



Arviointiohjeet SOTE-tekoälykokeilujen vaikutusten ja vaikuttavuuden arviointiin

Virtanen, L., Kortelainen, M., & Heponiemi, T.

30.8.2024

Sisällys

- Arviointiohjeiden tarkoitus ja perusta
- Tiivistelmä
- SOTE-tekoälykokeilujen arviointiohjeet
 1. Vaikuttavuuden arviointi
 - Määritelmä
 - Vaikuttavuusarvioinnin suunnittelu: tutkimusasetelmat, indikaattorit ja tiedonkeruun menetelmät
 2. Muu arviointi
 - Tekninen toimivuus
 - Turvallisuus ja eettisyys
 - Kokeilun onnistuminen
- Kirjallisuutta ja lisätietoja



Arviointiohjeiden tarkoitus ja perusta

Tekoälyn kokeiluhankkeissa on tärkeää pyrkiä arvioimaan tekoälyn käytöllä saatuja vaikutuksia ja vaikuttavuutta suhteessa kokeilulle asetettuihin tavoitteisiin. Tässä tiedostossa esitetty arviointiohjeistus toimii tiiviinä tukimateriaalina sosiaali- ja terveysministeriön osarahoittamille SOTE-tekoälykokeiluille¹.

Arviointiohjeet laadittiin THL:n tutkimusryhmän asiantuntemuksen sekä aiemman kansainvälisen ja kansallisen kirjallisuuden pohjalta. Koska tekoälyn vaikutuksia, vaikuttavuutta ja käyttöönoton onnistumista arvioivaa tutkimusta on vielä tehty niukasti²⁻⁴, ohjeet perustuvat tutkimusryhmän soveltamaan tietoon niistä arvioinnin kohteista, joiden voidaan olettaa olevan keskeisiä tekoälykokeiluille. Yleisiä arviointimenetelmiä on pyritty yksinkertaistamaan niin, että ne soveltuvat myös lyhyiden kokeilujen arviointiin. Vahvan tutkimusnäytön osoittaminen edellyttää kuitenkin riittävää resursointia arvioinnin suunnitteluun ja toteuttamiseen.



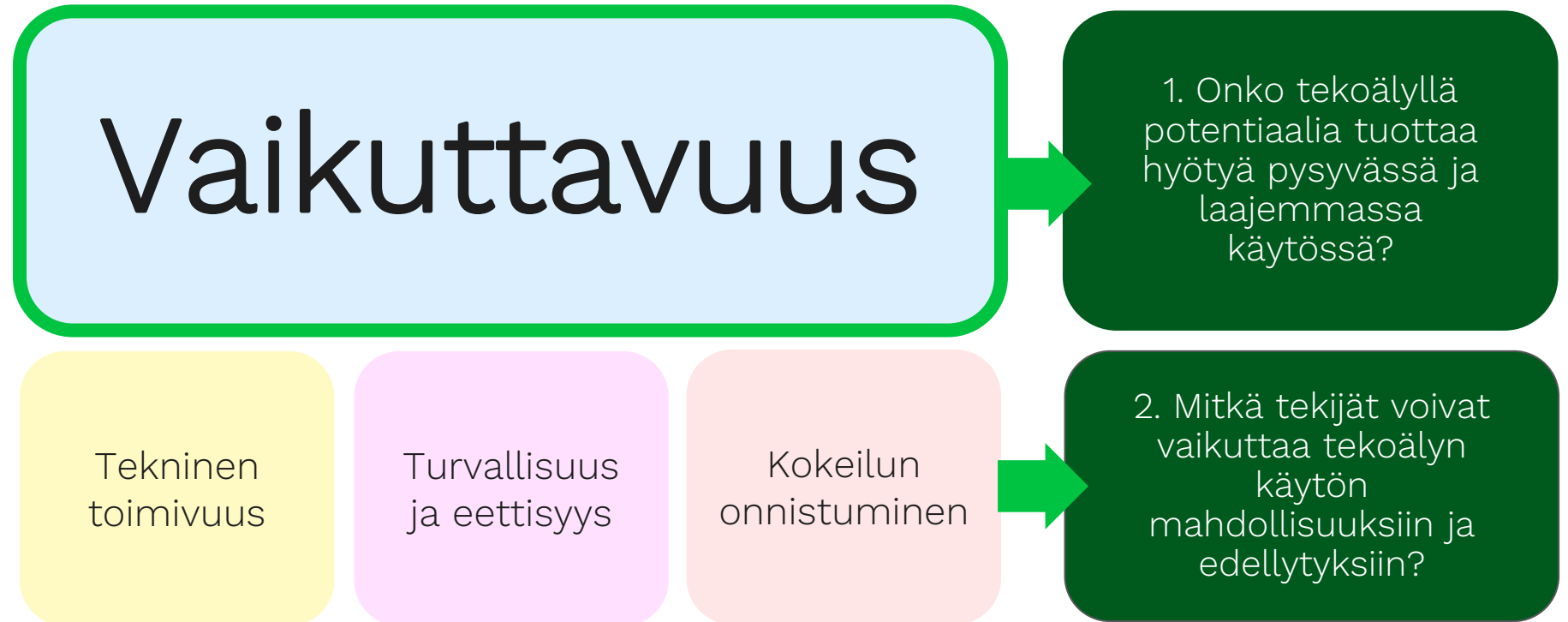
Tiivistelmä SOTE-tekoälykokeilujen arviointiohjeista

