. Järgmises sektsioonis on esitatud mõned praktilised soovitusel, mis ehk küll pole uued ega radikaalsed, kuid võivad siiski toimida kui väikesed arengukohad kehtivas, kohalikus teaduspraktikas.

5.3 Praktilised soovitusel

Järelduste usaldusväärsuse ja metoodilise selguse suurendamiseks on oluline, et raportites oleks selgelt välja toodud ka järelduste tegemise eeldused ja nende kontrollimise protseduurid. Käesoleva uurimistöo kodeerimistulemused näitasid, et analüüsitud rakendusuuringute raportites jäävad järelduste kehtivust puudutavad eeldused sageli kajastamata või ebamääraselt esitatuks. Kui mõelda sellele, miks statistiliste järelduste eelduste esitamine uuringutes tihti puudulikuks jääb, siis tundub, et asi ei pruugi alati olla uurijate teadmiste või oskuste puudumises. Pigem võib põhjuseks olla levinud arusaam, et eelduste kontrollimine ja nende põhjalik dokumenteerimine on midagi, mis kuulub akadeemiliste uurimuste juurde ning on rakendusuuringute kontekstis pigem tüütu formaalsus kui hädavajalik sisuline tegevus. See hoiak võib tuleneda ka uuringute tellijate survest või ootustest. Seega oleks vajalik raportite koostamise kultuuri muutus, kus selgelt dokumenteeritud eelduste kontroll ei ole enam lihtsalt „hea tava”, vaid elementaarne osa uuringu läbiviimisest ja aruande koostamisest.

Käesoleva uurimistöo tulemused viitavad selgelt metoodilistele nõrkustele, mis korduvad rakendusuuringute aruannetes ja vajavad süsteemset tähelepanu ning parandamist. Kõige sagedasemaks puuduseks on statistiliste tulemuste esitamise ja tõlgendamise pinnapealsus, tulemusi esitatakse sageli pelgalt p-väärtuse alusel, ilma et käsitletaks piisavalt tulemuste praktilist või sisulist olulisust, seose tugevust või piiranguid järelduste üldistatavusele.

Korrektsead teaduslikud järeldused nõuavad analüüsiprotsessi ja tulemuste täielikku läbipaistvust. P-väärtuste ja muude statistiliste näitajate valikuline esitamine moonutab tulemuste tõlgendamist ning võib viia eksitavate järeldusteni. Kui andmeid analüüsitakse mitmel viisil, ent raporteeritakse vaid need tulemused, mis saavutavad soovitud olulisuse taseme, muutuvad avaldatud p-väärtused sisuliselt ebausaldusväärseks. Selline praktika, mida tuntakse teisisõnu andmete kaevandamise või p-häkkimise nimede all, suurendab teaduskirjanduses näiliselt oluliste tulemuste hulka. Probleem

ei piirdu ainult formaalsete korduvate testimistega – piisab ka sellest, kui uurija valib tulemused välja nende statistiliste omaduste põhjal, ilma et lugejale oleks selgitatud valiku tegemise aluseid. Seepärast on oluline, et uuringutes oleks kirjas kõik testitud hüpoteesid, tehtud andmekogumise ja analüüsiotsused ning arvutatud p-väärtused. Ilma selle teabeta on p-väärtustel põhinevad järeldused sisuliselt usaldamatud (Wasserstein ja Lazar, 2016).

Selliste metoodiliste puuduste vältimiseks on hädavajalik muuta uuringuraportite koostamise praktikaid. Üheks praktiliseks soovitusena on metoodikakäsitluste standardiseerimine – kui mingi ministerium tellib väliselt asutuselt töö, siis nendele töödele peaks olema rakendatud ühised, selged nõuded raportis informatsiooni esitamise kohta nii järeltava analüüsi eelduste kui ka statistilise järelduste puhul. Selliste juhiste järgimine suurendab läbipaistvust ja metoodilist selgust ning vähendab väärtõlgenduste riski. Samas tuleb arvesse võtta, et antud rakendusuuringuviivad läbi mitmed erinevad asutused, kellel ilmselt on oma sisemised standardid. Minu ettepanek oleks koostada rakendusuuringute teostajatele juhendavad põhimõtted, mis loetlevad kõik olulised aspektid, mida uuringuaruandes tuleks käsitleda. Eesmärk ei ole kehtestada jätku reegleid, vaid pakkuda läbipaistvust ja kvaliteeti soodustavat raamistikku, mille järgimine muudab aruanded sisukamaks ja paremini hinnatavaks. Käesolevas töös ilmnes mitmeid juhtumeid, kus raporti põhjal polnud võimalik otsustada, kas mõnda protseduuri, näiteks kaalude kasutamist, oli üldse rakendatud. Sellistes olukordades piisaks ka lühikesest märkusest, näiteks „kaale ei kasutatud, sest...“, et vältida ebaselgust. Analüüsitud valim jättis sageli mulje, et kui midagi ei tehtud, siis seda ei mainitudki, mis omakorda raskendab usaldusvääruse ja metoodilise põhjalikkuse hindamist.

Teiseks praktiliseks soovitusena on kutsuda töid läbiviivad asutused ja uurijad üles olema põhjalikumad nii uuringute kavandamise kui ka raportite koostamise faasis. Käesolevas töös analüüsitud uuringuraportid olid käsitletud aspektide dokumenteerimise põhjalikkuses üldiselt ebaühtlased; sarnast taset tööprotsessi kirjeldamises oli märgata peamiselt sama asutuse koostatud tööde puhul. Seetõttu oleks soovitatav, et uuringuid läbiviivad asutused ja uurijad pööraksid süsteemselt tähelepanu sellele, kuidas järeltava analüüsi eeldusi ning statistilisi järeldusi dokumenteeritakse. Soovitatav on võtta aluseks sarnased kriteeriumid nagu käesolevas töös kasutatud hindamissüsteemis: kas on esitatud info valimi kujunemise kohta, kas on põhjendatud eelduste täitmine ja kas statistilised järeldused on esitatud üheselt ja arusaadavalt. Arusaadavalt ei pruugi kõiki detaile olla võimalik alati esitada, kuid just eelduste ja järelduste läbipaistvam käsitlemine oleks enamasti teostatav vähese lisapingutusega ning aitaks vältida tõlgendusprobleeme.

