

Soovitused SPIN- programmi mõju hindamiseks

2023

PRAXIS | MÖTTEKODA

Uuringu tellija on MTÜ SPIN.

Autorid:

Maris Vainre

Triin Edovald

Viitamine: Vainre, M., & Edovald, T. (2023). Soovitused SPIN-programmi mõju hindamiseks. Tallinn: Mõttekoda Praxis.

Ahtri 6a, Tallinn, 10151

www.praxis.ee, praxis@praxis.ee

praxis | mõttekoda

Sisukord

Lähenedised mõju mõõtmisele.....	4
Eeldused kvaliteetseks täismahuliseks juhuslikustatud kontroll-uuringuks.....	5
Leida mõõdik peamise tulemusnäitaja mõõtmiseks.....	5
SPINi mõju mõõtmise võimalikud vahendid.....	5
Vajalik osalejate arv: 960.....	7
Märkused.....	8
Osalejate värbamine.....	8
Olulised otsustuskohad.....	8
Juhuslikustamise protseduur.....	9
SPINi kontekst ja lahendusvõimalused.....	9
Kontrollgrupi olemus.....	9
Gruppide ristumise vältimine.....	10
SPINi mõju hindamiskvaliteedi parandamise võimalused.....	11
TAI tase 4: Täismahuline juhuslikustatud kontroll-uuring.....	11
Eeltöö TAI tase 4 saavutamiseks: Piloot juhuslikustatud kontroll-uuring.....	12
Alternatiivsed mõju-uuringu metodoloogiad.....	13
Realistilik hindamine (<i>realist evaluation</i>).....	14
Protsessi jälgimine (<i>process tracing</i>).....	15
Panusanalüüs (<i>contribution analysis</i>).....	15
Kvalitatiivne võrdlev analüüs (<i>Qualitative Comparative Analysis, QCA</i>).....	16
Võrdlev juhtumiuuring (<i>comparative case study</i>).....	17
JuKU meetodiga mõju-uuringu eeldatav ajakava.....	19
Tehniline info.....	21
Osalejate arvu arvutamine.....	21
Viited.....	23

Lähenedes mõju mõõtmisele

Mõju eeldab kausaalsust, st nt programmis osalemine põhjustab muutust mõnes tulemusnäitajas. Et tulemusnäitajat võivad mõjutada ka muud faktorid (nt aja jooksul võivad mõned probleemid leeveneda), on mõju-uuringutes kesksel kohal muude mõju avaldada võivate faktorite välistamine. Selline välistamine vajab teatud uuringutüüpi: see peaks olema prospektiivne (ehk kausaalsust on keeruline omistada tagant järele) ja võrdlev (mõju avaldub millegagi võrreldes). Juhuslikustamine, ehk kahe või enama sarnase grupi loomine nii, et igal uuritaval on sama suur võimalus sattuda igasse uurimisgruppi, võimaldab veenduda, et ainus erinevus gruppide vahel on sekkumine või programm, mida grupiliikmed saavad (Chambliss & Schutt, 2018), ehk teisiti – muud võimalikult mõju avaldavad faktorid esinevad mõlemas grupis. Kui kõiges muus, va sekkumise saamine, on grupid võrdsed, ja peale sekkumist on gruppide tulemusnäitaja erinev, siis ainus põhjus erinevuse tekkeks on uuritav sekkumine.

Sellist eksperimentaalset (või juhuslikustatud) uuringutüüpi saab kasutada erinevate uurimisküsimuste lahendamiseks. Näiteks võib uurida programmi üldist mõju: kas programm „töötab“ (ehk kas programm avaldab mõju konkreetsele tulemusnäitajale, nt riskikäitumiste sagedus)? Samuti võib mõõta erinevate programmi komponentide mõju, mis annab teadmisi, kuidas programmi üldist mõju suurendada: kas programmi komponent mõjutab soovitud tulemust (kas enesekehtestamise oskuste õpetamine vähendab riskikäitumiste sagedust)? Või teha valikuid erinevate komponentide osakaalu kohta, programmis: kumb programmi komponent on mõjusam, kas riskikäitumiste sageduse vähendamisel on suurema mõjuga aeroobne füüsiline aktiivsus või oskus ennast kehtestada? See, millist uurimisküsimust tasub parajasti esitada, sõltub sellest, mis programmi kohta juba teada on (Skivington et al., 2021).

Terve programmi efektiivsuse (ingl. k. *effectiveness*) kui mõju (ing. K. *efficacy*) hindamisel on nõ kuldstandardiks kasutada juhuslikustatud kontroll-uuringut (JuKUt) (Gottfredson et al., 2015; Skivington et al., 2021), mida on mõistlik läbi viia siis, kui selle läbi viimiseks on eeldused täidetud.

Eeldused kvaliteetseks täismahuliseks juhuslikustatud kontroll-uuringuks

Leida mõõdik peamise tulemusnäitaja mõõtmiseks

Juhuslikustatud kontroll-uuring (JuKU) läbi viimine eeldab hästi defineeritud mõju mudelit ja hoolikalt valitud tulemusnäitajat ja seega tulemusnäitaja mõõdiku valikut. Nii on võimalik kontrolli all hoida nii vale-negatiivsete kui ka vale-positiivsete tulemuste tase (Andrade, 2015; Wagenmakers et al., 2012) ja seeläbi olla kindlam tulemuste vettpidavuses.

Projekti käigus täiendatud programmeerimisest selgub, et peamine SPINi eesmärk on vähendada osalenud noorte riskikäitumist (Edovald et al., 2023). Seega on mõistlik hinnata SPINi mõju riskikäitumisele, mida omakorda on tarvis täpsemalt defineerida – millised riskikäitumised?, millist muutust näha soovitakse: riski-käitumiste sageduse vähenemine, riskikäitumiste riskiastme vähenemine (nt riskikäitumiste arv jääb samaks, aga need on teatud kriteeriumi alusel (nt oht avalikule korrale) vähem ohtlikud – väheneb autovarguste arv, aga suureneb nt suitsetamine – ehk üks käitumine asendatakse teisega), kombinatsioon sagedusest ja riskiastmest – nt väheneb ohtlikumate riskikäitumiste sagedus, samas kui vähem ohtlikud ei muutu.

SPINi mõju mõõtmise võimalikud vahendid

Enne uuringut on vaja defineerida riskikäitumine/riskikäitumised, milles SPINi abil muutus kõige tõenäolisemalt tekib. Näiteks suitsetamist võib pidada riskikäitumiseks ent kui SPIN ei tegele tubakatoode tarvitamise vähendamise, ei ole tõenäoline, et just selles käitumises muudatust märgataks.

Meil ei ole olnud võimalik tuvastada kvaliteetset (st valideeritud ja eestindatud) riskikäitumiste mõõdikut. Seega on hetkel riskikäitumise mõõtmiseks suures plaanis kaks varianti: registriandmed (nt koolist puudumise andmed e-koolist) või enesekohased hinnangud (kui mitu korda puudusid? Kui mitu korda tarbisid tubakatooteid?). Mõlemal on head ja vead.