Tehtävä	Arviointityyppi	Arviointimenetelmät	Arvioitavat indikaattorit	Arvioinnin ajankohta
Kokeilun ensisijaisena tehtävänä on tuottaa tietoa tekoälyn vaikuttavuudesta, jotta voidaan selvittää pysyvän ja laajemman käyttöönoton potentiaaliset hyödyt.	1. Vaikuttavuuden arviointi	Mahdolliset tutkimusasetelmat tutkimusnäytön vahvuuden mukaisessa järjestyksessä: A. Satunnaistettu kontrolloitu tutkimus (vahvin näyttö), B. Kontrolloitu tutkimus, C. Ennen-jälkeen-tutkimus, ja D. Jälkeen-tutkimus (heikoin näyttö). Tiedonkeruumenetelminä kyselyt, rekisteritiedot, aikamittaus ja/tai haastattelut.	Vähintään yksi keskeinen vaikuttavuusindikaattori sekä noin 1–2 muuta seurattavaa indikaattoria, jotka yhdessä kuvaavat toiminnan tehokkuutta ja sujuvuutta, palveluiden laatua ja/tai asiakkaiden tai potilaiden terveyttä ja hyvinvointia.	Tutkimusasetelmasta riippuen joko ennen kokeilua ja kokeilun loppuvaiheessa (suositeltu) tai ainoastaan kokeilun loppuvaiheessa.
Toissijaisena tehtävänä on arvioida muiden tekijöiden vaikutuksia tekoälyn käytön mahdollisuuksiin ja edellytyksiin.	2. Muu arviointi			
	2.1 Tekninen toimivuus	Lyhyet palautekyselyt tai haastattelut ja suorituskykyanalyysit	Tekoälyjärjestelmän suorituskykyä, johdonmukaisuutta, käytettävyyttä ja vakautta kuvaavat indikaattorit	Säännöllisesti kokeilun aikana
	2.2 Turvallisuus ja eettisyys	Lyhyet palautekyselyt tai haastattelut, haittatapahtumien seuranta, asiantuntija-arviot ja riskinarviointi	Tietosuojaa ja tietoturvaa, oikeudenmukaisuutta sekä päätösten läpinäkyvyyttä ja autonomiaa kuvaavat indikaattorit	Säännöllisesti kokeilun aikana perustuen ennen kokeilua tehtyyn riskienhallinta-suunnitelmaan
	2.3 Kokeilun onnistuminen	Palautekyselyt tai haastattelut sekä rekisteritietojen ja asiakirjojen analyysit	Tekoälyn käytön hyväksyttävyyttä ja omaksumista, toteutettavuutta, kokeilun suunnitelmanmukaisuutta ja kustannuksia kuvaavat indikaattorit	Kokeilun aikana



Tekoälykokeilun vaikuttavuuden ja muiden tekijöiden arviointi

- Kattava arvioiminen on tärkeää tekoälyn vaikutusten ja vaikuttavuuden selvittämiseksi sosiaali- ja terveydenhuollon todellisissa olosuhteissa²⁻¹⁹.
- Arviointimenetelmät esitellään tarkemmin seuraavilla sivuilla.





1.

Vaikutta- vuuden arviointi

Onko tekoälyllä
potentiaalia
tuottaa hyötyä
pysyvässä ja
laajemmassa
käytössä?



Vaikuttavuuden määritelmä

Vaikuttavuudella tarkoitetaan SOTE-tekoälyn välillistä tai välitöntä kykyä tuottaa sosiaali- ja terveydenhuollon todellisissa olosuhteissa merkittävää parannusta:

1. toiminnan tehokkuudessa ja sujuvuudessa,
2. palveluiden laadussa, ja/tai
3. asiakkaiden tai potilaiden terveydessä ja hyvinvoinnissa^{6,10,20,21}.



Vaikuttavuus- arvioinnin suunnittelu

Tekoälyn vaikutusten ja vaikuttavuuden arviointia voidaan helpottaa sitomalla mittaaminen voimakkaammin tavoitteisiin^{6,20}.

1. Arvioinnin suunnittelu alkaa tekoälyratkaisun **vaikuttavuustavoitteiden** eli tavoiteltujen muutosten ja parannusten asettamisesta.
2. Asetettujen tavoitteiden perusteella määritellään, millaisella **tutkimusasetelmalla** tavoitteiden toteutumista eli **vaikuttavuutta** voidaan arvioida.



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

Mihin sosiaali- ja tai terveydenhuoltoon liittyvään ongelmaan tekoälyratkaisu pyrkii vastaamaan?



Millaisia vaikutuksia tekoälyratkaisulla tavoitellaan?



Mitkä ovat tekoälykokeilun tarkat, mitattavat, kokeilujakson aikana saavutettavat, olennaiset ja aikasidonnaiset tavoitteet (engl. SMART goals)?



Millainen tutkimusasetelma ja mitkä tiedonkeruun ja analyysin menetelmät soveltuvat tavoitteiden arviointiin?

Tutkimusasetelma vaikuttaa tutkimusnäytön vahvuuteen

Etusijalla tekoälykokeilujen toteuttamisessa pitäisi olla sellaiset kokeilut, joissa vaikuttavuus pystytään vahvimmin todentamaan.