5.4 Töö piirangud

Käesoleva uurimistöö üheks põhiliseks piiranguks on kasutatud kontentanalüüsi meetod, mis põhineb üksnes raportite tekstide analüüsil. Selline lähenemine seab piirangud järelduste tegemisele, kuna tegelikke analüüsi ja algseid andmestikke ei olnud võimalik uuesti analüüsida ega kontrollida. Sellest tulenevalt sõltuvad käesoleva töö tulemused suuresti sellest, kuidas uuringu koostajad oma metoodikaid, analüüsi ja järeldusi raportites kirjeldasid, tegelik analüüs võib olla olnud korrektsem või vähem korrektne, kui aruannete tekstidest nähtub.

Lisaks piirab uuringu üldistatavust ka suhteliselt väike ja spetsiifiline valim. Käesolevas uuringus analüüsitud 30 raportit olid kõik tellitud Eesti Haridus- ja Teadusministeeriumi poolt, mis seab piirangud tulemuste laiendamisele teistele rakendusuringutele või teiste valdkondade uuringupraktikatele – on võimalik, et erinevate ministeeriumite poolt tellitud uuringud võivad oma metoodilises kvaliteedis erineda, mis tähendab, et käesolevas töös tuvastatud probleemid ei pruugi olla universaalsed. Samas viitavad varasemad rahvusvahelised uuringud, näiteks Hoekstra jt (2014), et järeldava analüüsi kirjeldamise probleemid on levinud. Kuigi erinevused antud probleemide ulatuses võivad esineda, on üldine tendents näidanud, et statistiliste järelduste tõlgendamise kitsaskohad on pigem endeemilised.

Tuleb arvestada ka hindamisprotsessi subjektiivsusega, mis on omane kvalitatiivsele kontentanalüüsi meetodile. Kuigi töö käigus koostati ja kasutati struktureeritud kodeerimisjuhendit, mille eesmärk oli tagada hinnangute võimalikult suur võrreldavus ja järjepidevus, sisaldab iga kontentanalüütiline hindamine paratamatult tõlgendusruumi. Näiteks võib ühe lugeja hinnang sellele, kas statistiline järeldus on „selgelt“ sõnastatud või „piisavalt“ põhjendatud erineda teisest lugejast, sõltudes kontekstist ja isiklikust arusaamast statistilise selguse või adekvaatsuse kriteeriumitest. Eriti keeruline on hinnata selliseid juhtumeid, kus mõisted või terminid on kasutusel, aga nende tähendus jääb ebamääraseks või on avatud mitmeti mõistmisele. Käesoleva uurimistöö metoodika puhul jääb võimalus, et teised uurijad, kasutades sama juhendit, oleksid üksikjuhtumite puhul jõudnud veidi erinevate hinnanguteni. See piirang ei vähenda töö üldist väärtust, kuid rõhutab vajadust käsitleda tulemusi kvalitatiivse tõendusmaterjalina, mis vajab kontekstipõhist ja kriitilist tõlgendamist.

5.5 Edasised uurimissuunad

Käesoleva uurimistöö tulemused loovad mitmeid perspektiive ja võimalusi edasisteks uuringuteks. Üheks oluliseks edasiseks uurimissuunaks oleks analüüsitud uuringuaruannete metoodilise kvaliteedi küsimuste laiendamine teiste Eesti ministeeriumite ja avaliku sektori asutuste tellitud rakendusuuringutele. Selline lähenemine võimaldaks hinnata, kas käesolevas töös tuvastatud metoodilised puudused, nagu statistiliste järelduste pinnapealne esitamine, p-väärtuste mehaaniline tõlgendamine ja seose tugevuse käsitlemata jätmine, on spetsiifilised ainult Haridus- ja Teadusministeeriumi uuringutele või kujutavad endast laiemat üldist tendentsi Eesti avaliku sektori tellitud uuringutes. Sellised võrdlused võimaldaksid selgemalt mõista institutsionaalsete tellimistingimuste, uuringu koostajate valiku või ministeeriumisiseste uurimistavade mõju uuringute metoodilisele kvaliteedile.

Lisaks Eesti-sisesele võrdlusele oleks väga informatiivne teha võrdlevaid uuringuid rahvusvaheliste rakendusuuringute raportitega. Võrreldes Eesti rakendusuuringute aruandeid näiteks Euroopa Liidu või Põhjamaade avaliku sektori poolt tellitud uuringute raportitega saaks hinnata, kas Eestis tuvastatud metoodilised probleemid erinevad oluliselt rahvusvahelistest. See võimaldaks paremini mõista Eesti uuringupraktika tugevusi ja nõrkusi rahvusvahelises perspektiivis ning pakkuda konkreetsemaid ja täpsemaid soovitusi selleks, kuidas parandada just Eesti uuringute kvaliteeti. Selline analüüs võiks hõlmata ka valdkondlikke võrdlusi, näiteks hariduse, sotsiaalpoliitika või tervisevaldkonna uuringute vahel, et tuvastada, millised valdkonnad on metoodilistes küsimustes tugevamad või nõrgemad ning miks see nii võiks olla.