1. Registriandmed

- a. Registriandmed võivad paremini peegeldada tegelikku käitumist ent nende kätte saamisega võib olla probleeme. Seega tuleb leppida kokku andmete jagamise korras ja taotleda ka vastavad load (nt Andmekaitse Inspektsioonilt).

- b. Registriandmete abil on võimalik välja arvutada norm – nt mis on see näitaja nende õpilaste seas, kes ei kuulu SPINi sihtgruppi või just nende seas, kes kuuluvad. See annab aluse hinnata kui suur on muutus, mille SPIN esile kutsub kui muutus esineb. Sellise normi välja arvutamine eeldab ligipääsu kõikidele andmetele, mitte vaid SPINis osalejate andmetele.
- c. Harvaesinevaid käitumisi (nt Politsei registritest leitavaid) võib olla väga vähe. Meie osalejate arvu arvutused ei kehti selliste mõõdikute kohta, mis on sisuliselt binaarsed (nt on väärteos kahtlustatav või mitte). Seega valides harva esinev ja binaarne tulemusnäitaja, tuleb osalejate arv uuesti välja arvutada. See tõenäoliselt suureneb.

2. Enesekohased andmed

- a. Enesekohased andmed võimaldavad teatavat paindlikkust, sest küsida saab just huvitava riskikäitumise kohta.
- b. Samas sisaldub vastuses ka mälu ebatäpsus, sotsiaalne soovitus (üle- või ala-hinnatakse riskikäitumise sagedust või tõsidust), küsimuse erinevad interpretatsioonid. Ehk teisiti andmetes on müra ja seetõttu on keerulisem signaali märgata.
- c. Valideeritud ja kohandatud skaalad maandavad punktis 2 kirjeldatud riske.
- d. Normeeritud skaalad annavad võimaluse võrrelda SPINi andmeid „keskmise“ noore omadega.

Alternatiivina võib hinnata mõju peamise tulemusnäitaja mediaatorile – konstruktile, milles toimuv muutus on eelduseks muutusele peamises tulemusnäitajas. Loodud programmiteooria seisukohast võiks mõõta toimetulekut toetavaid hoiakuid (enesetõhusus, enesehinnang, väärtushinnangud), enesejuhtimise oskuseid (toimetulek ja säilenõtkus (ingl. k. *resilience*), enesekontrolli, emotsioone) või sotsiaalseid oskuseid (Edovald et al., 2023). Paljude nende konstruktide hindamiseks puuduvad kvaliteetselt eestindatud mõõdikud (Edovald et al., 2023), ent on ka erandeid (Öeren et al., 2023).

Kui valida mediaator peamiseks tulemusnäitajaks, siis nii saab selgitada küll SPINi mõju mediaatorile ent mitte lõpp-eesmärgile – vähendada riskikäitumist. Ülekanne peamisele tulemusnäitajale oleks sellisel juhul teoreetiline – see tugineks teaduskirjandusele, mis näitab nende kahe konstrukti vahelist (põhjuslikku) seost.

Võimalikke mediaatoreid on mitu, ent mitme puhul ei ole kasutuses kvaliteetsed mõõtevahendid. Eestindatud mõõtevahendi valimisel võime küll kindlad olla, et see mõõdab seda, mida ta on loodud mõõtma, ent me ei tea, kas mõõtevahendi indekseeritav muutus on SPINi programmiteooria seiskohalt olulisim. (Näiteks, kui muutust enesetõhususes ei ole, kas siis SPINil pole mõju või toimib mõjumehhanism

mõne muu konstrukti kaudu, mida me ei mõõtnud, nt eneseregulatsiooni oskused?). Kui mõõdame mõju eestindamata vahenditega, siis on mõju puudumisel keeruline öelda, kas põhjus on SPINis või mõõtevahendis.

Vajalik osalejate arv: 960

Osalejate arvu leidmiseks kasutasime sisend-parameetritena Tabelis 1 kirjeldatud andmeid (vt nende valimise põhjendusi **Tõrge! Ei leia viiteallikat.**: Osalejate arvu arvutamine lk 21) ja seejärel simuleerisime valitud parameetrite alustel andmeid 10 000 korda¹ ja arvutasime iga andmestiku kohta p-väärtuse (Lakens & Caldwell, 2021). Leidsime, et antud parameetrite juures oleks meil vaja 800 osaleja andmeid, et tuvastada valitud efekti suurust².

Need on osalejate arv, kelle andmeid on vaja analüüsida. Arvestades, et osad noored langevad SPINist välja või nende andmeid ei saa kätte, peaks igaks juhuks uuringusse kaasama rohkem noori. Siinkohal arvestasime varu 20% ehk kokku peab uuringusse kaasama 960 noort, kes jagunevad kahte võrdsesse gruppi, pooled, kes saavad SPINi ja teine pool, kes ei saa.

Arvutuste järgi on seega vaja 480 noort juhuslikult suunata SPINi ja teine 480 hoida kontrollgrupis. SPINi gruppe on üle Eesti 32 ja nii saab treenida mitut gruppi noori korraga. Kui SPIN treenib igal hooajal 160 noort (ja leiaks uuringusse 320 noort), siis on vaja osalejate arvu kätte saamiseks 3 hooaega.

Tabel 1. Valimi arvutuse eeldused

Parameeter	Väärtus
Coheni d	0,2
r (enne-pärast)	0,5
α	0,05
β	0,2
Statistilise testi jõudlus = $1 - \beta$	0,8

¹ Simuleerisime olukorda, kus viisime antud parameetritele tuginedes läbi 10 000 uuringut ja arvutasime iga kord, kas leidsime statistiliselt olulise erinevuse sekkumis- ja kontrollgrupi vahel ($p < 0,05$, sest alfa = 0,05). Seejärel lugesime kokku, mitmel korral statistiliselt oluline erinevus esines ja veendusime, et see juhtus vähemalt 80% uuringutest, st vähemalt 8 000 korral (ehk statistiline jõudlus on 0,8).

² Seda tulemust kontrollisime programmiga G*Power (Faul et al., 2007), mis ei arvesta andmekogumisega kaasnevast müra seotud variatsioonidega ja seega on veidi vähem täpsem. Saime vajalikuks osalejate suuruseks 790.

Märkused

1. Efekti suurus on suure tõenäosusega väiksem kui $d = 0,2$ ja võib tuleneda mitte ainult programmi enda tõhususest vaid ka SPINi rakendustruudusest. Väiksem efekti suurus eeldab suuremat osalejate arvu.
2. Valenegatiivse tulemuse tõenäosus on tõhususuuringuks pigem kõrge (20%). Kui seda alandada, on vaja suurendada osalejate arvu.
3. Oluline on, et neil 960 uuringus osalejale poleks varasemaid kokkupuuteid SPINiga. Enne otsust uuring läbi viia, on tähtis teada, kas on võimalik nii palju noori uuringusse leida – kas on nii palju noori, kes on nõus osalema.
4. Selleks, et vähendada projekti kogukestust, võib kaaluda SPINi programmi lühendamist, nt poole võrra. Nii saab vajaliku arvu osalejaid varem kätte. Samas tähendab see seda, et mõõdetakse SPINi lühema versiooni mõju.