Vahvin näyttö ←

→ Heikoin näyttö

	A. Satunnaistettu kontrolloitu tutkimus (RCT)	B. Kontrolloitu tutkimus	C. Ennen–jälkeen-tutkimus	D. Jälkeen-tutkimus
Esimerkki toteutuksesta	Puolet yksikön hoitajista satunnaistetaan käyttämään tekoälyä, ja toiset jatkavat tavanomaista toimintaansa. Työkuormaa tai muuta valittua indikaattoria mitataan kaikilta yksikön hoitajilta ennen kokeilua ja kokeilun/vertailuryhmän seurannan loppuvaiheessa. Tuloksia verrataan tekoälyä käyttäneiden ja tavanomaista toimintaa jatkaneiden välillä.	Kaikki yksikön hoitajat ottavat tekoälyn käyttöön. Vertailu tehdään toiseen yksikköön, joka jatkaa tavanomaista toimintaansa. Työkuormaa tai muuta valittua indikaattoria mitataan molempien yksiköiden hoitajilta ennen kokeilua ja kokeilun/vertailuryhmän seurannan loppuvaiheessa. Tuloksia verrataan tekoälyä käyttäneiden ja tavanomaista toimintaa jatkaneiden välillä.	Kaikki yksikön hoitajat ottavat tekoälyn käyttöön, ja työkuormaa tai muuta valittua indikaattoria mitataan vain tässä yksikössä ennen kokeilun aloittamista ja sen jälkeen kokeilun loppuvaiheessa. Kokeilun loppuvaiheessa saatuja tuloksia verrataan ennen kokeilua saatuihin tuloksiin yksikössä.	Kaikki yksikön hoitajat ottavat tekoälyn käyttöön. Työkuormaa tai muuta valittua indikaattoria mitataan vasta kokeilun loppuvaiheessa arvioiden koettua muutosta tekoälyn käytön myötä. Vertailua aiempiin tuloksiin tai toiseen ryhmään ei tehdä.
Hyödyt	Vähentää merkittävästi virheiden ja vääristymien riskiä, mikä tekee siitä parhaan tutkimusasetelman tekoälyn todellisen vaikuttavuuden arvioimiseksi: <ul style="list-style-type: none"> Satunnaistaminen minimoi sekoittavien tekijöiden vaikutuksen, mikä parantaa tulosten luotettavuutta ja yleistettävyyttä. Mahdollistaa selkeän vertailun tekoälyn käytön ja tavanomaisen toiminnan välillä. 	Helpompi toteuttaa, koska ryhmien satunnaistamista ei tarvita. Vertailuryhmän olemassaolo kuitenkin mahdollistaa arvioinnin siitä, johtuuko ero tekoälystä.	Helppo toteuttaa, koska vertailuryhmää ei tarvita. Tarjoaa tietoa siitä, miten tekoäly vaikuttaa kokeiluun osallistuneessa ryhmässä.	Helppo ja nopea toteuttaa, koska tiedonkeruu toteutetaan vain kerran kokeilun jälkeen.
Haitat	Vaatii suurempia resursseja ja tarkkaa suunnittelua, erityisesti satunnaistamisen vuoksi, ja siksi toteuttamiseen voi liittyä käytännön haasteita tai myös eettisiä ongelmia.	Ilman satunnaistamista voi olla vaikea varmistaa, että ryhmät ovat alun perin vertailukelpoisia. Mahdolliset erot kokeilu- ja vertailuryhmän välillä voivat vääristää tuloksia.	Ei voida erottaa, johtuvatko muutokset todella tekoälystä vai muista ulkoisista tekijöistä, eivätkä tulokset välttämättä ole yleistettävissä.	Mahdollinen muutos perustuu kokeiluryhmän subjektiiviseen näkemykseen, johon voi vaikuttaa esim. asenteet ja odotukset. Tämä heikentää huomattavasti tulosten luotettavuutta.