Oluline edasine uurimissuund oleks ka süvitsi minev kvalitatiivne uurimus, mis keskenduks uuringuaruannete koostajate ja tellijate kogemustele ning seisukohtadele. Kvalitatiivsed intervjuud raportite koostajate ja tellijatega võimaldaksid välja selgitada, miks kirjeldatud metoodikaprobleemid tekivad, millised on nende tekkimise peamised põhjused ning millised institutsionaalsed ja praktilised takistused võivad mõjutada raportite metoodilist kvaliteeti. Selline kvalitatiivne lähenemine aitaks avada ka laiemaid küsimusi, näiteks seda, kuidas uuringu koostajad mõistavad metoodilise läbipaistvuse ja täpsuse nõudeid, milliseid juhiseid ja standardeid nad kasutavad või miks neid ei kasutata ning kuidas mõjutab nende tööd rahastusmodel, tähtsajad või muud praktilised piirangud. Lisaks võimaldaks selline kvalitatiivne analüüs paremini mõista uuringute tellijate ja tegijate vahelisi suhteid ja kommunikatsiooni, mis võivad oluliselt mõjutada uuringute lõppkvaliteeti.

KOKKUVÕTE

Käesoleva magistritöö eesmärk oli hinnata valitud Haridus- ja Teadusministeeriumi tellitud rakendusuuringuid, milles on kasutatud järeltavat analüüsi, metoodilist kvaliteeti, keskendudes sellele, kuivõrd on nendes täidetud järeltava statistika kasutamiseks vajalikud eeldused ning kas statistilised järeldused on esitatud selgelt ja sisukalt, võimaldades teha usaldusväärseid ja põhjendatud üldistusi. Ministeeriumite tellitud rakendusuuringutel on oluline roll avaliku poliitika ja ühiskondlike otsuste kujundamisel, mistõttu on uurimistulemuste usaldusväärsus ja üldistatavus teaduspõhise poliitikakujundamise seisukohalt kriitilise tähtsusega.

Järeldava analüüsi usaldusväärsus eeldab valimi esinduslikkust, tõenäosuslikku valikut, sihtpopulatsiooni kaetust ning statistiliste näitajate, nagu p-väärtuste ja usaldusvahemike, korrektset kasutamist ja tõlgendamist. Töö teoreetiline raamistik lähtuski neid aspekte puudutavast kontseptuaalsetest alustest. Magistritöö tugineb autori varasemale bakalaureusetööle, mis tõi esile metodoloogilisi puudusi HTM-i tellitud uuringute küsitlusraportites, käesolev töö laiendab seda käsitlust, hinnates süstemaatilisemalt järeltava analüüsi eeldusi ja statistiliste järelduste esitusviise.

Kontentanalüüsi meetodil analüüsiti 30 valimipõhist rakendusuuringut. Uurimuse läbiviimiseks koostati struktureeritud kodeerimisjuhend, mis võimaldas hinnata uuringuraportite metoodilist läbipaistvust ja analüütilist kvaliteeti tervikuna, kaasates nii tekstiosa kui lisad, tabelid ja joonised.

Tulemused näitavad, et osades uuringutes esinevad märkimisväärsed puudused valimi koostamise kirjelduse, valimiraami, osalusmäärade ning valikumeetodite selguse osas. Samuti kasutatakse mitmel juhul mittetõenäosuslikke valimeid ilma piisava kriitilise refleksioonita, mis seab kahtluse alla järelduste üldistatavuse. Statistiliste tulemuste esituses esines probleeme nii konteksti puudumise kui ka ebapiisavate näitajatega, näiteks usaldusvahemike esitamise puudumise ja p-väärtuste liigse rõhutamisega.

Antud uurimistöö pakub esmakordselt süsteemse ülevaate järeltava statistika rakendamisest ja probleemidest valitud aastatel HTM-i poolt tellitud rakendusuuringutes, kus kasutati järeltavat analüüsi. Selle töö metoodiliselt rangelt struktureeritud hinnang teeb nähtavaks levinud puudujäägid ning loob aluse paranduste planeerimiseks. Selles mõttes sarnaneb töö väärtus rahvusvaheliste püüdlustega tõsta uuringute läbiviijate ja tulemuste tarbijate teadlikkust statistiliste järelduste õigest tõlgendamisest.

SUMMARY

The aim of this master's thesis was to assess the methodological and analytical quality of applied research commissioned by the Estonian Ministry of Education and Research that employed inferential analysis, focusing on the extent to which the assumptions required for inferential statistics were met and whether the statistical conclusions were formulated clearly and meaningfully, enabling reliable and well-founded generalisations. Applied research commissioned by ministries plays a key role in shaping public policy and societal decision-making, making the reliability and generalisability of research results critical for evidence-based policymaking.

The reliability of inferential analysis requires sample representativeness, probability-based selection, population coverage, and the correct use and interpretation of statistical indicators such as p-values and confidence intervals. The theoretical framework of this thesis was based on these conceptual foundations. The thesis builds on the author's earlier bachelor's thesis, which identified methodological shortcomings in survey reports commissioned by the Ministry of Education and Research, with the present work expanding this perspective by systematically assessing the assumptions underlying inferential analysis and the presentation of statistical conclusions.

Thirty sample-based applied studies were analysed using the method of content analysis. A structured coding manual was developed for the study, enabling the assessment of the methodological transparency and analytical quality of research reports, covering the main text, appendices, tables, and figures.

The results show that a substantial proportion of the studies exhibited significant shortcomings regarding the description of sample formation, the sample frame, response rates, and the clarity of selection methods. In several cases, non-probability samples were used without sufficient critical reflection, calling into question the generalisability of the conclusions. Problems were also observed in the presentation of statistical results, including the lack of context, the absence of key indicators such as confidence intervals, and the excessive emphasis on p-values.

