SOSIAALI- JA TERVEYSPALVELUJEN KUSTANNUSVAIKUTTAVUUDEN EDISTÄMINEN DATA-ANALYTIIKAN AVULLA

Lappeenrannan–Lahden teknillinen yliopisto LUT

Tuotantotalouden kandidaatintutkielma

2023

Saara Säkäjärvi

Tarkastaja: Tutkijaopettaja Jouni Koivuniemi

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT Teknis-luonnontieteellinen

Tuotantotalous

Saara Säkäjärvi

Sosiaali- ja terveyspalvelujen kustannusvaikuttavuuden edistäminen data-analytiikan avulla

Tuotantotalouden kandidaatintyö

2023

48 sivua, 5 kuvaa ja 3 taulukkoa

Tarkastaja: Tutkijaopettaja Jouni Koivuniemi

Avainsanat: kustannusvaikuttavuus, vaikuttavuus, sosiaali- ja terveydenhuolto, data-analytiikka, vaikuttavuusperusteinen terveydenhuolto

Tämän kandidaatintyön tavoite on tutkia data-analytiikan hyödyntämistä sosiaali- ja terveyspalvelujen kustannusvaikuttavuuden edistämisessä. Työn tarkoituksena on perehtyä siihen, millaisilla data-analytiikan keinoilla kustannusvaikuttavuutta voidaan edistää. Työssä perehdytään myös analytiikan prosessiin ja esitellään sosiaali- ja terveydenhuollon datan hyödyntämisen tueksi teoreettinen viitekehys. Työ toteutetaan integroivana kirjallisuuskatsauksena.

Työssä todetaan, että hyvinvointivaltiomme rahoituspohja on heikentynyt, jolloin yhdenvertaiset ja laadukkaat sosiaali- ja terveyspalvelut tulee toteuttaa entistä tehokkaammin ja taloudellisemmin. Kustannusvaikuttavat palvelut tuottavat asiakkaille enemmän hyötyä käytössä olevilla resursseilla. Kustannusvaikuttavuuden mittaamisessa on tärkeää hyödyntää yhtenäisiä mittareita, jotta tulosten vertailukelpoisuus säilyy hyvänä. Kustannusvaikuttavuutta voidaan arvioida viidellä eri tasolla, joista yksittäisen intervention kustannusvaikuttavuus on tarkin ja yhteiskunnallinen vaikuttavuus laajin taso.

Sosiaali- ja terveydenhuollon arjessa asiakkaista sekä potilaista kerätään valtavia määriä tietoa. Kerätyn datan arvo perustuu siitä saatuun hyötyyn. Hyöty puolestaan rakentuu datan jalostamisen eli analytiikan prosessin kautta. Analytiikan prosessin vaiheet ovat kuvaileva, diagnostinen, ennakoiva ja ohjaileva analytiikka. Prosessia voidaan hyödyntää myös rakenteeltaan haastavaan sosiaali- ja terveydenhuollon dataan. Työssä esiteltyjen esimerkkien avulla huomataan, että sosiaali- ja terveyspalvelujen kehittämiseen hyödynnetään jo nyt monipuolisesti data-analytiikan eri menetelmiä. Kustannusvaikuttavuustutkimusta hidastaa kuitenkin se, että yhtenäisten mittareiden käyttö ei ole vielä riittävän kattavaa.

Sisällysluettelo

Tiivistelmä

1 Johdanto ... 3

1.1 Työn tavoite ja tutkimuskysymykset ... 4

1.2 Rajaukset ja tutkimusmenetelmä ... 5

1.3 Työn rakenne ... 5

2 Vaikuttavuus ja kustannusvaikuttavuus sosiaali- ja terveydenhuollossa ... 7

2.1 Vaikuttavuudesta saadaan tietoa tutkimalla ... 7

2.2 Kustannusvaikuttavuus käsitteenä ... 9

2.3 Vaikuttavuuden mittaaminen ja arviointi ... 11

2.4 Vaikuttavuus- ja tulosperusteinen sosiaali- ja terveydenhuolto ... 16

3 Data ja analytiikka ... 19

3.1 Datan ominaisuudet ja big data ... 19

3.2 Data-analytiikka ... 22

3.3 Data-analytiikan prosessi ... 23

4 Data-analytiikka kustannusvaikuttavuuden edistämisessä ... 27

4.1 Viitekehys kohti kustannusvaikuttavampaa sosiaali- ja terveydenhuoltoa ... 27

4.2 Sosiaali- ja terveydenhuollon data ja sen hyödyntämisen haasteet ... 29

4.3 Datasta päätöksentekoon ... 31

4.4 Esimerkkejä data-analytiikan avulla tuotetusta kustannusvaikuttavuudesta ... 33

5 Johtopäätökset ... 36

Lähteet ... 41

1 Johdanto

Sosiaali- ja terveysministeriön selvityksessä todettiin, että Suomessa julkaistaan vuosittain 10–20 vertaisarvioitua terveystaloustieteellistä arviointia, joista valtaosa käsittelee lääkkeiden kustannusvaikuttavuutta. (Booth, Aronen & Mäkelä 2017). Selvityksessä tarkasteltiin artikkeleiden määrää, kun keskeisessä terveydenhuollon PubMed-viitetietokannassa tehtiin haku hakusanoilla ”cost-effectiveness” ja ”Finland”. Vuonna 2017 näillä hakukriteereillä artikkeleita löytyi 320 ja nyt keväällä 2023 sama haku tuotti 555 tulosta. Suurin osa artikkeleista käsittelee yhä lääkkeiden tai tietyn intervention kustannusvaikuttavuutta, mutta hoitopolkujen ja järjestelmätason kustannusvaikuttavuuden tutkiminen on noussut yhä enemmän esille ja on saanut tilaa myös poliittisessa keskustelussa.

Vaikuttavuus ja kustannusvaikuttavuus on tunnistettu tärkeiksi aiheiksi suomalaisen sosiaali- ja terveydenhuoltojärjestelmän kehittämisessä ja ne on nostettu yhdeksi järjestelmän vuosien 2023–2026 kansallisista tavoitteista. Vaikuttavuuden tavoitteen avulla pyritään varmistamaan että sosiaali- ja terveydenhuollon toiminta on taloudellisesti kestävää ja kustannusten kasvua saadaan hidastettua suhteessa palvelutarpeen kasvuun. Toiminnan painotusta siirretään ennaltaehkäisevään työhön ja palvelujen ohjauksessa pyritään vaikuttavuusperusteisuuteen. Muita tavoitteita ovat palvelujen laadun parantaminen, sosiaali- ja terveydenhuollon yhdistäminen toimivaksi osaksi yhteiskuntaa sekä tiedon ja tutkimuksen merkityksen lisääminen (Sosiaali- ja terveysministeriö 2022, s. 16).

Huoltosuhteen heikkeneminen tuottaa haasteita hyvinvointivaltion rahoituspohjalle, jolloin laadukkaat ja yhdenvertaiset sosiaali- ja terveydenhuollon palvelut tulee pyrkiä tuottamaan tehokkaammin ja yhä taloudellisemmin (Sosiaali- ja terveysministeriö 2022, s. 22). Tietoa kustannuksista ja kustannusvaikuttavuudesta on jo tähän mennessä kerätty pilottihankkeiden avulla suhteellisen paljon, mutta kansallisella tasolla ei ole sopivia rakenteita kerätyn tiedon laajamittaiseen hyödyntämiseen (Pitkänen, Torkki, Tolkki, Valtakari & Leskelä 2020). Kehittyneiden data-analytiikan menetelmien tuottama arvo vaikuttavuuden kannalta on tunnistettu ja sen käyttöönottoa on syytä tarkastella monipuolisesti (Vuorela 2018).