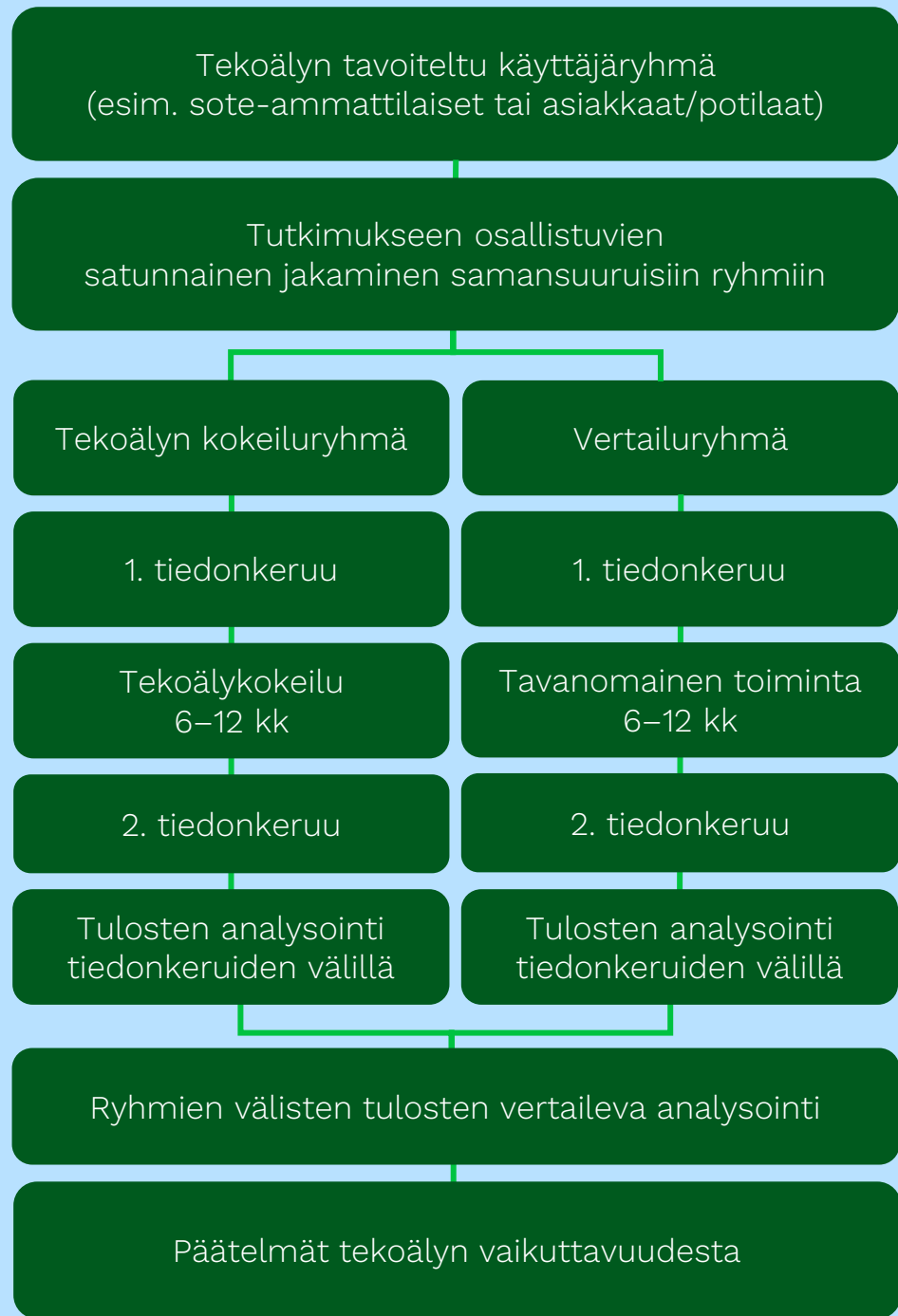


A. RCT

- Tavoitelluin asetelma tekoälyn vaikuttavuuden arviointiin^{10,17}.
- Seurataan vähintään 60–120 henkilöä, joista satunnaisella jaolla 30–60 aloittaa tekoälyn kokeilun ja samansuuruinen vertailuryhmä jatkaa tavanomaista toimintaa. Suuremmat ryhmät on suositeltavia. Satunnaistaminen voidaan toteuttaa yksilö-, tiimi-, yksikkö- tai organisaatiotasolla.
- Tiedonkeruut toteutetaan molemmissa ryhmissä ensin lähtötilanteessa ja sitten tekoälykokeilun ja vertailuryhmässä tavanomaisen toiminnan seurannan päätyttyä.
- Muutosta vaikuttavuusindikaattoreissa analysoidaan molempien ryhmien sisällä ja välillä.



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

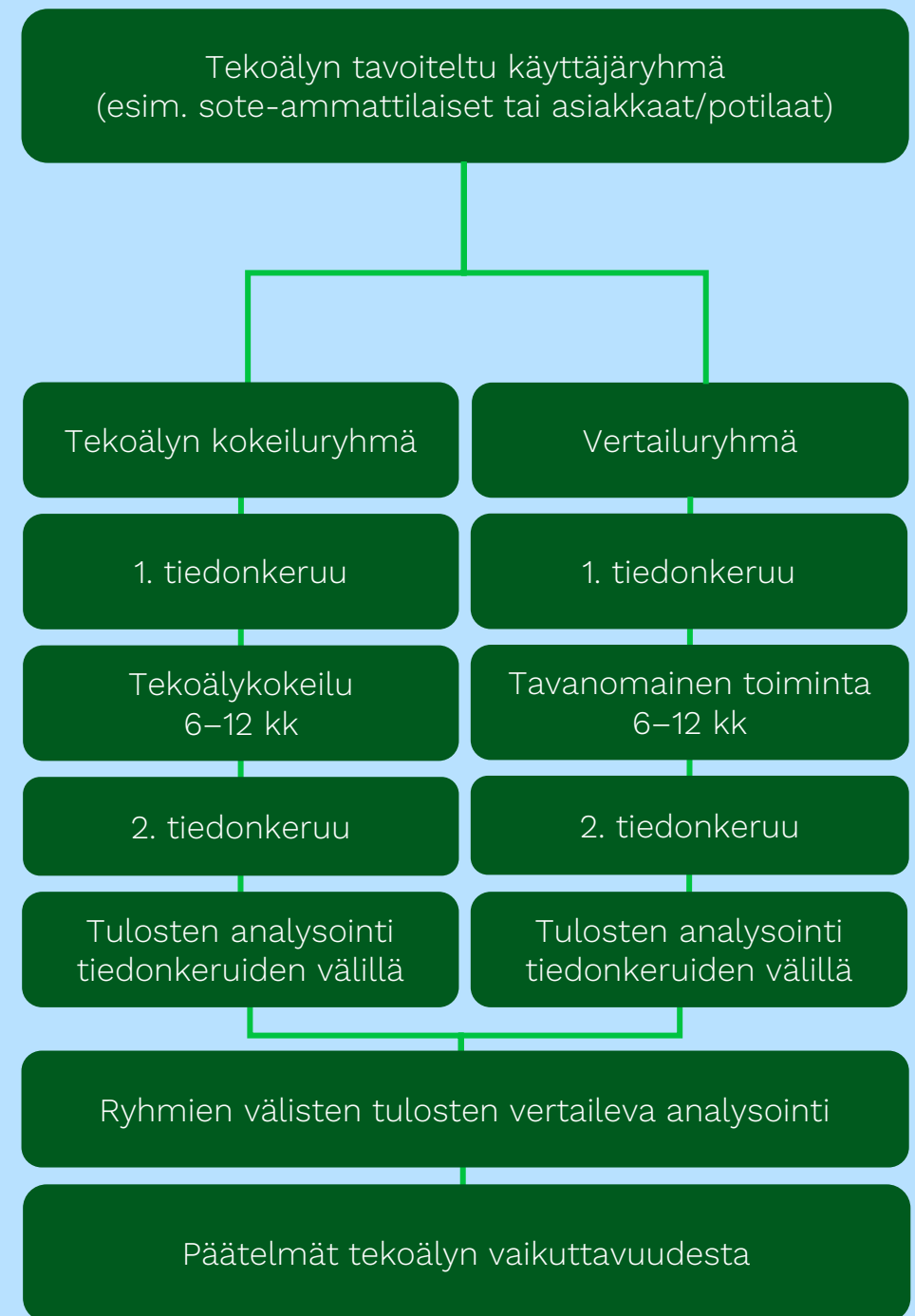


B. Kontrolloitu tutkimus

- Hyödyllinen asetelma tekoälyn vaikuttavuuden arvioinnissa, mikäli satunnaistaminen ei ole mahdollista^{10,17}.
- Seurataan vähintään 30–60 henkilön suuruista tekoälyn kokeiluryhmää. Valitaan samansuuruinen tavanomaisen toiminnan vertailuryhmä. Suuremmat ryhmät on suositeltavia.
- Tiedonkeruut toteutetaan molemmissa ryhmissä ensin lähtötilanteessa ja sitten tekoälykokeilun ja vertailuryhmässä tavanomaisen toiminnan seurannan päätyttyä.
- Muutosta vaikuttavuusindikaattoreissa analysoidaan molempien ryhmien sisällä ja välillä.