Osalejate värbamine

Uuringut ette valmistades tuleb täpselt kokku leppida uuringusse kaasamise tingimused (nt noore vanus, millised on esinevad riskifaktorid). See võib vähendada sobilike osalejate hulka, ent vähendab ka müra andmestikus ja võimaldab seega kergemini märgata programmi mõju, kui mõju on olemas.

Osalejate värbamisel on oluline roll spetsialistidel, kes tavaliselt noori SPINi suunavad (nt sotsiaalpedagoogid). Nende koostöövalmidus on uuringu õnnestumise seisukohast võtmetähtsusega.

Osalejate värbamisel tuleb läbi mõelda ka see, kuidas noori ja nende vanemaid uuringus osalemiseks motiveerida. Siin on kaks aspekti. Esiteks, tähendab uuringus osalemise otsus nõustumist riskiga, et nad ei saa (kohe või mitte kunagi) osaleda programmis, millest räägitakse kui potentsiaalselt mõjusast.

Teiseks tuleb motiveerida osalejaid ja eriti kontrollgrupis osalejaid oma andmeid jagama. Enne sekkumist kogutavate andmetega ei ole üldiselt probleeme, sest uuringusse kaasatakse vaid need, kes on uuringus osalemiselega nõus ja kes täidavad ära küsimustikud (või annavad ligipääsu oma andmetele). Sekkumisjärgsetes andmekogudes on aga tihti puuduvaid andmeid. Sageli motiveeritakse osalejaid andmekogumisel osalema väikese rahalise tänuga.

Olulised otsustuskohad

1. Kokku leppida uuringusse kaasamise tingimused (seda võib teha jätku-uuringut ette valmistades).

2. Hinnata, kas SPINil on piisavalt hea koostöö spetsialistidega, kes aitaksid osalejaid uuringusse värvata ja mil määral nad on nõus osalema. Neil võib olla pigem vahendajaroll – edastada vanema nõusolekul info uurijatele, kes välja selgitavad, kas noor sobib uuringusse ja kas ta on nõus osalema. Nende roll võib olla ka suurem – nt värvata osalejaid, hinnata nende sobivust, teostada andmekogumist.
3. Kuivõrd on vanematel ja lastel huvi osaleda uuringus? Kas seda huvi on võimalik toetada nii, et kokku tuleks vajalik arv osalejaid?

Juhuslikustamise protseduur

Juhuslikustatud kontrolluuringu kvaliteedi tagamiseks on osalejad vaja juhuslikustada nii, et kõigil oleks võrdne võimalus sattuda kontroll- ja sekkumisgruppi. Selleks, et see tõenäosus oleks tõepoolest võrdne, ei pruugi piisata vaid tõenäosusteooriast, vaid vajalik on ka arvestada inimlike kalduvustega olla vastutulelik ja abivalmis. Niisiis on oluline juhuslikustamise protsess „pimestada“, st keegi uuringu korraldajatest ega neist, kes viivad kokku juhuslikustamise tulemuse osalejaga (kumba gruppi konkreetne noor läheb) ei saaks mõjutada osalejate gruppidesse paigutumist.

SPINi kontekst ja lahendusvõimalused

1. Uuringut ette valmistades on vaja läbi mõelda kes värbab, kes kogub andmeid, kes juhuslikustab, kes viib kokku juhuslikustamise tulemuse ja osaleja ning kes kontrollib kas juhuslikustamisest on kinni peetud.
2. Probleemid selles osas vähendavad uuringu kvaliteeti ja alust kausaalsust hinnata.

Kontrollgrupi olemus

Sõltuvalt uurimisküsimusest tuleb valida kontrollgrupp. Näiteks, kui tahame teada, kas SPIN on parem kui nõ „tavapraktika“, siis tuleb tagada, et kontrollgruppi juhuslikult määratud osalejad osaleksid tavapraktikas (nt tugi sotsiaalpedagoogilt vms) ega osaleks mingis muus programmis. „Tavapraktikat“ peaksid saama sellisel juhul ka SPINis osalevad noored.

Kontrollgrupi olemus tuleb kokku leppida uuringut ette valmistades. Kui kontrollgrupp on alternatiivne sekkumine, tuleb selleks eraldada ka ressursse.

Arvestades, et SPINi mõju pole varem hinnatud võib olla mõistlik kasutada nõ „tavapraktika“ gruppi.

Gruppide ristumise vältimine

SPIN lubab praegu trenni sõpru kaasa võtta. Uuringu käigus tuleb aga veenduda, et kaasa võetud sõber ei oleks kontrollgruppi sattunud uuringus osaleja. Seega tuleb koolitada treenereid seda küsima ja vajadusel „ei“ ütleva.

Samuti võib olla vajalik selgitada noortele, vanematele ja spetsialistidele, miks kontrollgrupis olemine tähtis on.

SPINi mõju hindamiskvaliteedi parandamise võimalused

Sekkumiste tõendatusse kvaliteedi hindamise raamistikke on mitmeid (nt Guyatt et al., 2008; Puttick & Ludlow, 2013). Ühe neist on loonud Tervise Arengu Instituut (TAI) (Tervise Arengu Instituut, 2023), mis meile teadaolevalt on esimene rakendatud omataoline raamistik Eestis (vt Joonis 1). Tasemete juures on oluline nii selgus sihtrühmas, programmiteoorias, mõju-uuringute disainis, tõendatuse kvaliteedis³ kui ka mõju kestuses.

Tase	Eesmärgi, sihtrühma, ülesehituse ja sisu kirjeldus on olemas	Toimimise loogika lähtub teoreetilisest raamistikust	Uuringud	Tõendatuse kvaliteet	Tulemuste püsivus / jätkuvus
Tase 1	Jah	Ei	Ei	Ei kohaldu	Ei kohaldu
Tase 2	Jah	Jah	Ei	Ei kohaldu	Ei kohaldu
Tase 3	Jah	Jah	Vähemalt üks kohalik või rahvusvaheline uuring näitab tegevuse positiivseid tulemusi sihtrühmas	1	Järelhindamist pole tehtud
Tase 4	Jah	Jah	Vähemalt üks kohalik või rahvusvaheline (pool)eksperimentaalne uuring näitab tegevuse positiivset mõju peamistele tulemusnäitajatele sihtrühmas	1-2	Järelhindamine on tehtud kuni kuus kuud peale ennetustegevuse elluviimist
Tase 5	Jah	Jah	Vähemalt üks kohalik ja rahvusvaheline (pool)eksperimentaalne uuring näitab tegevuse positiivset mõju peamistele tulemusnäitajatele sihtrühmas	2-3	Järelhindamine on tehtud peale kuue kuu möödumist tegevuse lõpust

Joonis 1. TAI ennetustegevuste mõju tõendatuse tasemed (allikas: TAI, 2023).