1.1 Työn tavoite ja tutkimuskysymykset

Tämä kandidaatintyö on luotu yllä mainittuun tarpeeseen edistää data-analytiikan hyödyntämistä sosiaali- ja terveyspalvelujen kustannusvaikuttavuuden arvioimisessa. Työn tavoitteena on luoda viitekehys, jonka avulla suomalaista sosiaali- ja terveydenhuoltoa voidaan viedä kohti kustannusvaikuttavampaa suuntaa data-analytiikan prosessin kautta. Tarkoituksena on perehtyä siihen, millaisilla data-analytiikan keinoilla kustannusvaikuttavuutta voidaan arvioida ja edistää. Kandidaatintyö vastaa päätutkimuskysymykseen, joka on:

- *Miten sosiaali- ja terveyspalvelujen kustannusvaikuttavuutta voidaan edistää data-analytiikan avulla?*

Päätutkimuskysymys on kokonaisuutena laaja-alainen, joten se on jaettu kolmeen apututkimuskysymykseen. Apututkimuskysymykset auttavat tutkimusongelman jäsentämisessä ja ne määrittävät työn rakennetta. Niiden avulla pyritään muodostamaan selkeä katsaus työn tärkeimmistä osa-alueista. Työn apututkimuskysymykset ovat:

- *Miten kustannusvaikuttavuus muodostuu sosiaali- ja terveyspalveluissa?*
- *Millainen on yksinkertaistettu data-analytiikan prosessi?*
- *Mitä data-analytiikan menetelmiä voidaan hyödyntää kustannusvaikuttavuuden edistämisessä?*

Kandidaatintyön aihe on monitieteellinen ja se yhdistää niin terveystaloustieteen, tietojenkäsittelytieteen, tietojohtamisen sekä tiedolla johtamisen aihealueita. Laaja-alaisuutta silmällä pitäen kandidaatintyön tavoitteena on luoda selkeä kuvaus siitä mitä vaikuttavuustutkimuksella ja kustannusvaikuttavuustutkimuksella voidaan saada aikaan ja mitä hyötyä siitä on koko yhteiskuntamme kannalta. Data-analytiikkaa ehdotetaan monipuoliseksi työkaluksi, jonka avulla sosiaali- ja terveydenhuollon vaikuttavuutta voidaan edistää.

1.2 Rajaukset ja tutkimusmenetelmä

Työ on rajattu käsittelemään erityisesti terveystaloustieteellistä näkökulmaa eli kustannusvaikuttavuutta. Vaikuttavuutta käsitellään työssä laajempana kokonaisuutena, sillä kustannusvaikuttavuuden muodostumista ei voi lähestyä ymmärtämättä ensin vaikuttavuusajattelua ja -ketjua. Vaikuttavuutta ja kustannusvaikuttavuutta käsiteltäessä perusteiden lisäksi huomio keskittyy yhteiskunnalliselle tasolle ja tarkoituksena on rakentaa käsittelyä aina järjestelmätason vaikuttavuustutkimukseen asti. Sosiaalihuolto on tärkeä osa hyvinvointijärjestelmäämme, joten sitä ei ole rajattu käsittelyn ulkopuolelle. Käsitteiden määrittelyssä, mittaristossa ja yksittäisissä esimerkeissä painottuu lähteiden saatavuuden vuoksi terveydenhuollon osa-alueet. Data-analytiikkaa käsitellään työssä siltä osin kuin sen hahmottaminen kokonaisuutena vaatii. Työssä ei ole tarkoitus perehtyä yksittäisiin menetelmiin sen tarkemmin ja työ jättää tässä kohtaa tilaa jatkotutkimukselle.

Kandidaatintyö toteutetaan integroivana kirjallisuuskatsauksena ja siinä pyritään muodostamaan mahdollisimman laaja-alainen kokonaiskuva aiheesta. Työssä ilmenee kriittinen ajattelu käytettyihin aineistoihin. Työn lähteinä hyödynnetään sekä kotimaista että ulkomaista kirjallisuutta sekä tutkimuksia. Kotimaisessa kirjallisuudessa tärkeä lähde on Antti Malmivaaran Vaikuttavuus sosiaali- ja terveydenhuollossa (2022). Siinä kuvataan ajantasainen ja monipuolinen näkemys vaikuttavuudesta kotimaisessa hyvinvointijärjestelmässä käsitteiden määrittelystä aina strategian luomisen periaatteisiin asti. Työssä hyödynnetään myös sähköisiä lähteitä. Tärkeimmät tietokannat, joita aineiston keruussa hyödynnettiin, olivat LUT Primo, Scopus, Research Gate ja PubMed. Keskeisimpiä hakusanoja olivat vaikuttavuus, kustannusvaikuttavuus, value-based healthcare, cost-effectiveness, descriptive analytics, predictive analytics, diagnostic analytics, prescriptive analytics ja big data.

1.3 Työn rakenne

Työ alkaa johdannolla, jossa avataan aiheen ajankohtaisuutta ja kerrotaan työn tausta, tavoite, tutkimuskysymykset ja työhön tehdyt rajaukset. Työ koostuu viidestä luvusta, joista kolmessa käsitellään työn varsinainen sisältö. Johtopäätökset on työn viimeinen luku.

Ensimmäisessä käsittelyluvussa käydään läpi vaikuttavuuteen ja kustannusvaikuttavuuteen liittyvää teoriaa. Siinä avataan vaikuttavuuteen liittyviä käsitteitä, käydään läpi vaikuttavuustutkimusta sekä mittaristoa vaikuttavuuden ja kustannusvaikuttavuuden arviointiin. Luvun tarkoituksena on antaa kattava kuvaus vaikuttavuuden muodostumisesta sosiaali- ja terveysalalla.

Toisessa käsittelyluvussa perehdytään dataan ja analytiikkaan. Luvussa määritellään ensin dataan liittyviä käsitteitä ja datan ominaisuuksia, jotka on tärkeä hahmottaa sen hyödyntämisen kannalta. Tämän jälkeen tutustutaan tarkemmin data-analytiikkaan sekä big data -analytiikkaan ja määritellään analytiikan prosessi, jonka kautta datasta saadaan arvoa päätöksentekoprosesseihin.

Kolmannessa käsittelyluvussa yhdistetään teoriaosuuksien sisältöä ja esimerkkien avulla hahmotellaan, mitä käyttömahdollisuuksia data-analytiikalla on kustannusvaikuttavuuden edistämisessä. Luvussa esitellään viitekehys, joka esittää vaikuttavuusperusteisen terveydenhuollon muodostumisen datan ja analytiikan prosessin pohjalta.

Työn lopuksi esitetään johtopäätökset, jotka perustuvat teoriaosuuteen ja esitettyyn viitekehykseen. Johtopäätöksissä kootaan vastaukset työn apututkimuskysymyksiin ja muodostetaan vastaus päätutkimuskysymykseen. Viimeisenä avataan tulevaisuuden näkemyksiä ja pohditaan mihin suuntaan aiheen tutkimusta olisi kannattavaa viedä.

2 Vaikuttavuus ja kustannusvaikuttavuus sosiaali- ja terveydenhuollossa

Tässä luvussa määritellään vaikuttavuuden ja kustannusvaikuttavuuden keskeisimpiä käsitteitä ja tarkastellaan, miten vaikuttavuus muodostuu potilaan tai asiakkaan näkökulmasta. Terveydenhuollon palveluita käyttävästä henkilöstä käytetään tässä työssä termiä potilas ja sosiaalihuollossa vastaava termi on asiakas. Tätä jaottelua käytetään yleisesti sosiaali- ja terveydenhuollossa ja se käy ilmi myös lainsäädännössä potilaiden ja asiakkaiden oikeuksia määritellessä (Laki sosiaalihuollon asiakkaan asemasta ja oikeuksista 812/2000; Laki potilaan asemasta ja oikeuksista 785/1992). Käsitteiden määrittelyn jälkeen esitetään kustannusvaikuttavuuden arviointia varten keinoja ja mittareita. Lopuksi perehdytään vaikuttavuus- ja tulosperusteisen sosiaali- ja terveydenhuollon määritelmiin ja tarkennetaan erovaisuuksia niiden välillä.