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

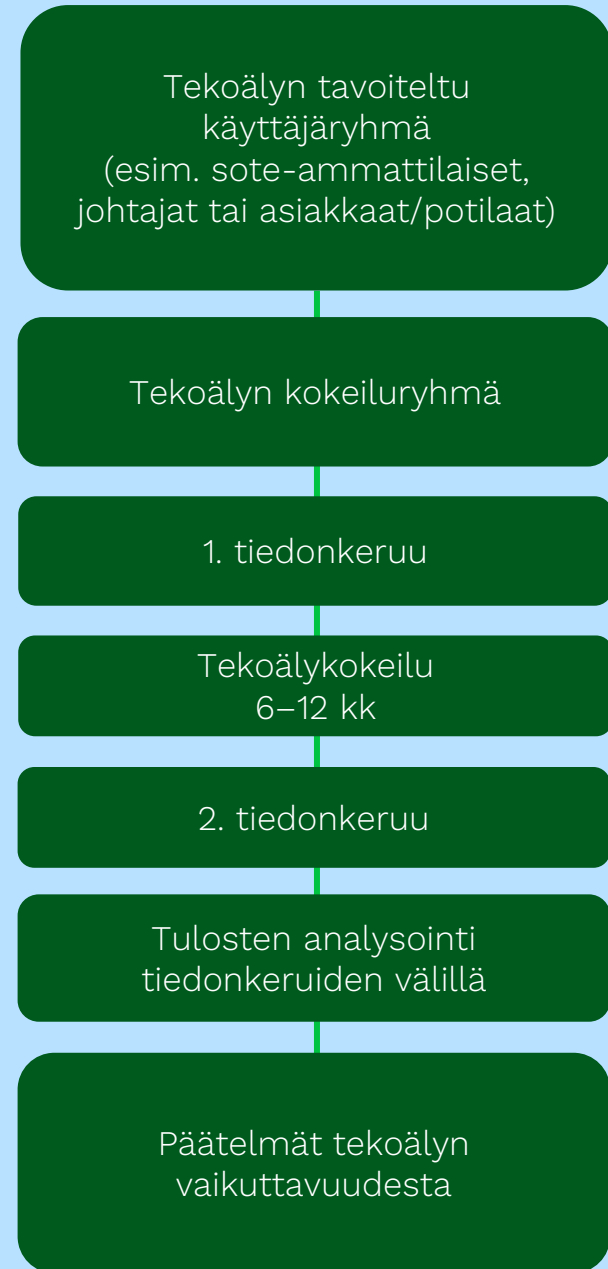


C. Ennen–jälkeen -tutkimus

- Sopii tekoälyn vaikuttavuuden arviointiin, mutta edellyttää toimenpiteitä kuten sekoittavien tekijöiden hallintaa tulosten luotettavuuden parantamiseksi^{10,17}.
- Vähintään 20–30 tekoälykokeiluun osallistuvaa henkilöä seurataan, mutta suurempi otos on parempi.
- Muutosta vaikuttavuusindikaattoreissa analysoidaan samassa kokeiluun osallistuvassa ryhmässä keräämällä tietoa ennen tekoälykokeilun aloitusta sekä kokeilun loppuvaiheessa.



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

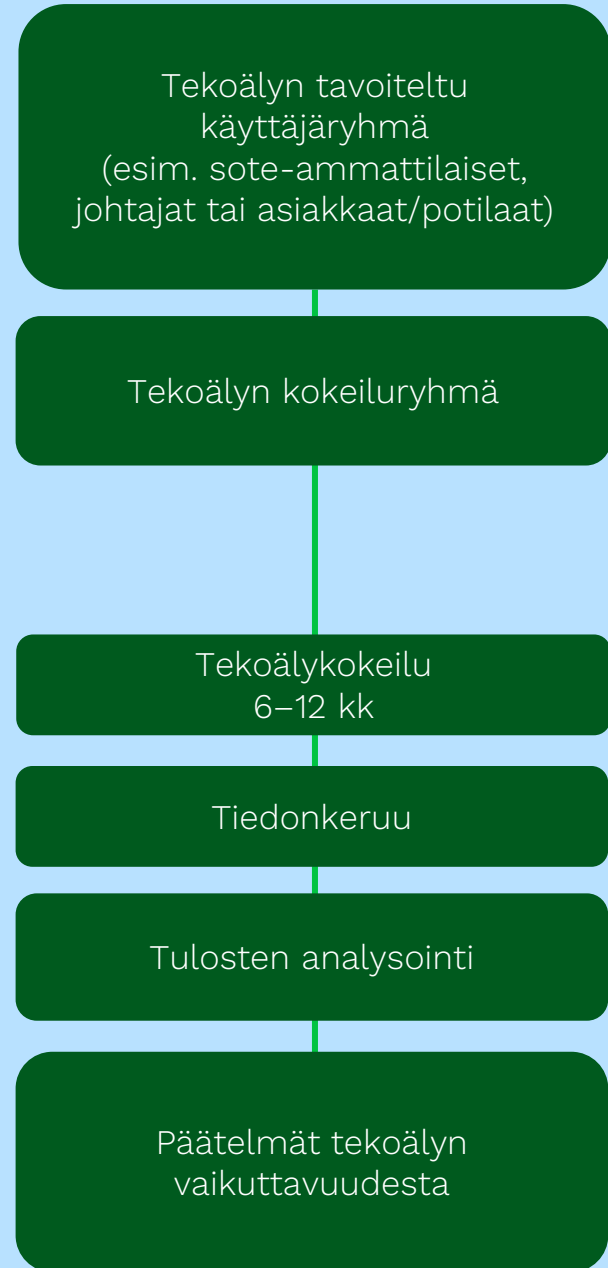


D. Jälkeen -tutkimus

- Sopii myös tekoälyn vaikuttavuuden arviointiin, mutta asetelmalla on huomattavasti heikompi luotettavuus edeltäviin tutkimusasetelmiin verrattuna¹⁰. Sekoittavat tekijät on huomioitava luotettavuuden parantamiseksi.
- Vaikutusten tarkastelua jälkikäteen voidaan tehdä esimerkiksi selvittämällä tekoälykokeiluun osallistuneilta henkilöiltä koettua muutosta vaikuttavuusindikaattoreissa kokeilun loppuvaiheessa tai heti sen päätyttyä (vähintään 20–50 henkilön kokemukset, mutta suurempi otos on parempi).