This research provides the first systematic overview of the application of inferential statistics and related issues in applied research studies commissioned by the Ministry of Education and Research during the selected years, where inferential analysis was used. The methodologically rigorous assessment highlights common shortcomings and provides a foundation for planning improvements. In this respect, the value of the work aligns with international efforts to enhance the awareness of researchers and users of research findings regarding the correct interpretation of statistical conclusions.

KASUTATUD KIRJANDUS

- Allison, P. (2004). Missing Data. Lewis-Beck, M., Bryman, A. E., & Liao, T. F. (toim.), *The Sage encyclopedia of social science research methods*. (lk 645-649). Sage publications.
- American Psychological Association. (2024). *Quantitative research design (JARS-Quant)*.
<https://apastyle.apa.org/jars/quantitative>
- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305-307. <https://doi.org/10.1038/d41586-019-00857-9>
- Bernardi, F., Chakhaia, L., & Leopold, L. (2017). ‘Sing me a song with social significance’: the (mis) use of statistical significance testing in European sociological research. *European Sociological Review*, 33(1), 1-15. <https://doi.org/10.1093/esr/jcw047>
- Bryman, A. (2012). *Social Research Methods 4th Edition*. Oxford: Oxford University Press
- Callegaro, M., Lozar Manfreda, K., Vehovar, V. (2015) *Web Survey Methodology*. [E-raamat]. London: Sage. doi:10.4135/9781529799651
- Clark, T., Foster, L., Sloan, L., & Bryman, A. (2021). *Bryman's social research methods* (6th ed.). Oxford University Press.
- Colquhoun, D. (2017). The reproducibility of research and the misinterpretation of p-values. *Royal society open science*, 4(12), 171085. <https://doi.org/10.1098/rsos.171085>
- De Leeuw, E. D., Hox, J., ja Dillman, D. (2008). *International handbook of survey methodology*. New York: Routledge
- Gallagher, P. (2004). Nonresponse. Lewis-Beck, M., Bryman, A. E., & Liao, T. F. (toim.), *The Sage encyclopedia of social science research methods*. (lk 740). Sage publications.
- Gelman, A., & Stern, H. (2006). The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *The American Statistician*, 60(4), 328–331. <https://doi.org/10.1198/000313006X152649>
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*, 31(4), 337-350. <https://doi.org/10.1007/s10654-016-0149-3>
- Groves, R.M., Fowler Jr, F.J., Couper, M.P., Lepkowski, J.M, Singer, E., Tourangeau, R. (2009). *Survey Methodology. Second Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc.

- Haridus- ja Teadusministeerium. (2025). *Uuringud ja statistika*.
<https://www.hm.ee/ministeerium-uudised-ja-kontakt/uuringud-ja-statistika/uuringud>
- Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic bulletin & review*, 21, 1157-1164.
<https://doi.org/10.3758/s13423-013-0572-3>
- Hu, Q., Zhao, S., Yang, J. (2010). Confidence intervals. Salkind, N. J. (toim.), *Encyclopedia of research design* (lk 211-215). Sage publications.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.
<https://doi.org/10.1371/journal.pmed.1004085>
- Lagerspetz, M. (2017). *Ühiskonna uurimise meetodid. Sissejuhatus ja väljajuhatus*. Tallinn: Tallinna Ülikooli Kirjastus
- Lang, T. A., & Altman, D. G. (2015). Basic statistical reporting for articles published in biomedical journals: the “Statistical Analyses and Methods in the Published Literature” or the SAMPL Guidelines. *International journal of nursing studies*, 52(1), 5-9. <https://www.equator-network.org/reporting-guidelines/sampl/>
- Lavrakas, P. J. (toim.) (2008). *Encyclopedia of Survey Research Methods*. Thousand Oaks, California: Sage Publications, Inc. <https://doi.org/10.4135/9781412963947>
- Lewis-Beck, M. (2004). Deletion. Lewis-Beck, M., Bryman, A. E., & Liao, T. F. (toim.), *The Sage encyclopedia of social science research methods*. (lk 244). Sage publications.
- Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data*. John Wiley & Sons.
<https://doi.org/10.1002/9781119482260>
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Chapman and Hall/CRC.
- Neuman, L.W. (2011). *Social research methods. Qualitative and quantitative approaches*. Boston: Pearson
- Pernet, C. (2017). Null hypothesis significance testing: a short tutorial. *F1000Research*, 4, 621.
<https://doi.org/10.12688/f1000research.6963.3>
- Põldoja, K. (2023). *Aastatel 2015-2022 Haridus-ja Teadusministeeriumi tellimusel läbiviidud küsitlusuuringute aruannete metoodikakirjelduste analüüs*. Bakalaureusetöö. Tartu Ülikool, infoühiskonna ja sotsiaalse heaolu õppekava.

- Rässler, S. (2004). Imputation. Lewis-Beck, M., Bryman, A. E., & Liao, T. F. (toim.), *The Sage encyclopedia of social science research methods*. (lk 477-480). Sage publications.
- Rotello, C. M., Myers, J. L., Well, A. D., & Lorch Jr, R. F. (2025). *Research Design and Statistical Analysis*. Routledge.
- The American Association for Public Opinion Research (AAPOR). (2023). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys 10th edition*.
<https://aapor.org/standards-and-ethics/standard-definitions/>
- Tiit, E-M. ja Tooding, L-M. (2019). *Statistikaleksikon*. Tartu: Tartu Ülikooli Kirjastus
- Toepoel, V. (2016). Doing Surveys Online. *Incentives for Respondents*. London: SAGE.
<https://doi.org/10.4135/9781473967243>
- Traugott, M. (2004). Representative sample. Lewis-Beck, M., Bryman, A. E., & Liao, T. F. (toim.), *The Sage encyclopedia of social science research methods*. (lk 967). Sage publications.
- Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). Sage publications.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133.
<https://doi.org/10.1080/00031305.2016.1154108>
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$.” *The American Statistician*, 73(sup1), 1-19. <https://doi.org/10.1080/00031305.2019.1583913>
- Wu, C. (2022). Statistical inference with non-probability survey samples. *Survey Methodology*, 48, 283-311.