2.1 Vaikuttavuudesta saadaan tietoa tutkimalla

Sosiaali- ja terveydenhuollossa vaikuttavuudella tarkoitetaan muutosta potilaan tai asiakkaan terveydentilassa, joka on saatu aikaan tuotetuilla palveluilla (Pitkänen, Torkki, Tolkki, Valtakari & Leskelä 2020). Vaikuttavuuden arviointi vaatii vertailua sekä luotettavaa arviointia, jonka avulla voidaan sulkea pois muut kuin itse intervention vaikutukset potilaan tai asiakkaan tilan kohentumisessa (Malmivaara 2022). Interventiolla tarkoitetaan lääketieteessä toimenpidettä, jolla pyritään vaikuttamaan yksilön tai ryhmän terveydentilaan tai käyttäytymiseen (Terveyskirjasto 2016). Vaikuttavuuden avulla pyritään vastaamaan siihen, miten sosiaali- ja terveydenhuollossa voidaan tarjota yhdenvertaisia ja laadukkaita palveluja, joilla voitaisiin edistää koko väestön terveyttä ja hyvinvointia parhaalla mahdollisella tavalla (Malmivaara 2022).

Vaikuttavuus on yläkäsite, jota voidaan tutkia kahdella eri tutkimusasetelmalla: sosiaali- ja terveydenhuollon arjessa sekä kokeellisessa ympäristössä satunnaistetussa kontrolloidussa kokeessa. Satunnaistettu kontrolloitu koe eli RCT (randomiced controlled trial) toteutetaan niin, että tutkittavien joukko jaetaan satunnaisesti kahteen tai useampaan ryhmään niin että

yhdelle ryhmälle annetaan tutkittavaa interventiota ja muut ryhmät saavat joko standardihoitoa, plaseboa tai ei mitään (Hinkka-Yli-Salomäki 2021). Satunnaistetun kontrolloidun tutkimuksen avulla saadaan luotettavimmaksi koettua tietoa, kun tutkitaan yhden toimenpiteen vaikuttavuutta (Malmivaara 2022). RCT tutkimuksessa vertailu toteutetaan ryhmien välillä, joiden vertailukelpoisuus on hyvä, eikä siinä käytetä niin kutsuttua ”historiallista kontrollia” eli interventiota saanutta ryhmää verrataan potilastietojen aikaisemmin standardihoitoa saaneisiin potilaisiin (Hinkka-Yli-Salomäki 2021). Tilanteessa, jossa potilaiden arpominen kontrolliryhmiin ei ole eettisesti oikein voidaan hyödyntää vertaiskontrolloitua tutkimusta eli BCT (benchmarking controlled trial), jossa ryhmät tehdään vertailukelpoisiksi valitsemalla mahdollisimman samankaltaisia potilaita hoitavat yksiköt ja hyödyntämällä tilastotieteiden apukeinoja (Malmivaara 2021).

Kokeellista vaikuttavuutta tutkitaan kontrolloiduissa ympäristöissä, jotka eivät ole reaalimaailmassa täysin mahdollisia. Sosiaali- ja terveydenhuollon arjen toimenpiteiden vaikuttavuuden seuraamiseksi Malmivaara (2013) on lanseerannut käsitteen arkivaikuttavuus, jonka tutkimiseen paras keino on havainnoiva vertaiskontrolloitu tutkimus (BCT). Vertaiskontrolloitujen tutkimusten tuottama tieto sosiaali- ja terveydenhuollon arkivaikuttavuudesta edellyttää, että käytössä on kansalliset laaturekisterit (Malmivaara 2022). Laaturekistereihin kerätään potilas- ja asiakasryhmäkohtaisesti tietoa potilaista/asiakkaista ja heidän saamistaan hoidoistaan ja hoitotuloksista (Jonsson & Pikkujämsä 2019). Laaturekisterien avulla kerättyä tietoa vaikuttavuudesta voidaan seurata ja arvioida luotettavasti sekä ne tekevät hoitoyksiköiden välisen vertaisarvioinnin mahdolliseksi (Jonsson & Pikkujämsä 2019).

Satunnaistetuissa kontrolloiduissa tutkimuksissa tutkimusryhmä koostuu vain pienestä joukosta potilaita, kun taas vertaiskontrolloitu tutkimus pohjautuu laaturekiesterien tietoon, joiden tietosisältömääräykset edellyttävät koko potilasaineiston sisällyttämisen. Sen vuoksi vertaiskontrolloitu tutkimusasetelma soveltuu hyvin, kun mitataan vaikuttavuutta, joka saadaan aikaan sosiaali- ja terveydenhuollon uudistuksiin ja kansanterveystyöhön kohdistuvilla interventioilla (Malmivaara 2022).

Oikean menetelmän löytämiseksi vaikuttavuustutkimusta tulisi lähestyä PICO-viitekehyksen kautta. PICO (patient/population, intervention, comparison, outcome) on järjestelmällinen katsaus, jonka avulla vaikuttavuustutkimuksen kysymyksenasettelu saadaan muotoiltua oikein. PICO-mallin perään voidaan lisätä myös kirjain S eli *study outcome* kuvaamaan tutkimusasetelmaa. Mallin mukainen tutkimuskysymyksen asettelu ja tulosten arviointi

varmistavat sen, että vaikuttavuustutkimukselle tärkeät elementit, potilasryhmän, intervention sekä hyödynnetyn vertailun ja mittariston kuvaukset, ilmenevät riittävällä tarkkuudella. (Methley, Campbell, Chew-Graham, McNally & Cheraghi-Sohi 2014)

2.2 Kustannusvaikuttavuus käsitteenä

Sosiaali- ja terveydenhuollon vaikuttavuutta olisi mahdollista kasvattaa lähes rajattomasti, jos käytössä olevat resurssit eivät tulisi esteeksi (Pitkänen, Torkki, Tolkki, Valtakari & Leskelä 2020). Vaikuttavuutta tutkittaessa on siis otettava huomioon myös terveystaloustieteellinen näkökulma ja hyvinvointialueiden rajalliset resurssit. Sosiaali- ja terveydenhuollon kustannusvaikuttavuudella tarkoitetaan eri yhteyksissä niin virallisissa esityksissä kuin myös vapaamuotoisemmassa viestinnässä hieman eri asioita ja kustannusvaikuttavuuteen liittyvät käsitteet sekoittuvat usein keskenään (Peura, Turunen, Purmonen, Happonen & Martikainen 2011). Työssä tarkennetaan seuraavaksi kustannusvaikuttavuuden käsitteitä niin kuin ne on kirjallisuudessa esitetty.

Pitkänen et al. (2020) määrittelevät kustannusvaikuttavuuden vaikuttavuuden suhteena kustannuksiin, jotka ovat syntyneet vaikutuksen aikaansaamisen seurauksena. Kustannusvaikuttavuus voidaan esittää kaavamuodossa seuraavan mukaisesti:

$$Kustannusvaikuttavuus = \frac{vaikuttavuus}{kustannukset} = \frac{asiakkaalle\ relevantit\ vaikutukset}{kustannukset\ näiden\ vaikutusten\ saavuttamiseksi}$$

Määritelmässä tuodaan esille sekä vaikuttavuus että vaikutus, joten tässä kohtaa on syytä tarkentaa käsitteiden välistä eroa. Vaikuttavuusajattelu kuten myös kustannusvaikuttavuusajattelu perustuvat logiikkaketjuun (Kuva 1), joka koostuu neljästä portaasta (Aistrich 2014). Alimmalla portaalla on panos eli resurssit kuten raha, henkilöstö ja osaaminen. Toisella portaalla on tuotos, joka on mitattava tehty työ kuten työtunnit, potilas-/asiakastapaamisten määrä ja kirjoitetut raportit. Tuotoksesta päästään vaikutukseen, joka kertoo, mitä konkreettisia muutoksia saatiin aikaan panoksen ja tuotoksen avulla. Terveydenhuollossa vaikutus voi kuvata esimerkiksi tietyn lääkkeen tehoa potilaan oireeseen. Viimeinen askelma eli vaikuttavuus on muutos hyvinvoinnissa, joka on saatu aikaan logiikkaketjun alempien

vaiheiden kautta. Vaikuttavuus pitää sisällään vaikutusta laajemman hyödyn ja yhteiskunnallisen näkökulman. Näin ollen voidaan myös todeta, että kustannusvaikutus eli yksinkertaisesti muutos kustannuksissa on eri asia kuin kustannusvaikuttavuus (Torkki 2019).