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).



Indikaattorien valitseminen vaikuttavuuden arviointiin

- On suositeltavaa, että tekoälyn vaikuttavuutta arvioidaan olennaisten käyttäjäryhmien sekä sellaisten toimintojen, palveluiden tai ei-käyttäjäryhmien osalta, joihin tekoälyn käyttö voi heijastua.

Vaikuttavuusindikaattorit^{6,10,14,17,20,21,22}

Toiminnan
tehokkuus ja
sujuvuus

Palveluiden laatu

Asiakkaiden/
potilaiden terveys ja
hyvinvointi

1. Määriteltävä vähintään yksi keskeinen tekoälyn vaikuttavuutta mittaava indikaattori, jonka osalta tavoiteltua muutosta arvioidaan kokeilun aikana.

2. Määriteltävä noin 1–2 muuta seurattavaa indikaattoria, jossa ei välttämättä tavoitella muutosta vaan sen voidaan odottaa pysyvän kokeilun aikana vähintään samalla tasolla kuin tavanomaisessa toiminnassa.



Toiminnan tehokkuuden ja sujuvuuden indikaattorit

Tekoälyn mahdollisten vaikutusten kannalta keskeisiä sosiaali- ja terveydenhuollon toimintaa kuvaavia indikaattoreita voivat olla^{6,10,14,17,20,21}:

- a. **Prosessien nopeus:** Kuinka nopeasti palvelut, hoitotoimenpiteet tai hallinnolliset tehtävät toteutetaan?
- b. **Henkilöstöressurit:** Minkä verran henkilöstöä tarvitaan tehtävien suorittamiseen?
- c. **Tuottavuus:** Kuinka paljon palveluita tai hoitoa pystytään tarjoamaan suhteessa käytettävissä oleviin resursseihin?
- d. **Työn kulku:** Kuinka sujuvasti työtehtävistä suoriudutaan?
- e. **Työkuorma:** Esiintyykö työssä jatkuvaa painetta esimerkiksi keskeneräisten töiden vuoksi, ja onko työtehtävien asianmukainen suorittaminen hankalaa annetussa työajassa?



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

Tiedonkeruun esimerkkejä:

- **Aikamittaus:** Pyydetään henkilöstöä kirjaamaan ylös käytettyä aikaa [valittuihin työtehtäviin tai palvelun toteuttamiseen]
- **Organisaation rekisteritiedot**
 - Keskimääräinen [valitun palvelun] käsittelyaika per asiakas [valitussa aikayksikössä]
 - Asiakasmäärät työntekijää kohden [valitussa palvelussa] [valitussa aikayksikössä]
- **Lyhyt kysely tai haastattelut**
 - Ammattilaisten kokemukset työn kulusta tai työkuormasta
 - Johtamistehtävissä olevien näkemykset tuottavuudesta

Palveluiden laadun indikaattorit

Tekoälyn mahdollisten vaikutusten kannalta keskeisiä palveluiden laatua kuvaavia indikaattoreita voivat olla^{6,10,14,17,20,21,22}:

- a. **Näyttöön perustuvat palvelut:** Kuinka hyvin hoito- ja toimintakäytännöt perustuvat parhaan saatavilla olevan tieteellisen tiedon harkittuun käyttöön?
- b. **Asiakas- ja potilasturvallisuus:** Kuinka hyvin palvelut toteutuvat ilman asiakkaille aiheutuvia riskejä tai haittoja?
- c. **Saatavuus ja saavutettavuus:** Kuinka hyvin tarpeenmukaisia palveluita on tarjolla oikea-aikaisesti oikeassa paikassa? Onko niiden [käyttö asiakkaalle mahdollista](#)?

Laatua voidaan myös mitata hoitotuloksilla tai asiakastavoitteiden saavuttamisella. Nämä tekijät löytyvät terveyden ja hyvinvoinnin indikaattoreista seuraavalta sivulta.



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

Tiedonkeruun esimerkkejä:

- Lyhyt kysely tai haastattelut
 - Hoitotyön tutkimussäätiön jatkokehittämät mittarit johdolle ja/tai ammattilaisille näyttöön perustuvan toiminnan [toteutumisesta](#), [osaamisesta](#) ja [asenteista](#)
 - [Asiakaspalautekysely](#)
- Organisaation rekisteritiedot
 - Raportoitujen asiakas- ja potilasturvallisuuteen liittyvien häirtatapahtumien määrä [valitussa palvelussa] [valittuna ajanjaksona] (HaiPro-raportit)
 - Perusterveydenhuollon avosairaanhoidon lääkärin kiireettömän käynnin odotusaika yli [valittu aikamääre] hoidon tarpeen arvioinnista, % toteutuneista käynneistä [valitun ajanjakson aikana]

Asiakkaiden tai potilaiden terveyden ja hyvinvoinnin indikaattorit

Tekoälyn mahdollisten vaikutusten kannalta keskeisiä terveyden ja hyvinvoinnin indikaattoreita voivat olla^{6,10,14,17,20,21}:

- a. **Hoitotulokset:** Kuinka hyvin hoidossa tai palveluissa on saavutettu toivotut terveysvaikutukset, kuten oireiden lievittyminen, sairauden paraneminen tai hoitotasapainon ylläpitäminen?
- b. **Asiakastavoitteiden saavuttaminen:** Kuinka hyvin sosiaalihuollon asiakkaat ovat saavuttaneet yhdessä sovitut henkilökohtaiset tavoitteet?
- c. **Fyysinen terveys, henkinen hyvinvointi ja toimintakyky:** Millainen on asiakkaiden a) yleinen terveydentila tai miten kroonisen sairauden hallinta sujuu; b) psyykinen hyvinvointi ja mielenterveys; tai c) toimintakyky arjen toiminnoista selviytymisessä?