TAI tase 4⁴: Täismahuline juhuslikustatud kontroll-uuring

Täismahuline juhuslikustatud kontroll-uuring (JuKU), st selline, mille tulemusena saab teha järeldusi SPINi mõju kohta, vajab teatud eeltingimusi (vt

³ Siin on oluline nii nihke tõenäosus (*risk of bias*), mõõtevahendi kvaliteet (kui hästi mõõdab seda, mida mõõtma peaks), millise sihtgrupi suhtes on mõju uuritud kui ka publitseerimata jäetud andmete olemasolu tõenäosus (sh nii eelretsenseeritavates ajakirjades kui ka mujal (nt raportid, avaandmed jms)).

⁴ Sõltub muuhulgas uuringu ja selles kasutatud mõõdikute kvaliteedist, ehk eeldab et varem kirjeldatud probleemkohad ületatakse.

Eeldused kvaliteetseks täismahuliseks juhuslikustatud kontroll-uuringuks, lk 5). Hetkel on palju teadmatust: nt millist mõõtevahendit kasutada, kuivõrd lihtne on huvilisi leida, kas juhuslikustamise protsess mõjutab valmidust uuringus osaleda ja kuivõrd on spetsialistid nõus uuringurühmaga koosööd teha. Need ebakindlused kätkevad nii riski uuringu õnnestumisele kui ka uuringu valenegatiivsete tulemuste esinemise tõenäosusele.

Täismahuline JuKU on ajamahukas (umbes 4,5 aastat, kui värvata saab 320 osalejat aastas) ja nõuab märkimisväärset rahastust: 250 000 – 350 000€⁵.

Kui soovida teha eeltööd SPINi tõendatuse TAI 5. taseme vääriliseks tõstmiseks, siis on vaja 5 aastast uuringut, et mõõta SPINi mõju 6 kuud peale programmi lõppu. Osalejate arv peaks siis olema suurem, sest peale kuuekuulist pausi SPINist, on osalejate uuringusse kaasamine keerulisem ja mõju suurus on väiksem (ehk vaja on veelgi enam osalejaid). Sellise uuringu tarbeks on mõistlik arvestada 270 000 – 375 000 euroga.

Kommenteerinud [MV1]: Ei teadnud eriti, kuis arvutada. Mu katse on siin https://praxisestonia.sharepoint.com/:x:/r/sites/Praxis/_layouts/15/Doc.aspx?sourcedoc=%7BB510BBB4-3B58-490F-80CC-2645A01B1C28%7D&file=SPIN_j%C3%A4tkuprojekti_maksumus_eeldatav.xlsx&action=default&mobileredirect=true

Kommenteerinud [MV2R1]: Vaatasin, et NIHR jagab 3-aastastele RCTdele kuni £500 000 ja pilootidele kuni £250 000. UKs on küll palgad kõrgemad

Eeltöö TAI tase 4 saavutamiseks: Piloot juhuslikustatud kontroll-uuring

Protseduuriliste ebakindluste välja selgitamiseks kasutatakse sageli piloot-JuKUsid (Bowen et al., 2009; Eldridge et al., 2016; Ester et al., 2022; Hallingberg et al., 2018; Vainre et al., 2022). Piloot-uuringu eesmärk ei ole mõju hindamine vaid tulevase mõju hindamise uuringu võimalikkuse välja selgitamine. Piloot-JuKU abil saab ühelt pool koguda infot selle kohta, kas täismahus JuKUt on üleüldse võimalik läbi viia. Teisalt annab see olulist infot uuringudisaini valikute kohta (nt tulemusnäitaja, värbamisprotseduur jms). Näiteks:

- millist tulemusnäitajat on kõige mõistlikum valida (nt millist riskikäitumist uurida ja/või millised mediaatorit)?
- kuidas toimub andmetele ligipääs (kui kasutada registriandmeid)?
- millised probleemid võivad tekkida andmeid kogudes?
- kuidas toimib koostöö spetsialistidega, kes abistavad uuringusse värbamisel?
- kuivõrd on osalejatel huvi uuringus kaasa teha?
- milliseid probleeme võib tekkida juhuslikustamise protseduuris?

⁵ Siin ja edaspidi: kulude arvutamisel ei ole arvestatud SPINi enda kuludega (nt tööjõud) ega SPINi läbi viimise kuludega (materjalid, saalide üür jms). Hinna sisse oleme arvestanud ka programmi rakendamise ja protsessi hindamise (*implementation and process evaluation*) kulud (vahemikus 65 000 – 115 000€). Programmi rakendamise ja protsessi hindamine aitab selgitada, miks sekkumine oli või ei olnud edukas, millised tegurid aitasid kaasa tulemusel, milliseid muudatusi tuleks teha ja milliseid on üldisemad õppetunnid sellise programmi ja uuringu läbiviimisel. Tuleb silmas pidades, et eelarve hinnangu andmisel oleme jäänud pigem konservatiivseks.

Piloot-JuKUsse ei pea kaasama palju osalejaid, sest selle eesmärk ei ole mõõta mõju, vaid nõ uuringu tegemist harjutada. Niisiis võib piisata nt 50-100st osalejast.

Selline uuring kestaks umbes 2 aastat (sellest 9 kuud on SPINi enda kestus) ja maksaks hinnanguliselt 190 000 – 270 000€.

Alternatiivsed mõju-uuringu metodoloogiad

Ülaltoodust nähtuvad erisugused kitsaskohad, mis JuKU läbiviimise võimalikku edukust võivad piirata. Üheks neist on vaieldamatult soovitusliku valimi suurus, mis SPINi tegevusulatust ja seeläbi võimalike uuritavate mahtu, värbamisega ja uurimusse värvatute juhuslikustamisega ja nende uurimuses hoidmisega seotud kitsaskohti silmas pidades, võib kättesaamatuks osutada. Seda enam, et eelnevalt kirjeldatud soovitusliku valimi suurus põhineb pigem soodsatel kui konservatiivsetel eeldustel (nt väikseim tuvastatava efekti suurus).

Selline olukord pole otseloomulikult tavatu ning väikeste kohortide puhul tuleb sageli kaaluda erineval metodoloogial põhinevaid mõju-uuringuid. Viimaseid kutsutakse ka 'väikse n-iga' (ingl.k. *small-n*) mõju-uuringuteks (White & Phillips, 2012). Olgu siinkohal öeldud, et väikse n-iga metodoloogia valiku puhul ei ole tegemist puhtalt valimi suurusest lähtuva pragmaatilise valikuga. Mainitud metodoloogiad lähtuvad pigem erinevast kausaalse ehk põhjusliku järeltuleva vaatenurgast (võrreldes JuKUga), pannes seejuures rõhuasetuse huvialustele küsimustele (Dawid, 2007). Näiteks, JuKut kui mõju-uuringu disaini soovime SPINi puhul ideaalis rakendada selleks, et uurida sotsiaalsete- ja enesejuhtimisoskuste (põhjused) mõju noorukite riskikäitumisele (tagajärg)⁶. Alternatiivseid mõju-uuringu disainide puhul uurime aga näiteks seda, et milline on SPINi kui sekkumise roll riskikäitumise vähenemisel. Teisisõnu, sekkumist nähakse kui ühte võimalikku mõjutegurit, mis tegutseb laiemas kontekstis, kus ilmnevad ka teised mõjutegurid ning fookus on tagajärgede (riskikäitumine) erinevatel põhjustel (sh sotsiaalsed- ja enesejuhtimisoskused), mis käivitavad muutuste mehhanismi.