LISAD

Lisa 1. Valim

Analüüsiüksuse number	Pealkiri	Läbiviija(d)	Link raportile	Aasta
A1	Kaasava hariduskorralduse tõhusus üldhariduses 2023	Tartu Ülikool - Haridusteaduste Instituut	Link	2024
A2	Õppepuhkuste kasutatavuse uuring	RAKE	Link	2024
A3	IEA 2022. aasta rahvusvahelise kodanikuhariduse uuringu Eesti tulemused	Tallinna Ülikool	Link	2023
A4	Eesti noortevaldkonna töötajaskonna töötingimused	MTÜ Balti Uuringute Instituut	Link	2023
A5	Kiusamist ennetavate ja vähendavate sekkumiste tõhusus ning haridusasutuste kogemused nende rakendamisel	SA Poliitikauuringute Keskus Praxis	Link	2022
A6	Eriolukorrast tingitud distantsõppe kogemused ja mõju Eesti üldharidussüsteemile	Tallinna Ülikool	Link	2021
A7	Eriolukorrast tingitud distantsõppe kogemused ja mõju kutseharidusele	Tallinna Ülikool	Link	2021
A8	Noorsootöös osalevate noorte rahulolu noorsootöoga 2020	MTÜ Balti Uuringute Instituut, HeiVäl OÜ	Link	2021
A9	Ettevõtlusõppe rakendamine üld-, kutse- ja kõrghariduses	Eesti Ettevõtluskõrgkool Mainor, Tallinna Ülikool	Link	2020

A10	Akadeemilised töötajad teadmusühiskonnas	Tartu Ülikool, SA Poliitikauuringute Keskus Praxis	Link	2019
A11	Organisatsioonipoolsete juhendajate rahulolu kutsehariduse praktikakorraldusega	Rakendusauuringute Keskus CentAR	Link	2019
A12	Elukestva õppe strateegia vahehindamine	SA Poliitikauuringute Keskus, Praxis Rakendusauuringute Keskus CentAR	Link	2019
A13	Noorsootöös osalevate noorte rahulolu noorsootöoga	Tartu Ülikool	Link	2018
A14	Töökohapõhise õppe lõpetajate uuring 2018	Kantar Emor	Link	2018
A15	Kutsehariduse maine üldhariduskoolide õpilaste, õpetajate ning lapsevanemate seas	Accaro Solutions OÜ	Link	2018
A16	Eesti kõrgkoolide 2015. aasta vilistlaste uuring	Ernst & Young Baltic AS	Link	2017
A17	Keelehoiakute uuring	Turu-uuringute AS	Link	2017
A18	Eesti kutseõppe 2015.-2016. aasta vilistlaste uuring	Ernst & Young Baltic AS	Link	2017
A19	Õpetajate töötaja kasutuse ja töö lisatasustamise praktikate uuring	Ernst & Young Baltic AS	Link	2016
A20	Muutunud õpikäsituse rakendamise ja selleks erinevate valikute pakkumise seire korraldamine, sh gümnaasiumis valikkursuste rakendamise osas	Tartu Ülikool - Haridusuuenduskeskus	Link	2016

A21	Teema "Keskond ja jätkusuutlik areng" ja teiste läbivate teemade rakendamine üldhariduses	Tartu Ülikool	Link	2016
A22	Eesti põhikooliõpilaste koduse õpikoormuse uuring 2013-2015	Tallinna Ülikool	Link	2015
A23	Vajaduspõhise õppetoetuse uuring	SA Poliitikauuringute Keskus Praxis	Link	2015
A24	Programmi TULE uuringu „Haridustee valikud ning õpingute katkestamise asjaolud Eesti kõrghariduses“ aruanne	Tartu Ülikool	Link	2015
A25	Projekt Üldpädevused ja nende hindamine "Ülevaade projektist ja mõningaid 7.-9. klasside uurimuse tulemusi"	Tallinna Ülikool	Link	2014
A26	Doktoriõppe tulemuslikkuse analüüs	Tartu Ülikool	Link	2014
A27	Eesti kõrgkoolide 2012. aasta vilistlaste uuring	OÜ Cumulus Consulting	Link	2014
A28	Muukeelne laps Eesti koolis: Jätkuuuring	Tallinna Ülikool	Link	2014
A29	Uusimmigrantõpilaste akadeemiline ja sotsiaalne toimetulek Eesti üldhariduskoolis	MindPark	Link	2013
A30	Kutseharidus ja muutuv tööturg	SA Poliitikauuringute Keskus Praxis	Link	2013

Lisa 2. Kodeerimisjuhend

1: Analüüsiüksuse (küsitlusuuringu raporti) number

AI-A99

2: Link küsitlusuuringu aruande failile

Link

3: Pealkiri

Sõnaline tunnus

4: Ilmumise aasta

Neljakohaline number vahemikus 2015–2022

5: Uuringu läbiviija nimetus

Asutuse nimetus

6: Kas on selgelt sõnastatud, kas valim on võetud kasutades tõenäosuspõhiseid või mittetõenäosuslikke meetodeid?

6a Vastuse aluseks olev osa raportist

6b Hinnang – 1 Jah Selgelt välja toodud, missugust valikumeetodit kasutati ja tõenäosuslikkus; 2 Ei Valikumeetodit ei ole kirjeldatud; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

6c Selgitus

7: Kas on välja toodud valimimahud, sh valimiraamiti?