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

Tiedonkeruun esimerkkejä:

- Lyhyt kysely
 - Asiakkaiden itsearviointi vointimittareilla (engl. PROM), esim. yleinen vointimittari [EQ-5D](#) tai toimintakyvyn mittari [WHODAS 2.0](#)
- Organisaation rekisteritiedot
 - Hoitotulokset tai asiakastavoitteiden toteutuminen
 - [Valitun] vastaanottopalvelun käytön määrä [valitun ajanjakson aikana]

Harhojen ja vääristymien välttäminen

- **Sekoittavat tekijät:** Tietoa on kerättävä osallistujien taustatekijöistä (esim. ikä, koulutus) ja ulkoisista tekijöistä (esim. organisatoriset tekijät), jotka voivat vääristää tuloksia. Nämä tekijät on kontrolloitava tilastollisin menetelmin tai huomioitava muuten tulosten tulkinnassa.
- **Tietolähteiden luotettavuus:** Osallistujien kokemuksiin perustuvat mittaukset ovat alttiita vääristymille, kuten odotuksille ja ennakkoasenteille, mikä on otettava huomioon tuloksia tulkitessa.
- **Tiedonkeruun menetelmät ja ajoitus:** Tiedonkeruut on toteutettava johdonmukaisesti samoilla menetelmillä ja indikaattoreilla sekä samanaikaisesti eri ryhmissä vertailukelpoisuuden varmistamiseksi. Toinen tiedonkeruu on ajoitettava kokeilun loppuun tai heti sen päätyttyä, jotta vaikutuksista saadaan tarkkaa tietoa ennen kokeiluryhmän palaamista tavanomaiseen toimintaan.





2.

Muu arviointi

Mitkä tekijät voivat vaikuttaa tekoälyn käytön mahdollisuuksiin ja edellytyksiin?

Tekninen toimivuus

Turvallisuus ja eettisyys

Kokeilun onnistuminen



2.1

Tekninen
toimivuus



Teknisen toimivuuden indikaattorit

Tekoälyn teknisen toimivuuden selvittämiseksi on suositeltavaa arvioida kokeilun aikana seuraavia tekijöitä, jotta mahdollisiin epäkohtiin voidaan puuttua^{10,12,14,18}:

1. **Suorituskyky:** Kuinka hyvin tekoälyjärjestelmä suoriutuu tehtävistään? Tuottaako se käytännössä hyödyllisiä tuloksia? Kuinka usein tekoälyn tuottamia ehdotuksia otetaan käyttöön?
2. **Johdonmukaisuus:** Ovatko tekoälyn tekemät päätökset tai ennusteet johdonmukaisia ja luotettavia?
3. **Käytettävyys:** Kuinka helppokäyttöinen tekoälyjärjestelmä on sosiaali- ja terveydenhuollon käytännön työssä tai asiointissa?
4. **Vakaus:** Kuinka usein järjestelmässä ilmenee teknisiä häiriöitä kuten kaatumisia? Millaisia vaikutuksia häiriöillä on ollut?



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

Tiedonkeruun esimerkkejä:

- Lyhyet kyselyt tai haastattelut jatkuvan käyttäjäpalautteen keräämiseksi
- Tekoälyjärjestelmän **lokitietojen ja suorituskykyraporttien** analysoiminen kokeilun aikana yhteistyössä teknisen kehittäjän kanssa
- [CLEAR –mittari](#) tekoälyn tuottaman terveystiedon laadun arvioimiseksi



2.2

Turvallisuus ja eettisyys



Turvallisuuden ja eettisyyden indikaattorit

Tekoälyn turvallisuuden ja eettisyyden määrittelemiseksi on suositeltavaa arvioida kokeilun aikana seuraavia tekijöitä, jotta mahdollisiin epäkohtiin voidaan puuttua^{3,5,10–14,17,19,23}:

1. **Tietosuoja ja tietoturva:** Onko kokeilun aikana havaittu tietoturvaloukkauksia tai -riskejä? Miten niihin on reagoitu ja varmistettu, etteivät vastaavat tilanteet toistu?
2. **Oikeudenmukaisuus:** Onko tekoälyn päätöksissä tai ennusteissa havaittu ennakkoluuloja tai syrjiviä käytäntöjä?
3. **Läpinäkyvyys ja autonomia:** Ymmärtävätkö käyttäjät tekoälyn päätösten tai ennustusten perusteet? Voiko niihin tarvittaessa pyytää lisäselvitystä? Kokevatko käyttäjät, että heillä on mahdollisuus kyseenalaistaa päätöksiä ja tehdä päätökset itse?

Euroopan tietosuojaneuvoston asiantuntijoiden tukiryhmässä julkaistussa [AI auditing -tarkistuslistassa](#) on tarkempia kysymyksiä, jotka auttavat arvioimaan tiedonhallinnan (s. 13–15), prosessien jäljitettävyyden ja tietoturvan (s. 17–18), oikeudenmukaisuuden (s. 18–19) sekä läpinäkyvyyden (s. 8) yhteensopivuutta yleisen tietosuoja-asetuksen kanssa²⁴.



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

Tiedonkeruun esimerkkejä:

- Tietosuojaan, tietoturvaan tai eettisyyteen liittyvien **haittatapahtumien tai ihmisen väliintulojen säännöllinen seuraaminen** kokeilun aikana perustuen ennen kokeilua tehtyyn riskienhallintasuunnitelmaan
- **Lyhyet kyselyt tai haastattelut** jatkuvan käyttäjäpalautteen keräämiseksi
- **Asiantuntija-arviot:** Lainsäädännön ja etiikan asiantuntijoiden säännölliset tarkastelut kokeilun aikana



2.3

Kokeilun
onnistuminen



Kokeilun onnistumisen indikaattorit

Kokeilun onnistumista voidaan arvioida jo kokeilun aikana, jolloin mahdollisiin epäkohtiin voidaan puuttua. Arvioinnissa on suositeltavaa pohtia seuraavia näkökulmia^{7,8,12,14}:

1. **Tekoälyn hyväksyttävyys ja omaksuminen:** Kuinka hyvin käyttäjät ovat vastaanottaneet tekoälyn? Kuinka laajasti ja hyvin tekoäly on omaksuttu käyttöön kokeiluun osallistuvissa toiminnossa, palveluissa tai käyttäjäryhmissä?
2. **Toteutettavuus:** Kuinka sujuvasti tekoälyn käyttöönotto on toteutunut? Esim. onko käyttäjille tarjottu koulutus ja tuki ollut riittävää, ovatko tekniikka ja resurssit toimineet suunnitellusti, ja kuinka hyvin mahdolliset ongelmat on saatu ratkaistua?
3. **Kokeilun suunnitelmanmukaisuus:** Kuinka tarkasti tekoälyn käyttöönotto on noudattanut alkuperäisiä suunnitelmia?
4. **Kustannukset:** Kuinka hyvin tekoälyn käyttöönotto on pysynyt asetetussa budjetissa?