Kuva 1: Vaikuttavuusajattelun logiikkaketju (mukaillen Aistrich 2014)

Vaikuttavuusketjua voidaan tarkastella kahdesta eri suunnasta niin sanotusti ”jalanjäljen” tai ”kädenjäljen” kautta. Jalanjäljen mukaan tarkastelu aloitetaan käytössä olevista panoksista ja tuotoksesta arvioimalla, millainen yhteiskunnallinen vaikutus niillä on. Kun puolestaan tavoitteena on luoda kustannusvaikuttavampaa sosiaali- ja terveydenhuoltoa, tarkastelu lähtee ketjun ylimmältä askeleelta ja tarkoituksena on kehittää alempia askelmia tavoitteen mukaisiksi. (Heliskoski, Humala, Kopola, Tonteri & Tykkyläinen 2018)

Samoin kuin vaikuttavuutta, myös kustannusvaikuttavuutta voidaan arvioida sen eri tasojen kautta. Tasot auttavat hahmottamaan, miten kustannusvaikuttavuus muodostuu kullakin sosiaali- ja terveydenhuollon kokonaisuuden muodostavalla osa-alueella. Mittaamisen eri tasoista on hyötyä vaikuttavuustutkimuksen ja johtamisen eri informaatiotarpeisiin (Pitkänen, Haavisto, Vähäviita, Torkki, Leskelä ja Komssi 2018). Malmivaara (2020) jakaa kustannusvaikuttavuustutkimuksen tutkimuskysymykset seuraavaan viiteen pääluokkaan:

1. Yksittäisen biologisen intervention kustannusvaikuttavuus
2. Toimenpiteen kustannusvaikuttavuus

3. Hoitoketjun kustannusvaikuttavuus
4. Vertaisten välinen kustannusvaikuttavuus
5. Sosiaali- ja terveydenhuoltojärjestelmän ja kansanterveystyön kustannusvaikuttavuus

2.3 Vaikuttavuuden mittaaminen ja arviointi

Vaikuttavuuden mittaaminen ei ole yksiselitteistä, sillä usein ei pystytä määrittämään suoraviivaista polkua suoritteesta vaikutukseen. Arviointia haastavat epävarmuustekijät, kuten diagnoosin oikeellisuus, potilaan/asiakkaan geneettinen profiili, hoitovaste, hoitomyöntyvyys ja motivaatio hoitoon (Pitkänen et al., 2018). Sosiaali- ja terveydenhuollossa pieni joukko potilaita tai asiakkaita käyttää suurimman osan palveluista (Kivipelto, Suhonen, Koivisto, Tiirinki & Miikki, 2019). Erityisesti paljon palveluita hyödyntävien kohdalla on haasteellista tunnistaa yksittäisen palvelun vaikutukset, sillä potilaan tai asiakkaan tilan muutokseen vaikuttavat palvelujen lisäksi perimä, fyysinen ja sosiaalinen elinympäristö, elintavat sekä sattuma (Pitkänen et al., 2018). Vaikuttavuuden mittaamiseen tarvitaan yhtenäiset mittarit ja olisikin suositeltavaa käyttää jo olemassa olevia mittareita, jotta vertailukelpoisuus säilyy hyvänä (Pitkänen et al., 2018).

Vaikuttavuudesta ja kustannusvaikuttavuudesta saadaan konkreettista tietoa satunnaistettujen kontrolloitujen ja vertaiskontrolloitujen tutkimusten avulla. PICOS-viitekehyksen mukaisesti jo tutkimuksen asettelussa tulee määrittää, mitä mittareita tutkimuksessa hyödynnetään. Sosiaali- ja terveydenhuollon mittarit ovat hyvin monipuolisia, jolloin mittaaminen voidaan tehdä tutkimuksista riippuen eri näkökulmista. Laaja-alaisuudesta johtuen mittareiden luokittelulle on useita eri vaihtoehtoja ja tässä työssä hyödynnetään Valtioneuvoston Reittiopas vaikuttavuuteen (Pitkänen, Tolkki, Torkki, Leskelä ja Valtakari 2020) mukaista ryhmittelyä (Taulukko 1), joka perustuu tiedon lähteeseen.

Taulukko 1: Mittareiden ryhmittely (mukaillen Pitkänen, Tolkki, Torkki Leskelä & Valtakari 2020)

	PROM (Patient-reported Outcome Measures)	**PREM (Patient-reported Experience Measures)**	**Kliiniset mittarit ja toimintakykytestit**	**Ammattilaisten raportoimat mittarit**
Tiedon lähde	Asiakkaan arvio	Asiakkaan arvio	Objektiivinen mittaus	Ammattilaisen arvio
Kuvaus	Asiakkaan käsitys hyvinvoinnistaan, toimintakyvystään ja terveydentilastaan	Asiakkaan raportoima kokemus saamastaan palvelusta tai hoidosta	Esimerkiksi laboratoriokoe ja kävelytesti	Ammattilaisen arvio asiakkaan toimintakyvystä ja terveydestä
Mihin tieto perustuu	Asiakkaan arvio	Asiakkaan arvio	Asiakastietojärjestelmät ja rekisterit	Asiakastietojärjestelmät ja rekisterit
Esimerkki mittarista	15D ja EQ-5D	NPS	10 metrin kävelytesti	RAI

Jotta vaikuttavuudesta saadaan kokonaisvaltaista tietoa, tarvitaan mittareita, jotka kuvaavat vaikuttavuutta kolmesta eri näkökulmasta. Nämä näkökulmat ovat asiantuntijan raportoimat mittarit, asiakkaan omat arviot sekä kliiniset mittarit. Lääketieteessä suosittiin aikaisemmin lähes pelkästään objektiivisia mittareita, jotka antoivat helposti yleistettäviä ja toistettavia tuloksia hoidon vaikuttavuudesta (Malmivaara 2022). Potilaan ja asiakkaan hoitoon osallistumisen kasvaessa biologisten mittareiden rinnalle on otettu myös potilaan omaan arvioon perustuvia PROM-mittareita. Asiakkaan raportoimia (Patient-reported Outcome Measures) mittareita ovat terveyteen liittyvää elämänlaatua kuvaavat mittarit kuten 15D ja EQ-5D (Nilsson, Orwelius & Kristenson 2016; Malmivaara 2022). 15D on Suomessa kehitetty

elämänlaatukysely, jossa tuloksia mitataan 15 eri kysymyksen avulla. Kysymysten eli toisin sanoen ulottuvuuksien perusteella potilaan terveydentila arvotetaan numeraalisesti ja arvioinnin avulla voidaan seurata eri ulottuvuuksilla tapahtuvia muutoksia yksityiskohtaisesti (Sintonen, Roine, Ryynänen, Blom & Hytönen 2022). Viisi ulottuvuutta sisältävän EQ-5D:n tarkoitus on puolestaan täydentää muita mittareita ja kartoittaa terveydentilojen arvotusten eroja eri maiden välillä (Sintonen, Roine, Ryynänen, Blom & Hytönen 2022). Toinen PROM-mittarityyppi on tiettyyn tautiin kehitetyt tilaspesifit mittarit kuten lonkkakipujen hoidossa hyödynnettävä toimintakykyindeksi Oxford Hip Score, joka toteutetaan myös kyselylomakkeen avulla (Black 2013).