Alljärgnevas tabelis on kirjeldatud mõned alternatiivid JuKULE, mis autorite hinnangul edasistel SPINi uuringutel võivad kasulikuks osutada ja mida SPINi juhtkond saab huvi korral kaaluda. Olgugi et alternatiivsed disainid on mõju-uuringute liigituse vaatenurgast erinevad, ilmnevad neil teatud kattuvused: (1) nad on teooriapõhised ja

⁶ Mõttekäik kehtib üldjoontes ka muude kontrafaktuaalsete mõju hindamise disainide puhul, kus kasutatakse kontroll- või võrdlusgruppi. JuKU eeliseks on see, et juhuslikustamise tulemusena saadud statistiliselt samaväärsete gruppide võrdlemisel on võimalik mõju kõige kindlamalt põhjuslikule tegurile (st sekkumisele) omistada.

Kommenteerinud [TE3]: See on ilmselt see 'mitte-eksperimentaalsete uuringudisainide' osas ülal?

Ma panin väga üldjoonelise kirjelduse ja valisin mõned ehk eelistatavamad lähenemised, nii et see pole ammendatud nimistu. Ja neist ma isegi ei pakuks midagi välja konkreetselt, vaid olen esitanud need neile kaalumiseks. Ma arvan, et oluline on nendega arutada, missugustele küsimustele nad vastuseid tahaks saada ja siis saab aidata neid valikuid ka teha.

keskenduvad teooria kinnitamisele ja alternatiivsete hüpoteeside kummutamisele ja (2) kesksel kohal on 'juhtum' kui keerukas 'üksus', kus mitmed mõjutegurid üksteist mõjutavad ja mida ei tohiks segi ajada näidisjuhtumi või üksikjuhtumi mõistega (Stern jt, 2012).

Realistlik hindamine (*realist evaluation*)

Fookus: Generatiivne mehhanism (st kuidas sekkumine töötab). Täpsemini keskendub küsimusele 'Mis töötab, millistes aspektides, mil määral, millistes kontekstides ja kuidas?' (Pawson & Tilley, 1997).

Kirjeldus: Keskendub programmeerimise eri komponentide hindamisele ja ennekõike mehhanismidele, mis kutsuvad esile tulemusi konkreetses keskkonnas. Lähtekohaks on, et põhjuslikud mehhanismid ja nende tagajärjed ei ole fikseeritud, vaid sõltuvad kontekstist. Realistliku hindamise puhul on kesksel kohal kontekst-mehhanism-tulemus (*context-mechanism-outcome*) konfiguratsioonid, mis moodustavad realistliku analüüsi peamise struktuuri. Uuringu raames kaardistatakse tulemused ja andmete pinnalt kujunevad tulemuste mustrid, seejärel analüüsitakse tulemusi genereerivad mehhanismid ja seejärel vaadeldakse, millises kontekstis tuvastatud mehhanismid aktiveerusid või mitte. Kontekstid võivad siinkohal olla seotud näiteks kasusaajate alagruppidega, sekkumise elluviimise protsessiga, kultuurilise ja poliitiliste tingimustega jmt.

Meetod: Realistlik hindamine on meetodi osas neutraalne. Kasutatakse nii kvalitatiivseid kui kvantitatiivseid meetodeid, kuid kvalitatiivsed meetodid kalduvad tavalisemad olema.

Metodoloogia puhul tuleks silmas pidada järgmist:

- Antud metodoloogia on sobilik hindamiseks, kuidas sekkumised just keerulistes olukordades töötavad.
- Annab infot selle kohta, kuidas sekkumine töötab ja millised tingimused peavad kehtima, et konkreetsed mehhanismid töötaksid (st konteksti eriomased jooned).
- Realistliku hindamise ulatus tuleb määratleda nagu iga uuringu puhul ning see pole ilmingimata ressursirohkem lähenemine kui muud teooria-põhised lähenemised, kuid see võib kujuneda kallimaks kui näiteks kontrollgrupita eel- ja järeluuringu disain.
- Sedalaadi uuringuid viivad sageli läbi multidistsiplinaarsed uurimismeeskonnad, eriti siis, kui on tegemist ulatuslike või keerulisemate

realistlike hindamistega, kuid pole välistada väiksemate realistlike hindamiste läbiviimine ühe või paar uurija poolt.

Huvi korral tasub vaadata intervjuud ühe realistliku hindamise metodoloogia loojaga, Ray Pawsoniga, mis on jaotatud neljaks osaks ning mille esimene osa on saadaval veebilingil: <https://vimeo.com/84215487>

Protsessi jälgimine (*process tracing*)

Fookus: Generatiivne mehhanism (st kuidas sekkumine detailselt töötab). Rõhuasetus küsimusel 'Kas ja kuidas on võimalik(ud) põhjus(ed) mõjutas(id) täpsustatud muutust või muutusi?'

Kirjeldus: Protsessi jälgimise näol on tegemist juhtumipõhise lähenemisega põhjuslikule järeldusele, mille raames kasutatakse juhtumisiseseid 'vihjeid' (põhjusprotsess vaatlusi, ingl. k. *causal-process observations*, CPOs), et otsustada, milline võimalikest alternatiivsetest selgitustest kehtib (Protsessi jälgimist saab kasutada nii selleks, et näha, kas tulemused on kooskõlas programmi teooriaga, kui ka selleks, et näha, kas on võimalik välistada alternatiivseid selgitusi (vaata lisaks Brady & Collier, 2010; Collier, 2011).

Meetod: Valdavalt kvalitatiivsed meetodid, kuid kasutatakse ka kvantitatiivseid meetodeid.

Metodoloogia puhul tuleks silmas pidada järgmist:

- Protsessi jälgimine võimaldab (detailselt) tuvastada põhjuslikkuse suuna.
- Samuti võimaldab antud metodoloogia võrrelda tuvastatud põhjuslikkust alternatiivsete põhjuslikkuse selgitustega.
- Võimalik rakendada mistahes muutuse põhjuste uurimisele eeldusel, et meil on piisavalt tõendust asetleidnud muutuste kohta.
- Protsessi jälgimiseks vajalike testide rakendamine ei ole nii lihtne kui esmapilgul paistab.
- Metodoloogia rakendamine on osaluspõhine ning eeldab märkimisväärset määral programmiga seonduva personali aega ja tähelepanu.
- Programmi jälgimise raames on vaja genereerida rohkesti tõendusmaterjali muutuse mehhanismi selgitamiseks.
- Metodoloogiat saab kombineerida alternatiivsete metodoloogiatega nagu panusanalüüs ja võrdlev juhtumiuuring.

Panusanalüüs (*contribution analysis*)

Fookus: Kontributsioon ehk panus. Rõhuasetus on küsimusel 'Millise panuse on sekkumine andnud muutus(t)ele?' (Mayne, 1999, 2008).