Virtanen, L., Kortelainen, M., & Heponiemi, T. (2024).

Tiedonkeruun esimerkkejä:

- Kyselyt tai haastattelut käyttäjäryhmille ja/tai mahdollisesti myös johdolle tekoälyn käytön hyväksyttävyyden ja omaksumisen sekä toteutettavuuden arvioimiseen
- Käyttötilastot ja käytön seuranta omaksumisen laajuuden arvioimiseen
- Projektin aikataulujen ja resurssien tarkastelu toteutettavuuden ja uskollisuuden arvioimiseksi
- Budjetin seuranta kokeilun toteutuneiden kustannusten määrittelymiseksi

Kirjallisuutta

1. Lähesmaa ym. (2024). [Suunnitelma SOTE-tekoälyn hyödyntämisen ekosysteemistä](#). STM.
2. Sanmark, J., & Sanmark, E. (2024). [Mitä tiedämme generatiivisen tekoälyn hyödyistä terveydenhuollossa?](#) *Lääketieteellinen Aikakauskirja Duodecim*, 140(12), 1023–1030.
3. Sharma, M. ym. (2022). [Artificial Intelligence Applications in Health Care Practice: Scoping Review](#). *Journal of medical Internet research*, 24(10), e40238.
4. Gomez Rossi, J. ym. (2022). [Evaluation of the Clinical, Technical, and Financial Aspects of Cost-Effectiveness Analysis of Artificial Intelligence in Medicine: Scoping Review and Framework of Analysis](#). *JMIR medical informatics*, 10(8), e33703.
5. Haverinen, J. ym. (2019). [Digi-HTA: Health technology assessment framework for digital healthcare services](#). *Finnish Journal of EHealth and EWelfare*, 11(4), 326–341.
6. Koivisto, J. (2021). [Esiselvitys sosiaali- ja terveydenhuollon kansallisten digitalisaatio-ohjelmien arviointikehikon kehittämiseksi](#). Työpaperi 28/2021. THL.
7. Smith, J. D., Li, D. H., & Rafferty, M. R. (2020). [The Implementation Research Logic Model: a method for planning, executing, reporting, and synthesizing implementation projects](#). *Implementation science*, 15(1), 84.
8. Pearson, N. ym. (2020). [Guidance for conducting feasibility and pilot studies for implementation trials](#). *Pilot and feasibility studies*, 6(1), 167.
9. Curran, G. M. ym. (2012). [Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact](#). *Medical care*, 50(3), 217–226.
10. Jin, M. F., Noseworthy, P. A., & Yao, X. (2024). [Assessing AI solution effectiveness: The role of pragmatic trials](#). *Mayo Clinic Proceedings: Digital Health*. Advance online publication.
11. Gama, F. ym. (2022). [Implementation Frameworks for Artificial Intelligence Translation Into Health Care Practice: Scoping Review](#). *Journal of medical Internet research*, 24(1), e32215.
12. Truong, T. ym. (2019). [A Framework for Applied AI in Healthcare](#). *Studies in health technology and informatics*, 264, 1993–1994.
13. Stahl, B. C. ym. (2023). [A systematic review of artificial intelligence impact assessments](#). *Artificial intelligence review*, 1–33.
14. Reddy, S. ym. (2021). [Evaluation framework to guide implementation of AI systems into healthcare settings](#). *BMJ health & care informatics*, 28(1).
15. Magrabi, F. ym. (2019). [Artificial Intelligence in Clinical Decision Support: Challenges for Evaluating AI and Practical Implications](#). *Yearbook of medical informatics*, 28(1), 128–134.
16. Park, Y. ym. (2020). [Evaluating artificial intelligence in medicine: phases of clinical research](#). *JAMIA open*, 3(3), 326–331.
17. de Hond, A. A. H. ym. (2022). [Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review](#). *NPJ digital medicine*, 5(1), 2.
18. Abbasian, M. ym. (2024). [Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI](#). *NPJ digital medicine*, 7(1), 82.
19. Amann, J. ym. (2020). [Explainability for artificial intelligence in healthcare: a multidisciplinary perspective](#). *BMC medical informatics and decision making*, 20(1), 310.
20. Pennanen, P. ym. (2023). [Digitaalisten palvelujen vaikutukset sosiaali- ja terveydenhuollossa](#). Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2023:52. Valtioneuvosto.
21. Bodenheimer, T., & Sinsky, C. (2014). [From triple to quadruple aim: care of the patient requires care of the provider](#). *Annals of family medicine*, 12(6), 573–576.
22. Nilsen, P. ym. (2024). [Towards evidence-based practice 2.0: leveraging artificial intelligence in healthcare](#). *Frontiers in health services*, 4, 1368030.
23. Ojanen, A. ym. (2022). [Algoritminen syrjintä ja yhdenvertaisuuden edistäminen: Arviointikehikko syrjimättömälle tekoälylle](#). Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2022:54. Valtioneuvoston kanslia.
24. Galdon Clavell, G. (2023). [AI Auditing: Checklist for AI Auditing](#). Support Pool of Experts Programme. Euroopan tietosuojaneuvosto.



Lisätietoja

tutkija Lotta Virtanen
tutkimusprofessori Mika Kortelainen
tutkimusprofessori Tarja Heponiemi

hyvinvointivaltion tutkimus –yksikkö
Terveyden ja hyvinvoinnin laitos

etunimi.sukunimi@thl.fi