7a Vastuse aluseks olev osa raportist

7b Hinnang – 1 Jah Täielikult netovalimi maht välja toodud; 2 Ei Netovalimi mahtu ei ole kirjeldatud; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

7c Selgitus

8: Kas on kirjeldatud valimiraam või sellele vastav andmekogum?

8a Vastuse aluseks olev osa raportist

8b Hinnang – 1 Jah Täielikult kirjeldatud valikuraami olemust, vajadusel selle väljastajat ja selgitatut populatsiooni kaetust (sh millised sihtpopulatsiooni segmendid pole kaetud); 2 Ei Valimiraami või sellele vastavat andmekogumikku pole täielikult kirjeldatud; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

8c Selgitus

9: Kas on välja toodud planeeritud ja saavutatud valimite mahud, nende eripärade võrdlus ja erinevuste põhjused ning kuidas nendega seotud probleemidega tegeleti?

9a Vastuse aluseks olev osa raportist

9b Hinnang – 1 Jah Olemas planeeritud ja saavutatud valimite mahud ja eripärade võrdlus ning probleemidega tegelemine (kui kohane); 2 Ei ole olemas planeeritud ega saavutatud valimi mahte või on olemas ainult saavutatud valimite mahud; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

9c Selgitus

10: Kas on selgelt kirjeldatud meetodid, millega osalejatega kontakteeruti, millega nad välja valiti või värvati, sh kirjeldatud kõik valikukriteeriumid ja võimalik üle-esindatus valimis?

10a Vastuse aluseks olev osa raportist

10b Hinnang – 1 Jah Selgelt kirjeldatud meetodid, millega osalejatega kontakteeruti, millega nad välja valiti või värvati, valikukriteeriumid ja võimalik üle-esindatus valimis; 2 Ei Igasugune värbamismeetodite kirjeldus puudub või on ebaselge; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

10c Selgitus

11: Kas on täielikult ja täpselt määratletud populatsioon, sh asukoha, vanuse, muude sotsiaalsete ja demograafiliste omaduste, aja lõikes (kui see on uuringu spetsiifikat arvestades relevantne?)

11a Vastuse aluseks olev osa raportist

11b Hinnang – 1 Jah täielik Selgelt ja korrektselt määratletud sihtpopulatsioon välja toodud; 2 Ei Sihtpopulatsiooni määratluse kohta informatsiooni ei leidu või see on ebaadekvaatne; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

11c Selgitus

12: Kas on välja toodud kuivõrd hästi valim sihtpopulatsiooni esindab ja selle järeldused?

12a Vastuse aluseks olev osa raportist

12b Hinnang – 1 Jah Kuivõrd hästi valim sihtpopulatsiooni esindab on kirjeldatud ning selle järeldused on selged; 2 Ei Esinduslikkust või selle tähendust analüüsitulemuste jaoks pole kirjeldatud või on, arvestades esinduslikkusega seotud probleeme, kirjeldatud ebapiisavalt; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

12c Selgitus

13: Kas on kirjeldatud, kuidas kaalud arvutati, sh välja toodud kasutatud tunnused ja kaalumisel kasutatud parameetrite allikad?

13a Vastuse aluseks olev osa raportist

13b Hinnang – 1 Jah Kaalude koostamispõhimõtted ja kaalude tunnused selgelt kirjeldatud; 2 Ei Kaale on kasutatud, aga nende olemust ega koostamispõhimõtteid pole selgelt kirjeldatud; 9 Ei kehti Kaalude kasutamist ei olnud võimalik raportist tuvastada

13c Selgitus

14: Kas on käsitletud andmete imputeerimist, andmestikust eemaldamist või asendamist?

14a Vastuse aluseks olev osa raportist

14b Hinnang – 1 Jah Täielikult kirjeldatud, millised andmed, millises ulatuses ja kuidas imputeeriti, eemaldati või asendati; 2 Ei Imputeerimist/eemaldamist/asendamist on rakendatud, aga pole selgitatud, millised andmed, millises ulatuses ja kuidas; 9 Ei kehti Imputeerimist/eemaldamist/asendamist ei ole võimalik tuvastada

14c Selgitus

15: Kas on ära toodud üksikasjad koostöö saavutamiseks kasutatud strateegiate kohta (nt eelkontaktid, kirjad, skriptid, rahalised motivaatorid ja nende kättetoimetamise viis, kontaktid keeldujate ümberveenmiseks)?

15a Vastuse aluseks olev osa raportist

15b Hinnang – 1 Jah Antud on täielik kirjeldus üksikasjadest koostöö saavutamiseks kasutatud strateegiate jaoks; 2 Ei Selgub, et selliseid strateegiaid rakendati, aga neid pole kirjeldatud või kirjeldus on ebaadekvaatne; 9 Ei kehti Selliseid strateegiaid ei rakendatud või neid ei olnud võimalik tuvastada.

15c Selgitus

16: Kas on kirjeldatud kvootide kasutamine?

16a Vastuse aluseks olev osa raportist

16b Hinnang – 1 Jah Kvootide kasutamise puhul on seda detailselt kirjeldatud; 2 Ei Kvootide kasutamise puhul pole seda detailselt kirjeldatud; 9 Ei kehti Kvootide ei kasutatud või neid ei olnud võimalik tuvastada

16c Selgitus

17: Kas paneelide puhul on kirjeldatud paneeliliikmete värbamine ja paneeli haldamine?