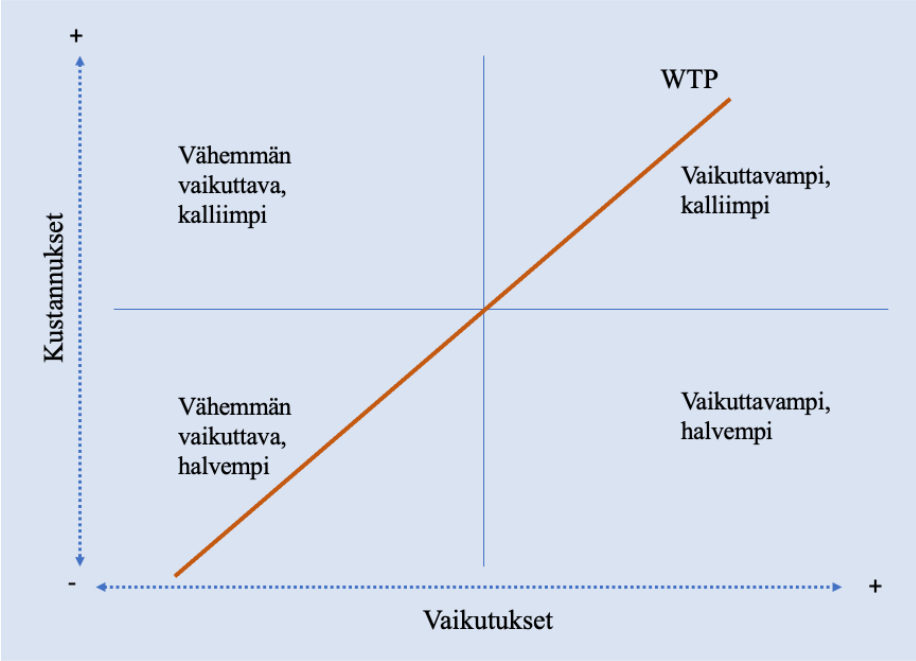
Siinä missä PROM-mittarit ovat tulosmittareita, PREM-mittarit (Patient-reported Experience Measures) keskittyvät kuvaamaan potilaan tai asiakkaan kokemusta (Nilsson, Orwelius, Kristenson 2016). PREM-mittareilla saadaan tietoa asiakkaan tyytyväisyydestä hoitoon sekä kokemuksia siihen liittyen (Pitkänen et al., 2018). NPS eli Net Promoter Score on yleisesti hyödynnetty PREM-mittari, joka perustuu yksinkertaisuudessaan kysymykseen siitä, miten todennäköisesti asiakas suosittelisi samaansa palvelua (Netpromoter 2023). PREM-mittarit ovat tärkeitä lähteitä, kun halutaan parantaa hoidon vaikuttavuutta ja laatua. Positiivisen hoitokokemuksen on huomattu parantavan hoitoon sitoutumista, joka puolestaan on tärkeä tekijä kliinisen vaikuttavuuden kannalta (Doyle, Lennox & Bell 2013). Pitkänen et al. (2018) toteavat kuitenkin, että aina positiivinen kokemus hoidosta ei lisää suoraan vaikuttavuutta ja huonoksi koettu hoito voi silti olla vaikuttavaa.

Asiakkaan terveydentilastaan tekemän arvion lisäksi vaikuttavuuden mittaamiseen tarvitaan Pitkänen et al. (2018) mukaan yhä myös terveydenhuollon ammattilaisen raportoimia sekä objektiivisia mittareita. Objektiiviset kliiniset mittarit kuten laboratoriokokeet ja toimintakykytestit ohjaavat hoitohenkilökunnan tekemiä hoitopäätöksiä. Toimintakykytestien tulokset eivät kuitenkaan ole aina relevanteimpia asiakkaan näkökulmasta, sillä usein potilaalle itselleen on merkittävämpää, suoriutuuko hän arjen toiminnoistaan kuin mikä kävelytestin avulla arvioitu nivelen liikeradan laajuus on. Ammattilaisen tekemällä arviolla tarkoitetaan kaikkia ammattilaisen potilaasta tai asiakkaasta tekemiä arvioita. Esimerkkinä ammattilaisten tekemästä arvioinnista on ikäihmisten hoidossa hyödynnettävä RAI-arviointi, jolla kartoitetaan ikääntyneiden hoidon tarvetta, hoidon laatua sekä kustannusvaikuttavuutta. (Pitkänen et al., 2018)

Kustannusvaikuttavuutta voidaan arvioida kustannusvaikuttavuusanalyysin (cost-effectiveness analysis, CEA) avulla. Kustannusvaikuttavuusanalyysi on terveystaloustieteessä paljon hyödynnetty analyysi, joka vertaa kahden tai useamman vaihtoehtoisen strategian kustannuksia saman toimenpiteen tai tehtävän vaikutuksiin. (Bang & Zhao 2012). Tärkeäksi tunnusluvuksi kustannusvaikuttavuusanalyysissä on muodostunut englanninkielinen incremental cost-effectivenes ratio eli ICER (Bang & Zhao 2012). ICER kuvaa vertailutoimenpiteen tuomaa kustannusvaikuttavuuden kasvua:

$$ICER = \frac{Kustannus\ A - Kustannus\ B}{Hoitotulos\ A - Hoitotulos\ B}$$

Kaava on esitetty Malmivaaran (2022) teoksessa ja siinä A tarkoittaa indeksitoimenpidettä ja B vertailutoimenpidettä. Kaavan osoittaja kuvaa toimenpiteeseen A liittyvää kustannusvaikutusta ja nimittäjä hoidolla A saavutettavaa vaikuttavuutta (Malmivaara 2022). Kustannusvaikutuksen ja vaikuttavuuden suhteesta saadaan inkrementaalinen (vähittäin kasvava) kustannusvaikuttavuus. Tunnuslukua hyödynnettäessä on hyvä huomata se, että vaikka kokonaiskustannukset pysyvät samana ICER tulee arvioida jokaisen vaikuttavuusmuuttujan osalta erikseen (Malmivaara 2022).



Kuva 2: ICER nelikenttä

ICER kuvataan usein yllä olevan kuvan (Kuva 2) mukaisen nelikentän avulla, joka kuvaa mitattavan toimenpiteen tai asian vaikuttavuuden muodostumista verrattuna kustannuksiin. Usein tutkimuksen tulokset asettuvat kohtaan, jossa uusi hoito on vaikuttavampaa, mutta myös kalliimpaa. Origon kautta kulkeva jana (willingness to pay, WTP) kuvastaa yhteiskunnan halukkuutta maksaa tietyn tasoisesta kustannusvaikuttavuudesta, joten janan sijoittuminen kuvaajalla vaihtelee tilanteesta riippuen. (Malmivaara 2022)

Kustannusvaikuttavuusanalyysin lisäksi voidaan hyödyntää kustannus-utiliteettianalyysia, jossa sanalla utiliteetti tarkoitetaan tavoiteltavaa terveyshyötyä tai hoidon tulosta (Heiskanen, Rannanheimo & Härkönen 2018). Kustannus-utiliteettianalyysit perustuvat kustannusvaikuttavuuden kuvaamiseen elämänlaatua kuvaavien mittareiden avulla ja siinä kustannuksia verrataan saavutettuihin laatupainotettuihin elinvuosiin (quality adjusted life year, QALY) (Vuorela 2017). QALY lasketaan arvioimalla potilaan elinvuodet intervention jälkeen niin että jokainen vuosi painotetaan elämänlaatupisteillä (Vuorela 2017). Laatupainotettuihin elinvuosiin on suhtauduttu ympäri maailmaa kriittisesti, joten myös kustannus-utiliteettianalyysiin on syytä suhtautua samalla varovaisuudella (Rand & Kesselheim 2021).

Kustannushyötyanalyysin avulla voidaan kuvata paremmin arvioitavan intervention arvoa yhteiskunnalle. Siinä vaikutuksia ei mitata terveyteen perustuvien mittareiden avulla vaan arviointi tapahtuu rahayksiköissä (Drummond, Schulper, Torrance O'Brien & Stoddard 2005). Kustannushyötyanalyysin avulla voidaan siis ottaa kantaa allokatiiviseen tehokkuuteen, eli siihen, mitkä hyvinvointia edistävät toimenpiteet tulee valita, jotta kustannusvaikuttavuus paranee (Valtonen 2011). Kustannushyötyanalyysin heikkous on se, että terveysvaikutusten arvioiminen kustannusten perusteella koetaan ongelmalliseksi. (Drummond et al., 2005) Inkrementaalisen kustannusvaikuttavuuden yhteydessä esitelty maksuhalukkuus (WTP) on myös yksi terveysvaikutusten rahalliseen arvottamiseen käytetty keino.

2.4 Vaikuttavuus- ja tulosperusteinen sosiaali- ja terveydenhuolto

Kuten esitellystä mittaristosta huomataan, vaikuttavuuden ja kustannusvaikuttavuuden mittaaminen on menossa yhä enemmän siihen suuntaan, jossa potilaan yksilölliset tarpeet pyritään huomioimaan laajemmin niin lääketieteessä kuin hoitopolkujen määrittelyssä. Myös sosiaali- ja terveyspalvelujen johtamisessa on hypätty mukaan tähän trendiin ja se näkyy vaikuttavuus tai arvoperusteisen terveydenhuollon yleistymisenä (Laine 2021).