Kirjeldus: Panusanalüüsi eesmärgiks on pigem produtseerida usutav, tõenduspõhine kirjeldus sekkumise panuse kohta, millega üks mõistlik inimene nõus on, kui et pakkuda vaieldamatut tõendust sekkumise panuse osas. Panusanalüüs käib käsikäes programmeerimisega, mis on panusanalüüsi aluseks. Viimase raames hinnatakse, kas ja mil määral programmeerimise teostatud sammud ja muutused reaalsusega kattuvad.

Meetod: Kvalitatiivsed ja kvantitatiivsed uurimismeetodid.

Metodoloogia puhul tasub silmas pidada järgmist:

- Panusanalüüsi näol on tegemist võrdlemise uue metodoloogiaga, mis pole mittetulundussektoris nii Eestis kui väljaspool laialdast kasutust leidnud.
- Panusanalüüs selgitab, kuidas ja miks muutus(ed) on toimunud, mis on eriti abiks siis, kui on soov tegevust laiendada (st sekkumist korrata).
- Antud metodoloogia ei vaja sekkumise alguses tuvastatud lähteandmeid ega kontrollgruppi, kuna programmi muutuste teooria võib kindlaks teha ka peale sekkumist.
- Ideaalis tehakse panusanalüüsi korduva protsessina, mille käigus toimub korduv tõenduse kogumine ja analüüsimine ning seeläbi täiustatakse panuse kirjeldust. See on aga aja ressursimahukas protsess, mis võib keerukaks osutuda, kuna enamik hindamisi viiakse läbi piiratud ressursside ja fikseeritud ajakavadega.
- Sarnaselt protsessi jälgimisega nõuab panusanalüüs rohkelt programmiga seotud personali aega, mistõttu tasub kaaluda, kui kasulikuks panusanalüüsist saadud tulemused osutuvad (st kulu ja tulu).
- Panusanalüüs eeldab selget programmeerimist, kus on väga vähe varieeruvust selles osas, kuidas sekkumist ellu viiakse. Metodoloogia on vähem sobilik juhtudel, kui huvialas on alternatiivsete programmeerimiste uurimine või arendamine.

Kvalitatiivne võrdlev analüüs (*Qualitative Comparative Analysis, QCA*)

Fookus: Konfiguratsioon.

Kirjeldus: Tegemist on metodoloogiaga, mis võimaldab keerukas olukorras analüüsida mitut juhtumit ja seeläbi selgitada, miks mõningatel juhtudel toimub muutus ja mõningatel juhtudel mitte (Befani, 2016). Antud lähenemine ei välista, et erinevate

juhtumite puhul kehtib erinev põhjuslikkuse edasikandumistee. Metodoloogia on kavandatud kasutamiseks juhtumite arvuga vahemikus 10-50, mil pole piisavalt juhtumeid tavapärase statistilise analüüsi rakendamiseks (ja samas on juhtumeid liiga palju, et puhtal kvalitatiivselt läheneda). Sisuliselt aitab metodoloogia otsida ja leida juhtumiüleseid mustreid, et paremini aru saada sellest, miks mõned muutused toimuvad ja mõned mitte. Kvalitatiivse võrdleva analüüsi puhul järgitakse ranget protsessi, mida kõikide juhtumite puhul võrdväärselt rakendatakse ning mis ideeliselt peaks võimaldama saadud tulemuste korratavust.

Meetod: Kvalitatiivsed ja kvantitatiivsed meetodid.

Metodoloogia puhul tasub silmas pidada järgmist:

- Kvalitatiivne võrdlev analüüs võimaldab juhtumitega minna süvitsi, kui samaaegselt samuti tuvastada juhtumite üleseid mustreid. Antud metodoloogia puhul on võimalik vältida kitsaskohti, mis esinevad puhtalt kvalitatiivsete (tulemuste korratavus) või kvantitatiivsete lähenemist puhul (kontekstide erisustega arvestamine).
- Range metodoloogia tingimustes, kus pole võimalik moodustada piisava statistilise jõuga valimit kvantitatiivse uuringu läbiviimiseks.
- Teatud tingimustes võimaldab metodoloogia testida programmiteooriat.
- Programmi seire and hindamise kontekstis abiks vahend õppetundide ja soovitude genereerimiseks, kuna metodoloogia ei keskendu otseselt muutuste hindamisele.
- Metodoloogia rakendamiseks on vajalik miinimumjuhtumite arvu, milleks sageli peetakse 10 juhtumit.
- Kvalitatiivse võrdleva analüüsi puhul ei saa kasutada andmeid, mis on puudulikud. Viimane aga tekitab riski, et oluline info jääb uurimise alt välja just puuduliku andmestiku tõttu.
- Uuringut alguses on sageli raske hinnata, kui palju aega ja ressursi uuring võtab, kuna metodoloogia raames tuleb korduvalt käia edasi-tagasi mööda analüüsi, juhtumeid ja programmiteooriat. Seega, antud lähenemine pole ilmtingimata sobilik piiratud ressursside ja fikseeritud ajakava korral.

Võrdlev juhtumiuuring (*comparative case study*)

Fookus: Konfiguratsioon.

Kirjeldus: Võrdlev juhtumiuuring kätkeb endast kahe või enama juhtumi analüüsimist moel, mis genereerib üldistatavamast teadmist põhjuslike küsimuste kohta (võrreldes üksikjuhtumiga), eriti just selles osas, mis puudutab seda, kuidas ja kas teatud

sekkumised töötavad või ei tööta (Kaarbo & Beasley, 1999). Võrreldavaid juhtumiuuringuid viiakse läbi teatud ajaperioodi vältel ja rõhuasetus on kontekstisisesel ja kontekstiülestel võrdlustel. Mõneti sarnaselt kvalitatiivse võrdleva analüüsiga on uuringu fookuses sarnase eesmärgiga juhtumite sarnasuste, erinevuste ja mustrite analüüsimisel ja sünteesimisel. Võrdleva juhtumiuuringu puhul on rõhuasetus põhjuslikkuse uurimisel: mil määral sekkumine põhjustas saadud tagajärgi, sealjuures arvestades kontekstuaalsete mõjudega (vt ka Goodrick, 2017).

Meetod: Kvalitatiivsed ja kvantitatiivsed meetodid.

Metodoloogia puhul tasub silmas pidada järgmist:

- Väidetavalt võib võrdlevat juhtumiuuringute kasutada nii põhjusliku omistamise kui panuse identifitseerimiseks. Seda siis olukordades, kus kontroll- või võrdlusgrupi moodustamine ei ole võimalik ja kus konteksti mõistmist nähakse olulise komponendina sekkumise edu või ebaedu selgitamisel.
- Lähenemine eeldab, et sekkumise elluviimise periood vältel on võimalik korratav andmekogumine ja analüüsi.
- Sobilik kombineerimiseks kvalitatiivse võrdleva analüüsiga (vt ülal).
- Eeldab oskuslikku uurimismeeskonda, kes valdab mitte ainult kvalitatiivseid ja kvantitatiivseid uurimismeetodeid, vaid kes oskavad ka arendada teooriaid ning sünteesida nii vastuolulist kui kattuvad tõendust, mis uurimise käigus ilmneb.
- Võrdlev juhtumiuuring võib osutuda väga ressursirohkeks. Üks võimalus seda leevendada on valida madalam arv juhtumeid.