17a Vastuse aluseks olev osa raportist

17b Hinnang – 1 Jah Paneelide kasutamise puhul on kirjeldatud värbamisprotsessi ja paneeli haldamise detaile; 2 Ei Paneelide kasutamise puhul pole kirjeldatud värbamisprotsessi ega/või paneeli haldamise detaile; 9 Ei kehti Paneelide pole kasutatud või nende kasutamist ei olnud võimalik tuvastada

17c Selgitus

18: Kas tõenäosusliku valimi puhul on välja toodud vastamismäär, selle definitsioon ja arvutuskäik? Kas mittetõenäosusliku valimi puhul on välja toodud osalemismäär?

18a Vastuse aluseks olev osa raportist

18b Hinnang – 1 Jah; täielik Vastamismäär/osalemismäär välja toodud, olemas selge arvutuskäik (kui see ei ole enesestmõistetav) 2 Ei Vastamismäär/osalemismäär ei ole välja toodud või on seda ebaadekvaatselt kirjeldatud; 9 Ei kehti Vastamismäär ega osalemismäär ei kehti antud uuringu meetoodika puhul

18c Selgitus

19: Kas on välja toodud avaldus, milles kirjeldatakse kitsendusi ja piiranguid uuringudisainis ja andmete kogumises?

19a Vastuse aluseks olev osa raportist

19b Hinnang – 1 Jah Selgesõnaliselt uuringudisaini ja andmete kogumise piirangud kirjeldatud; 2 Ei Antud piirangute avaldust ei ole või kirjeldus on väga pealiskaudne; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

19c Selgitus

20: Kas statistilised tulemused on esitatud piisava täpsusega (nt täpsed p-väärtused, seose tugevus ja usaldusvahemikud)?

20a Vastuse aluseks olev osa raportist

20b Hinnang – 1 Jah Statistilised tulemused on esitatud täpselt: p-väärtused on esitatud täpse arvuna (nt $p = 0.035$, mitte lihtsalt $p < 0.05$), seose tugevust kirjeldatakse asjakohaste mõõdikute (nt korrelatsioonikordaja, regressioonikoefitsient) kaudu ning usaldusvahemikud on välja toodud standardiseeritud viisil (näiteks 95% usaldusnivool, näiteks $[0.12; 0.45]$); 2 Ei Statistilised tulemused on esitatud ebapiisava täpsusega: p-väärtused on esitatud ebamääraselt (nt $p < 0.05$), puuduvad seose tugevuse näitajad või need on esitatud ilma kontekstita, ja usaldusvahemikud on jäetud lisamata või pole märgitud usaldusnivood; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

20c Selgitus

21: Kas p-väärtused on esitatud ja tõlgendatud viisil, mis väldivad väärtõlgendusi?

21a Vastuse aluseks olev osa raportist

21b Hinnang – 1 Jah P-väärtused on esitatud ja tõlgendatud viisil, mis peegeldab nende tegelikku tähendust, neid käsitletakse kui andmete ja nullhüpoteesi vahelise kooskõla mõõdupuud, mitte kui tõenäosust, et nullhüpotees on tõene või väär. Kirjeldusest on selge, et väike p-väärtus tähendab vaid seda, et saadud andmed on ebatavalised, kui nullhüpotees kehtib, mitte tõendina, et nullhüpotees on vale. Järelduste tegemisel on arvestatud, et ainuüksi madal p-väärtus ei ole piisav tõestus sisulise seose / erinevuse olemasolu väitmiseks; 2 Ei P-väärtusi on tõlgendatud vääralt, neid on käsitletud kui tõenäosust, et nullhüpotees on tõene või väär, või on p-väärtusele omistatud tegelikkusest suurem tähendus – levinud vead hõlmavad näiteks järeldust, et suur p-väärtus tähendab seose või erinevuse puudumist või et väike p-väärtus on kindel tõend seose olemasolu kohta, samuti on probleemiks, kui p-väärtust kasutatakse tulemuste sisulise olulisuse või tähenduse ainsa indikaatorina; 9 Ei kehti P-väärtuseid ei ole kasutatud

21c Selgitus

22: Kas võtmetulemuste puhul on esitatud usaldusvahemikud?

22a Vastuse aluseks olev osa raportist

22b Hinnang – 1 Jah Võtmetulemuste ehk põhijärelduste puhul on esitatud adekvaatselt arusaadaval viisil usaldusvahemikud; 2 Ei Võtmetulemuste juures, kus need oleks kasulikud, puuduvad usaldusvahemikud; 9 Ei kehti Usaldusvahemikud ei ole antud analüüsi puhul asjakohased

22c Selgitus

23: Kas usaldusvahemikud on esitatud ja tõlgendatud viisil, mis kajastab määramatust tulemustes ja väldib eksitavaid üldistusi?

23a Vastuse aluseks olev osa raportist

23b Hinnang – 1 Jah Usaldusvahemikud on esitatud koos relevantse, praktilise kontekstiga. Usaldusvahemike tõlgendused on esitatud sellisel viisil, et nende kirjeldamisel välditakse ebakorrektselt „tõenäosuse“ väitmist, usaldusvahemiku esitamisel ei väideta seda, et tõeline väärtus on kindlasti antud vahemikus, usaldusnivood ei rakendata võimaluse või tõenäosuse näitajana; 2 Ei Usaldusvahemike tõlgendused on tehtud ebaadekvaatselt, näiteks on jäetud mulje, et vahemik sisaldab tõelist väärtust mingi tõenäosusega, või ei ole selgitatud, mida vahemiku laius praktiliselt tähendab; 9 Ei kehti Usaldusvahemikud ei ole antud analüüsi puhul asjakohased

23c Selgitus

24: Kas statistilised järeldused on kooskõlas analüüsi põhjal leitud tõenditega?