Michael Porter ja Elizabeth Teisberg määrittelivät teoksessaan Redefining Health Care (2006) käsitteen value-based health care (VBHC) eli arvoperusteinen terveydenhuolto, jolla viitataan kustannusvaikuttavuuteen pyrkivään terveydenhuollon johtamiseen. Arvoperusteinen terveydenhuolto siirtää sosiaali- ja terveydenhuoltoa järjestelmäkeskeisestä lähestymistavasta kohti potilaskeskeisempää mallia, jossa pääasiallinen tavoite on pyrkiä kulujen karsimisen sijaan luomaan arvoa asiakkaille ja potilaille. Arvo määritellään teoksessa kustannusvaikuttavuuden tapaan eli potilaalle tai asiakkaalle tuotetun hoidon vaikuttavuus suhteutettuna siihen käytettyihin kustannuksiin (Porter & Teisberg 2006). Porterin ja Teisbergin mallissa asiakaskeskeisyys ilmenee asiakkaiden segmentoinnilla eikä niinkään yksilöllisellä tasolla. Mallissa mittareiden oletetaan olevan relevantteja koko asiakasryhmälle esimerkiksi tiettyä sairautta sairastaville, joten asiakaskeskeisyyden yksilöllisyydessä olisi vielä parantamisen varaa (Pitkänen et al., 2018).

Kotimaisessa kirjallisuudessa arvoperusteisesta terveydenhuollosta käytetään usein ristiin myös käsitettä vaikuttavuusperusteinen terveyenhuolto. Sosiaali- ja terveydenhuollossa kustannusvaikuttavuudesta ja arvon arvioinnista puhuttaessa tulee kuitenkin tiedostaa näiden

käsitteiden eroavaisuudet, sillä niiden arviointikehykset eroavat toisistaan ja arvoa mitataan eri lähtökohdista (Laine 2021). Malmivaara tekee teoksessaan Vaikuttavuus sosiaali- ja terveydenhuollossa (2022) selkeän erottelun käsitteiden välille käyttämällä arvoperusteisesta terveydenhuollosta käsitettä tulosperusteinen terveydenhuolto. Tulosperusteisessa terveydenhuollossa terveyshyödyn mittaaminen on sairaus- ja potilasryhmä kohtaista eli siinä pyritään mittaamaan muutos potilaan terveydentilassa ennen ja jälkeen hoitojakson ja tämä muutos suhteutetaan hoitoon käytettyihin kustannuksiin (Laine 2021; Malmivaara 2022). Kun huomio keskitetään hoitotuloksiin, sen seurauksena tuotetun arvon määrä, hoidon laadukkuus, vaikuttavuus ja kustannusvaikuttavuus paranevat (Malmivaara 2022). Tulosperusteisen terveydenhuollon määritelmä vastaa siis Iso-Britanniassa hyödynnettävää Porterin ja Teisbergin (2006) value-based health care määritelmää.

Vaikuttavuusperusteinen lähestymistapa sopii puolestaan paremmin, kun halutaan vertailla hoitoyksiköiden välisiä tuloksia. Kuten kaikkeen vaikuttavuuden arviointiin, myös vaikuttavuusperusteiseen terveydenhuollon mittaamiseen tarvitaan aina jokin kohde, johon tuloksia voidaan verrata. Arviointi toteutetaan vertaiskontrolloituna tutkimuksena, jossa vertaillaan kahta tai useampaa hoitolaitosta. Tutkimuksella pyritään selvittämään hoidon eroja riippuen siitä, mihin hoitolaitokseen yksittäinen asiakas tai potilas hakeutuu. Samalla määritetään tietoa palvelujen laadukkuudesta, yhdenvertaisuudesta, vaikuttavuudesta ja kustannusvaikuttavuudesta. (Malmivaara 2022)

Tulosperusteinen terveydenhuollon järjestelmä eroaa edellä mainitusta siten, että pelkkä hoitotulosten arviointi ja mittaaminen tulosperusteisessa terveydenhuollossa ei ota huomioon sitä, mikä on todellisuudessa sosiaali- ja terveydenhuollon toimenpiteen osuus potilaan tai asiakkaan ennusteen paranemisessa. Tällöin jää huomioimatta toiminnan laadukkuus ja saatetaan tehdä virheellisiä tulkintoja. Tulosperusteisen sosiaali- ja terveydenhuollon hyödyt perustuvat siihen, että sen avulla tuotetaan tärkeää tietoa arkikustannusvaikuttavuudesta ja arkivaikuttavuudesta palvelujen tuottajille. Vaikuttavuusperusteinen sosiaali- ja terveydenhuolto puolestaan tuottaa potilaille, palvelun tilaajille ja tuottajille, organisaatioille, hyvinvointialueille sekä päättäjille tärkeää tietoa yksiköiden ja alueiden vertaiskehittämiseen. Alla olevaan taulukkoon (Taulukko 2) on koottu tiivistelmä vaikuttavuus- ja tulosperusteisista sosiaali- ja terveydenhuollon malleista. (Malmivaara 2022)

Taulukko 2: Tulos- ja vaikuttavuusperusteiset mallit (mukaillen Malmivaara 2022)

	Tulosperusteinen sosiaali- ja terveydenhuolto	**Vaikuttavuusperusteinen sosiaali- ja terveydenhuolto**
Tavoitteet	Hoitojen tulosten ja kustannusten arviointi	Hoitoyksiköiden välinen vertaisarviointi
Toimintatavat	Hyödynnetään laajaa mittaristoa, jonka avulla tuotetaan luotettavaa tietoa hoitotuloksista ja kustannuksista.	Määritetään tietoa palvelujen laadukkuudesta, yhdenvertaisuudesta, vaikuttavuudesta ja kustannusvaikuttavuudesta
Mihin pyritään?	Keskitetään huomio hoitotuloksiin, jolloin asiakkaille/potilaille tuotetun arvon määrä, hoidon laadukkuus, vaikuttavuus ja kustannusvaikuttavuus paranevat	Tuotetaan laadukasta tietoa, jolla voidaan vertailla eri hoitoyksiköitä sekä alueita
Ketkä hyötyät?	Palvelujen tuottajat	Potilas, palvelujen tilaajat, palvelujen tuottajat, organisaatiot, hyvinvointialueet, päättäjät

3 Data ja analytiikka

Tämän luvun tarkoituksena on määritellä, mitä datalla tarkoitetaan ja mitkä ominaisuudet vaikuttavat sen käyttömahdollisuuksiin. Luvussa avataan datan ja big datan käsitteitä sekä perinteisen data-analytiikan eroja massadata-analytiikkaan. Lopuksi määritellään prosessi, jonka avulla datasta saadaan tuotettua relevanttia tietoa päätöksenteon tueksi.

3.1 Datan ominaisuudet ja big data

Datalla tarkoitetaan tietojärjestelmiin tallennettua tietoa, joka on koneellisesti käsiteltävissä ja jolla ei ole sellaisenaan vielä arvoa. Datan arvo ja järjestys syntyvät vasta jalostamisen yhteydessä, jolloin arvon syntyminen on sidonnaista sekä aikaan että kontekstiin (Rowley 2006). Toisin sanoen datan arvo pohjautuu siihen, miten sen tulkitsija onnistuu käsittelemään sitä ja mitä relevanttia tietoa datasta saadaan. Tieto tarkoittaa suomen kielessä informaatiosta jalostettua ja sisäistettyä tietämystä, ymmärrystä ja viisautta, mutta yhdyssanoissa käytetään usein sanaa tieto viitaten kuitenkin dataan. Tässä työssä ilmenee myös tällaisia yhdyssanoja, kuten tietovarasto ja tietoallas. (Digitaalinen Helsinki 2023)

Datan määrä ja kompleksisuus ovat kasvaneet kiihtyvää vauhtia 1990-luvulta alkaen, mikä on ollut pohjana massadata eli big data käsitteen muodostumiselle (Barnes 2013). Vaikka big datasta on sittemmin tullut muotisana, tarjoaa se uusia keinoja kehittää data-analytiikkaa, liiketoimintaa sekä päätöksentekoa tehokkaammaksi (Waller & Fawcett 2013). Big datan määritelmä vaihtelee hieman lähteestä riippuen ja siinä on tulkinnan varaa. Yleisesti käsitteellä big data viitataan useista eri tietolähteistä kerättyyn dataan, joka on määrältään ja muodoltaan haastavaa käsitellä (Waller & Fawcett 2013).