Ülaltoodu näidete puhul ei ole me välja toonud alternatiivsete lähenemiste detailseid metodoloogilisi samme, vaid pigem üritanud ilmestada metodoloogia rõhuasetust, mis puudutab võimalike uurimisküsimuste käsitlemist. Järgmiseks sammuks võiks siinkohal olla võimalike uurimisküsimuste formuleerimine ja seejärel koos otsustamine, kas ja milline alternatiivne moodus JuKUle võiks kõne alla tulla.

JuKU meetodiga mõju-uuringu eeldatav ajakava

Ajakava võib ootamatuste tõttu nihkuda, ent kui kõik sujub, võib ajakava välja näha umbes selliselt:

Uuringu ülesanded	Kuluv aeg	Selgitus, mõned riskid ajakavale
Ettevalmistus		
Tulemusnäitajate valik, kohandamine	6-7 kuud	Kui tulemusnäitajaid on vaja kohandada või testida, eeldab see eraldi uuringut, st ka uuringu planeerimist, valimi arvutamist, eetikakomisjoni luba, andmeanalüüsi jne.
Värbamisprotseduuris ja kaasamiskriteeriumites kokku leppimine		See osa eeldab koostööd programmi suunajatega, sh nende koolitamist ja suunajate nõusoleku saamist uuringus osaleda (andmeilt neilt ei koguta; vajalik on v nende valmisolekut uuringule kaasa aidata)
Juhuslikustamisprotseduuris kokku leppimine		Siin on mitu tegevust koos. Esiteks, kuidas genereeritakse juhuslikud jaded, mis näitavad gruppidesse jaotumist. Teiseks, kuidas need jaded viiakse kokku uuringus osalejatega nii, et osaleja satuks just sellesse gruppi, kuhu ta määrati, mitte sellesse grupi, kuhu määraja (nt sotsiaalpedagoog) arvab ta paremini sobivat. Seega on tarvis teatav infovahetus, nt et kui leitakse osaleja, siis uurijad ütlevad suunajale, kumba gruppi see osaleja läheb.
Meetodid tagamaks osalejate maksimaalne kohalkäik SPINi trennides		
Meetodid tagamaks andmetäius		Et oleks võimalikult vähe puuduvaid andmeid ja võimalikult täpsed andmed (nt et osalejad ei vastaks suvaliselt).
Andmekogumise protseduuris kokku leppimine		
Andmeanalüüsi protseduuris kokku leppimine		
Rakendusuuringu kavandamine		Hea oleks mõju kõrval uurida, kuidas SPINi rakendati ja kuivõrd rakendustruu on SPINi läbi viimine.

Uuringu eelregistreerimine		Oluline samm ent see ei lisa oluliselt aega kogu uuringu ajakavale. See vajab küll ajalist panust, ent selle võrra on raporti kirjutamine ja eetikakomisjoni taotluse kirjutamine hõlpsamad.
Eetikakomisjoni loa taotlemine	2-3 kuud	Komisjon on hõivatud projektidega ega saa kõiki uuringuid läbi vaadata. Komisjonil on kahtlusi selles osas, kuidas osalejate huve kaitstakse.
Värbamine		
	1 kuu	Osalejaid on raske leida, need ei tule piisavalt palju kokku. Mida lühem on värbamisperiood, seda intensiivsemalt tuleb noori värvata. Samas neid muidugi mitte sundides. Vajalik võib olla mitu värbamisperioodi, et uuritavate arv kokku saada.
Läbiviimine		
Üks kohort	9 kuud	Sõltub doosi suurusest ja vajaliku kohortide arvust. Täismahulise JuKU tarbeks läheb vaja mitut kohorti.
Analüüs		
Andmete puhastamine	3 kuud	Sõltuvalt andmete kvaliteedist ja allikatest võib võtta rohkem või vähem. Tavaliselt alahinnatakse aega, mis see võtab.
Mitmene imputatsioon puuduvate andmetega arvestamiseks	1 kuu	Võib vahele jätta ent selle kasutamine suurendab statistilist võimsust planeeritule lähemale. Eeldab ligipääsu super-/klasterarvutile. Selle ligipääsu ostmise peaks olema eelarvestatud, kui ligipääs ei ole tasuta (nt ülikoolidel võib olla tasuta).
Andmete analüüs	1-2 kuud	Sõltub tulemusnäitajate arvust, andmestiku omadustest ja sobivate statistiliste meetodite olemusest. Sõltuvalt imputeeritud andmestike arvule, võib vajada ligipääsu klasterarvutile.
Raporti kirjutamine	1-2 kuud	
Programmi rakendamise ja protsessi hindamine		

Tehniline info

Osalejate arvu arvutamine

Statistiliste järelduste tegemine tugineb sageli tõenäosusteoorial. Selleks, et kogutud andmetelt oleks võimalik teha järeldusi nii, et saame hinnata riski, et need ei pea paika, on tähtis teada nii vale-positiivse kui ka vale-negatiivse mõju avaldumise tõenäosust. Nende tõenäosuste kontrolli all hoidmiseks tuleb uuringut ette valmistades hinnata, kui palju osalejaid on vaja, et uurijate (ja/või SPINi meeskonna) aktsepteeritud tõenäosusega oleks võimalik märgata SPINi mõju juhul, kui SPINil on mõju (Lakens, 2022).

Selle arvutuskäigu juures on oluline omada selget ettekujutust sellest, milline on sekkumise oodatava mõju suurus. Ennetavate programmide mõju on väike, võrreldes sekkumisprogrammidega, mida viiakse läbi nende hulgas, kellel probleem on juba avaldunud.

Efekti suuruse väljendamiseks on mitu lähenemist ent eksperimentaalse uuringu kontekstis käsitletakse efekti suurust sageli⁷ kui standardiseeritud keskmiste erinevust (ingl. k. *standardised mean difference*, *SMD*) kahe juhuslikustatud kontrollgrupi vahel. Vahel nimetatakse seda ka Coheni d-ks (ingl. k. *Cohen's d*).

SPIN on ennetusprogramm, täpsemalt sihistatud ennetusprogramm. Allolev arvutus põhineb eeldusel, et tulemusnäitajat saab mõõta vähemalt ordinaalskaalal (nt esineb üldse mitte, harva, sageli, iga päev) (st mitte nominaal skaala (milline riskikäitumine) ega binaarne tunnus (esineb/ei esinenud)).