24a Vastuse aluseks olev osa raportist

24b Hinnang – 1 Jah Statistilised järeldused peegeldavad analüüsi tulemusi, on arvesse võetud seose tugevust, usaldusvahemikke ja p-väärtusi, ning järeldused on tehtud andmete tugevuste ja piirangute kontekstis – pole esitatud põhjendamatuid järeldavaid väiteid, näiteks ainult statistilise olulisuse põhjal; 2 Ei Statistilised järeldused ei vasta analüüsi tulemustele, näiteks tehakse liiga kategoorilisi või üleliigselt üldistavaid väiteid ainult p-väärtuse põhjal või liialdatakse seose tugevusega; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

24c Selgitus

25: Kas välditakse statistiliselt mitteoluliste tulemuste tõlgendamist tõendina seose või erinevuse puudumise kohta?

25a Vastuse aluseks olev osa raportist

25b Hinnang – 1 Jah Statistiliselt mitteolulisi tulemusi ei tõlgendata täielikult efekti puudumisena, vaid tunnistatakse, et mitteolulisus võib tuleneda ka väikesest valimist või ebapiisavast testimise võimsusest; 2 Ei Statistiliselt mitteoluline tulemus on ekslikult tõlgendatud tõendina selle kohta, et mõju puudub, eirates võimalust, et efekt võib siiski eksisteerida, kuid seda pole piisavalt täpselt hinnatud; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

25c Selgitus

26: Kas välditakse terminite „oluline“ ja „mitteoluline“ mitmetähenduslikul või eksitaval viisil kasutamist?

26a Vastuse aluseks olev osa raportist

26b Hinnang – 1 Jah Termineid „oluline“ ja „mitteoluline“ on kasutatud selgelt, on alati täpsustatud, kas viidatakse statistilisele olulisusele (nt p-väärtuse põhjal) või sisulisele olulisusele; 2 Ei Termineid on kasutatud mitmetähenduslikult või eksitavalt, pole selge, kas „oluline“ või „mitteoluline“ viitab statistilisele või sisulisele olulisusele või on jäetud mulje, et statistiline olulisus otseselt tähendab sisulist olulisust; 9 Ei kehti Uuringus ei kasutata neid termineid või nende tähendus pole asjakohane

26c Selgitus

27: Kas välditakse mudelist muutujate väljajätmise põhjendamist ainult nende hinnangute statistilise olulisuse põhjal?

27a Vastuse aluseks olev osa raportist

27b Hinnang – 1 Jah Muutujate kaasamine või väljajätmine mudelist põhineb teoreetilistel kaalutlustel või varasemal uurimistööl, mitte ainult statistilise olulisuse olemasolul või puudumisel – on arvestatud, et isegi statistiliselt mitteoluline muutuja võib olla sisuliselt tähtis; 2 Ei Muutujad on jäetud mudelist välja nende statistilise mitteolulisuse tõttu, eirates võimalikku sisulist tähtsust ja teoreetilist relevantust; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

27c Selgitus

28: Kas koefitsiendid on esitatud otstarbekal ja arusaadaval kujul, mis hõlbustab seose tugevuse mõistmist?

28a Vastuse aluseks olev osa raportist

28b Hinnang – 1 Jah Koefitsiendid on esitatud viisil, mis võimaldab hinnata seose tugevust, näiteks koefitsiendid on esitatud tabelites, joonistel või tekstis koos lisainfoga, mis teeb nende tõlgendamise lihtsaks. Kui on kasutatud raskesti tõlgendatavaid näitajaid, nagu logit-suhe, riskimäär või teisendatud skaalaga tunnuse regressioonikordaja, siis on lisatud täiendav selgitus või teisendus, mis muudab seose tugevuse praktiliselt mõistetavaks – näiteks on esitatud regressioonikordaja eksponentastmel või marginaalmõjud ning prognoositud tõenäosused, mis hõlbustavad tulemuste tõlgendamist; 2 Ei Koefitsientide esitus on ebaselge või raskesti tõlgendatav – seose tugevust on keeruline hinnata ja tulemuste tähendus ei ole selgelt mõistetav; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

28c Selgitus

29: Kas arutletakse tulemuste nagu teststatistikute, seosekordajate ja koefitsientide sisulise olulisuse üle, mitte ainult statistilise olulisuse üle?

29a Vastuse aluseks olev osa raportist

29b Hinnang – 1 Jah Arutelu keskendub tulemuste sisulisele tähendusele ja praktilisele olulisusele, mitte ainult sellele, kas need on statistiliselt olulised – näiteks võrreldakse seoste tugevust teiste uuringutega või hinnatakse, kas mõju on tegelikult ühiskondlikult või teaduslikult tähtis; 2 Ei Arutelu piirdub statistilise olulisuse kinnitamisega, jättes tähelepanuta tulemuste sisulise tähenduse ja praktilise tähtsuse; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

29c Selgitus

30: Kas välditakse statistikute võrdlemist pelgalt nende statistilise olulisuse või olulisuse tõenäosuste põhjal?

30a Vastuse aluseks olev osa raportist

30b Hinnang – 1 Jah Ei võrrelda lihtsalt erinevate statistikute statistilist olulisust või olulisuse tõenäosusi, vaid hinnatakse võrdlusefekti (erinevuse määra) esinemist ja suurust; 2 Ei Tulemusi võrreldakse peamiselt nende statistilise olulisuse põhjal, eirates, et olulisuse taseme erinevus ei

tähenda alati olulist erinevust tulemustes; 9 Ei kehti Küsimus ei ole antud uurimisraporti puhul asjakohane

30c Selgitus

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Karlis Põldoja,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose “Järeldava analüüsi eeldustele vastavus ja statistiliste järelduste adekvaatsus valitud Haridus- ja Teadusministeeriumi tellitud rakendusuuringutes aastatel 2013–2024”, mille juhendaja on Indrek Soidla, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Karlis Põldoja

27.05.2025