Vaikka big datalle ei ole yksiselitteistä määritelmää, laajojen tietomäärien käsittelyä voi lähestyä big datan arvon luomisen pohjalla olevan seitsemän yleisesti tunnistetun ulottuvuuden eli 7 V:n kautta (Garcia & Alvarez-Fernandez 2022). Nämä ominaisuudet eli määrä (volume), moninaisuus (variety), nopeus (velocity), arvo (value), todenmukaisuus (veracity), vaihtelevuus (variability) ja visualisointi (visualization) havainnollistavat eriteltyinä hyvin

big datan laajuutta sekä kompleksisuutta. Alla olevaan kuvaan (Kuva 3) kuvatut ulottuvuudet luovat myös pohjaa big data -analytiikalle.

Kuva 3: Big datan ominaisuudet

Määrä – Big datan ensimmäinen ominaisuus on datan määrä. Big datan määrää havainnollistaa parhaiten konkreettiset luvut ja arviot. On arvioitu, että datan määrä on 500 kertaistunut vuosien 2011–2020 aikana ja datan varastointikapasiteetti kaksinkertaistuu yhä joka kolmas vuosi (Garcia & Alvarez-Fernandez 2022). Big datan raportoinnin suuruusluokka on tera- ja petatavuissa (Balali, Nouri, Nasiri & Zhao 2020).

Moninaisuus – Big datan moninaisuudella tarkoitetaan sitä, että valtava määrä dataa on peräisin lukuisista eri lähteistä ja sitä on saatavilla useissa eri muodoissa (Sagiroglu & Sinanc 2013). Data voidaan luokitella muodoltaan strukturoiduksi, puolistrukturoitu tai strukturoimattomaksi. Strukturoitu data on helpointa käsitellä, sillä se on muodoltaan valmiiksi jäsenneltyä (Eberendu 2016). Strukturoidulla datalla on selkeä rakenne ja sen tulkitseminen tietokoneella on mahdollista. Strukturoidun datan analysointia helpottaa myös se, että sitä voidaan helposti ryhmitellä relaatioiden eli erilaisuussuhteiden avulla (Eberendu 2016). Strukturoitua dataa voivat tuottaa sekä koneet että ihmiset ja sitä on esimerkiksi maksuliikennetiedot, osoitetiedot ja sijaintitiedot (Agrawal & Kadadi 2022).

Strukturoimaton eli jäsentelemätön data on haastavampaa käsitellä, sillä sitä ei voida varastoida taulukoihin ja sen käsittely vaatii enemmän työtä (Agrawal & Kadadi 2022). Agrawalin ja Kadidin (2022) mukaan strukturoimatonta tietoa ovat esimerkiksi viestit, kuvat ja HTML-tunnisteet. Puolistrukturoitu data on muodoltaan jäsennellyn ja jäsentelemättömän välissä. Puolijäsennelty data luokitellaan jäsennellyn tiedon muodoksi, mutta se ei noudata relaatiotietokantojen ja muiden tietomallien taulukkorakennetta (Javatpoint 2023). Se sisältää kuitenkin tunnisteita ja hierarkioita, jonka vuoksi se voidaan luokitella itseään kuvaavaksi rakenteeksi (Javatpoint 2023). Strukturoidun, puolistrukturoidun ja strukturoimattoman datan analyyttinen käsittely vaatii eri toimenpiteitä sekä järjestelmiä.

Nopeus - Big datan kolmas ominaisuus on nopeus, jolla kuvataan puolestaan sitä, miten nopeasti uutta dataa syntyy ja miten nopeasti se liikkuu tietokannassa tai tietokantojen välillä (Garcia & Alvarez-Fernandez 2022). Datan nopeus mahdollistaa sen, että big datan käsittely on hyvin reaaliaikaista ja uusin tieto saadaan aina välittömästi analysoitavaksi (Bhadani & Jothimani 2016).

Arvo – Big datalla katsotaan olevan konkreettinen ja mitattavissa oleva arvo. Arvo määritellään organisaation tai yrityksen datasta saaman hyödyn perusteella (Bhadani & Jothimani 2016). Data ei siis ole itseisarvoltaan hyödyllistä vaan sen arvo syntyy sen mukaan, miten sitä pystytään hyödyntämään (Bhadani & Jothimani 2016).

Todenmukaisuus – Todenmukaisuudella tarkoitetaan sitä, miten oikeaa ja kattavaa data on ja sisältääkö se virheitä (Garcia ja Alvarez-Fernandez 2022). Garcia ja Alvarez-Fernandez (2022) toteavat myös, että tiedon todenmukaisuudella ei niinkään viitata datan laatuun vaan juuri täsmällisyyteen. Heidän mukaansa datan lähteellä, eli mistä data on kerätty, on suuri vaikutus siihen, miten todenmukaista data on.

Vaihtelevuus – Vaihtelevuus ja visualisointi on useissa lähteissä jätetty käsittelemättä, mutta kattavamman käsityksen luomiseksi ne on hyvä ottaa mukaan tarkasteluun. Tietovirtojen epäjohdonmukaisuus eli vaihtelevuus aiheuttaa haasteita datan käsittelyssä, sillä tietovirat ovat harvoin tasaisia. Kuormittavia huippuja syntyy esimerkiksi lisääntyneen sosiaalisen median käytön seurauksena. (Katal, Wazid & Goudar 2013)

Visualisointi – Datan visualisointi on big datan ominaisuuksista se, joka on eniten esillä analytiikkaan perehtymättömälle. Visualisoinnin avulla pyritään tekemään big datan monimutkaisista kaavoista näkyvämpiä ja helpommin havaittavia (Miller 2017). Millerin (2017)

mukaan big datan visualisoinnin avulla pystytään lisäämään datan selkeyttä, ratkaisemaan ongelmia, korostamaan ja kuvaamaan muuten näkymätöntä tietoa sekä luomaan ennustavia kuvaajia. Yhteenvetona voidaan todeta, että big datan ominaisuudet ja näin ollen hyödyntämismahdollisuudet ovat hyvin laajat. Jotta kompleksisesta tietomäärästä saadaan tuotettua hyödyllistä tietoa, tulee myös analytiikan keinojen olla entistä tehokkaampia.

3.2 Data-analytiikka

Tiedon luominen ja varastoiminen on pitkään ollut helpompaa kuin sen hyödyntäminen, mutta tosi asia on se, että jalostamattomasta datasta ei ole hyötyä tietoaltaisiin ja tietovarastoihin säilöttynä. Data-analytiikan keinoin datasta pyritään löytämään kaavoja ja riippuvuuksia, joiden avulla voidaan luoda organisaation toimintaa kuvaavia malleja (Balali, Nouri, Nasiri & Zhao 2020). Data-analytiikka kuvaa siis joukkoa erilaisia tekniikoita, työkaluja ja prosesseja, jotka auttavat tiedon sisäistämisessä ja sen saattamisessa päätöksentekoa helpottavaan muotoon (Runkler 2012). Analytiikka voidaan jakaa kahteen lähestymistapaan, jotka ovat datalähtöisyys ja mallilähtöisyys.

Analytiikka käsitteenä yhdistetään usein suurien tietomäärien eli massadatan käsittelyyn sekä tietokone ohjelmiin, jotka käsittelevät tätä dataa eli toisin sanoen datalähtöiseen analytiikkaan (data-driven analysis) (Greasley 2019). Datalähtöinen analytiikka pyrkii tuottamaan kuvauksen siitä, miten jokin systeemi toimii eri skenaarioissa eli millainen tuotos saadaan aikaam tietyllä syötteellä (Greasley 2019). Datalähtöiseen analyysiin voidaan lisätä mahdollisuus luoda ennustuksia hahmontunnistuksen (pattern recognition) avulla (Greasley 2019). Hahmontunnistus on koneoppimisen osa-alue, joka pyrkii sekvensoinnin, luokittelun ja klusteroinnin avulla löytämään malleja ja kaavoja ”kohinaisesta tiedosta” eli datasta, jossa on paljon hyödyttömäksi katsottua tietoa ja jopa vikoja (Liu, Sun ja Wang 2006). Koneoppimisen lisäksi datalähtöisen analyysin tekniikoita ovat tiedonlouhinta, prosessilouhinta ja visuaalinen analytiikka (Greasley 2019).