Me hindame, et mõistlik oodatav SPINi mõju suurus on $d = 0,2$ ehk SPINi tulemusel on sekkumisgrupi ja kontrollgrupi vaheline erinevus tulemusnäitajas on 0,2 standardhälvet. See on pigem optimistlikult suur (Gutman & Schoon, 2015). Et mõju arvestamine oleks täpsem, peaks tulemusnäitajat mõõtma nii enne kui ka pärast sekkumist. See võimaldab arvestada gruppide algtasemega. Enne ja pärast mõõtmised aga korreleeruvad omavahel. Puuduvad head valdkonnaspetsiifilised andmed ennustamiseks kui võrd need korreleeruvad, seda enam, et see sõltub mõõdiku kvaliteedist. Oma arvutustes lähtusime teaduskirjanduses välja pakutust: $r = 0,5$ (Walters et al., 2019). Valimi suuruse arvutamisel on oluline paika panna ka tõenäosus,

⁷ Kuigi olemas on ka teisi efekti suuruse väljendusi, siis need on omavahel konverteeritavad ja valik millist kasutada, ei mõjuta oluliselt lõpptulemust – valimi suurust.

et leiame valepositiivse tulemise (nn α). Selle valik on subjektiivne ehk sõltub vajadusest. Ütleme, et lepime 5% tõenäosusega, et saame valepositiivse tulemise ($\alpha = 0,05$). Seejärel beeta, ehk tõenäosus, et me ei märka gruppide vahelist erinevust, kuigi see on olemas (valenegatiivse tulemus). Ütleme, et see on 20% ($\beta = 0,2$, mis on uuringutes levinud praktika, ent nagu kõik konventsioonid, võib seda murda).

Viited

- Andrade, C. (2015). The primary outcome measure and its importance in clinical trials. *The Journal of Clinical Psychiatry*, 76(10), e1320-1323.
<https://doi.org/10.4088/JCP.15f10377>
- Befani, B. (2016). *Pathways to Change: Evaluating development interventions with Qualitative Comparative Analysis (QCA)*. Sztokholm: Expertgruppen för biståndsanalys (the Expert Group for Development Analysis). <http://eba.se/en/pathways-to-change-evaluating-development-interventions-with-qualitative-comparative-analysis-qca>
- Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C., & Fernandez, M. (2009). How We Design Feasibility Studies. *American journal of preventive medicine*, 36(5), 452–457.
<https://doi.org/10.1016/j.amepre.2009.02.002>
- Brady, H. E., & Collier, D. (Toim). (2010). *Rethinking social inquiry: Diverse tools, shared standards* (2. ed). Rowman & Littlefield.
- Chambliss, D. F., & Schutt, R. K. (2018). Chapter 6. Causation and Experimental Design. *Making Sense of the Social World. Methods of Investigation*. SAGE.
- Collier, D. (2011). Understanding Process Tracing. *PS: Political Science & Politics*, 44(4), 823–830. <https://doi.org/10.1017/S1049096511001429>
- Dawid, A. P. (2007). *Fundamentals of statistical causality* (279; RSS/EPSRC Graduate Training Programme). Department of Statistical Science, University College London.
https://www.researchgate.net/publication/242495222_Fundamentals_of_statistical_causality
- Edovald, T., Ester, J., & Öeren, M. (2023). *SPIN-programmi programmiteooria*. Mõttekoda Praxis.
- Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Hopewell, S., Coleman, C. L., & Bond, C. M. (2016). Defining Feasibility and Pilot Studies in Preparation for Randomised Controlled Trials: Development of a Conceptual Framework. *PLOS ONE*, 11(3), e0150205. <https://doi.org/10.1371/journal.pone.0150205>

- Ester, J., Koppel, K., Vainre, M., & Öeren, M. (2022). *Alarm bell: Preventing upper secondary school dropouts in Estonia – a feasibility randomised controlled trial*. <https://doi.org/10.17605/OSF.IO/26TW9>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Goodrick, D. (2017). *Comparative Case Studies: Methodological Briefs - Impact Evaluation No. 9* (9; Methodological Briefs - Impact Evaluation, lk 17). UNIFEC. <https://www.unicef-irc.org/publications/754-comparative-case-studies-methodological-briefs-impact-evaluation-no-9.html>
- Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of Evidence for Efficacy, Effectiveness, and Scale-up Research in Prevention Science: Next Generation. *Prevention Science*, *16*(7), 893–926. <https://doi.org/10.1007/s11121-015-0555-x>
- Gutman, L. M., & Schoon, I. (2015). Preventive Interventions for Children and Adolescents: A Review of Meta-Analytic Evidence. *European Psychologist*, *20*(4), 231–241. <https://doi.org/10.1027/1016-9040/a000232>
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*, *336*(7650), 924–926. <https://doi.org/10.1136/bmj.39489.470347.AD>
- Hallingberg, B., Turley, R., Segrott, J., Wight, D., Craig, P., Moore, L., Murphy, S., Robling, M., Simpson, S. A., & Moore, G. (2018). Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: A systematic review of guidance. *Pilot and Feasibility Studies*, *4*(1), 104. <https://doi.org/10.1186/s40814-018-0290-8>
- Kaarbo, J., & Beasley, R. K. (1999). A Practical Guide to the Comparative Case Study Method in Political Psychology. *Political Psychology*, *20*(2), 369–391. <https://doi.org/10.1111/0162-895X.00149>
- Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, *8*(1), 33267. <https://doi.org/10.1525/collabra.33267>

- Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, 4(1), 251524592095150. <https://doi.org/10.1177/2515245920951503>
- Mayne, J. (1999). Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly. *The Canadian Journal of Program Evaluation*. <https://www.semanticscholar.org/paper/Addressing-Attribution-through-Contribution-Using-Mayne/0a8ef4389853342e52adebf20518cd3dald24f74>
- Mayne, J. (2008). *Contribution analysis: An approach to exploring cause and effect*. he Institutional Learning and Change (ILAC) Initiative. <https://www.betterevaluation.org/tools-resources/contribution-analysis-approach-exploring-cause-effect>
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. Sage.
- Puttick, R., & Ludlow, J. (2013). *An approach that balances the need for evidence with innovation*.
- Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. *BMJ*, 374, n2061. <https://doi.org/10.1136/bmj.n2061>
- Tervise Arengu Instituut. (2023). *Ennetustegevuste tõendatuse taseme hindamine*. <https://www.tai.ee/sites/default/files/2023-02/T%C3%B5endatuse%20taseme%20hindamine.pdf>
- Vainre, M., Galante, J., Watson, P., Dalglish, T., & Hitchcock, C. (2022). Protocol for the Work Engagement and Well-being Study (SWELL): A randomised controlled feasibility trial evaluating the effects of mindfulness versus light physical exercise at work. *BMJ Open*, 12(4), e050951. <https://doi.org/10.1136/bmjopen-2021-050951>
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, 7(6), 632–638. <https://doi.org/10.1177/1745691612463078>

- Walters, S. J., Jacques, R. M., Dos Anjos Henriques-Cadby, I. B., Candlish, J., Totton, N., & Xian, M. T. S. (2019). Sample size estimation for randomised controlled trials with repeated assessment of patient-reported outcomes: What correlation between baseline and follow-up outcomes should we assume? *Trials*, *20*(1), 566. <https://doi.org/10.1186/s13063-019-3671-2>
- White, H., & Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework* (Working paper no 15). International Initiative for Impact Evaluation.
- Öeren, M., Edovald, T., Ester, J., Jaanson, K.-L., Kõiv, Kurmiste, A., Mürk, A., Saks, K., & Vainre, M. (2023). *SPIN-programmi valideerimisuring*. Mõttekoda Praxis.