Mallilähtöisen analyysin (model-driven analysis) avulla voidaan syötteen ja tulosteen välisen yhteyden sijasta selittää tarkemmin systeemin sisäistä rakennetta. Mallilähtöiset analyysityökalut ovat selittävää analytiikkaa ja niiden avulla pyritään yksinkertaistamaan

järjestelmä sen keskeisempien elementtien avulla. Mallilähtöinen analyysi vaatii vähemmän dataa kuin datalähtöinen analyysi, sillä mallilähtöisyys pyrkii kuvaamaan rakenteita yksinkertaisesti. Mallilähtöiset analytiikan keinot soveltuvat monimutkaisten järjestelmien selittämiseen helpommin ymmärrettävässä muodossa. (Greasley 2019)

Datalähtöinen analyysi sopii siis mallilähtöistä analyysia paremmin massadatan käsittelyyn ja monet big data -analytiikan menetelmät perustuvat juuri datalähtöisen analyysin tekniikoihin kuten koneoppimiseen ja datan visualisointiin. Tekniikoita ei voida kuitenkaan luokitella perinteisen ja massadata-analytiikan keinoihin, sillä massadata-analytiikan pohja on perinteisen analytiikan keinoissa. Syy, minkä takia massadata-analytiikka on rakentunut perinteisen data-analytiikan rinnalle, on se, että se mahdollistaa strukturoimattoman ja nopeasti muuttuvan datan käsittelyn. Massadata-analytiikka on perinteistä data-analytiikkaa syvällisempää ja sen avulla voidaan tutkia rajatun mittarimäärän sijaan laajempia kokonaisuuksia (Bhadani & Jothimani 2016). Sen avulla on esimerkiksi mahdollista tehdä tutkimus koko populaatiosta yksittäisen otoksen sijaan (Bhadani & Jothimani 2016).

Monimutkainen ja haastava data edellyttää analytiikalta uutta ajattelua ja pääpaino asettuu suurien datamassojen hallintaan ja reaaliaikaiseen käsittelyyn (Katal, Wazid & Goudar 2013). Data-analytiikka tulisi ryhmittelyistä huolimatta hallita kokonaisuutena. Big data -analytiikkaa ei voi ymmärtää, jos ei ole perehtynyt perinteisen data-analytiikan malleihin. Datalähtöinen ja mallilähtöinen analyysi puolestaan tarjoavat yhdessä uusia ratkaisuja analyyttisten ongelmien ratkaisemiseen.

3.3 Data-analytiikan prosessi

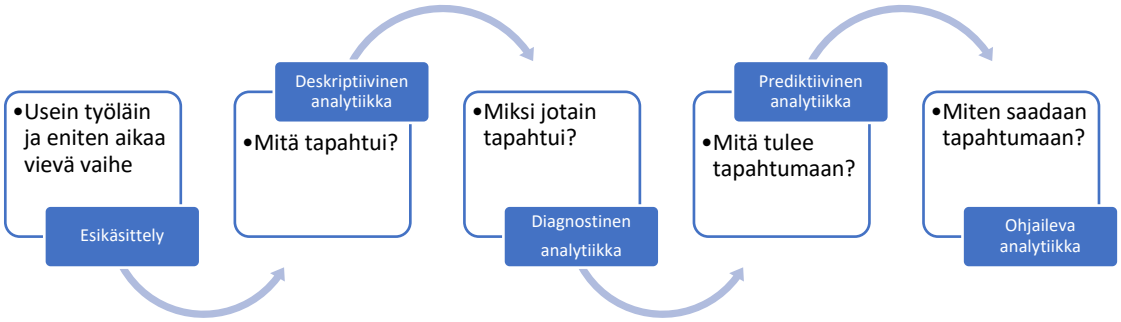
Datan jalostaminen hyödynnettävään muotoon tapahtuu monivaiheisen prosessin kautta. Prosessi jakaa data-analytiikan sen kompleksisuuden sekä arvon pohjalta ja sen vaiheista voidaan myös käyttää termiä data-analytiikan taso. Data-analytiikan prosessi alkaa datan keräämisestä ja varastoimisesta, mutta tässä työssä pääpaino on jo kerätyn datan työstämisessä arvoa luovaan muotoon.

Data-analytiikan työläin ja eniten resursseja vievä vaihe on usein datan esikäsittely. Kuten aiemmin jo todettiin datan määrä, on lisääntynyt ja lisääntyy valtavasti uusien kanavien ja teknologioiden myötä. Lisääntynyt datan määrä ei kuitenkaan automaattisesti viittaa

tarpeellisen tiedon määrän kasvamiseen, joten datasta on osattava löytää ne osat, jotka ovat kutakin käyttötarkoitusta ajatellen relevantteja. Data ei myöskään ole aina suoraan hyödynnettävissä, sillä se voi olla haasteellisessa muodossa tai sisältää virheitä, duplikaatteja tai puuttuvia osia (Garcia et al., 2016).

Datan esikäsittelyllä varmistetaan, että tietovirtojen laatu on riittävällä tasolla. Esikäsittely koostuu kahdesta osasta, jotka ovat datan valmistelu ja datan vähentäminen. Eri tietoläheistä kerätyn datan yhdistäminen, datan normalisointi, skaalaaminen, muuntaminen ja puhdistaminen ovat pakollisiksi katsottuja datan valmistelukeinoja (Garcia et al., 2016). Datan vähentämisellä pyritään puolestaan selkeyttämään käsiteltävää dataa, jotta turhat ominaisuudet saadaan poistettua ja käsittelystä tulee vaivattomampaa. Garcia et al. (2016) esittelevät datan vähentämisen työkaluiksi muun muassa ominaisuuksien ja instanssien valinnan sekä diskretisoinnin. Mainittujen esimerkkien lisäksi datan esikäsittelemiseen on useita eri menetelmiä.

Data-analytiikan prosessi jakautuu neljään vaiheeseen ja kun prosessi etenee näiden vaiheiden mukaisesti datasta saatava arvo kasvaa (Balali, Nouri, Nasiri & Zhao 2020). Arvon lisääntymisen lisäksi prosessista on hyvä huomioida erityisesti sen eri vaiheiden aikasidonnaisuus. Kaksi ensimmäistä vaihetta käsittelevät menneisyyteen viittaavaa dataa ja kahdessa viimeisessä vaiheessa pyritään ennustamaan tulevaa. Balali et al. (2020) mukaan analytiikan prosessia voidaan soveltaa sekä perinteiseen analytiikkaan että big data -analytiikkaan. Big data vaikuttaa kuitenkin osaltaan analytiikkaprosessin tehokkuuteen erityisesti, silloin kun hyödynnettävän datan epätarkkuus ja monimutkaisuus kasvavat (Balali, Nouri, Nasiri & Zhao 2020). Seuraavaksi esitellään data-analytiikan prosessin vaiheet, jotka on esitetty myös alla olevassa kuvassa (Kuva 4).



Kuva 4: Data-analytiikan prosessi

Ensimmäinen data-analytiikan prosessin vaiheista on deskriptiivinen eli kuvaileva analytiikka. Kuvaileva analytiikka vastaa kysymykseen ”Mitä tapahtui?” ja sen avulla tunnistetaan mennyttä toimintaa kuvaavasta datasta tärkeää tietoa organisaation prosesseista (Jeyanthi 2022). Datalähtöinen deskriptiivinen analytiikka pitää sisällään tilastollisia kuvauksia muuttujista ja nykyisistä sekä menneistä trendeistä (Greasley 2019). Mallilähtöistä deskriptiivistä analytiikkaa hyödynnetään esimerkiksi uusien tuotteiden ja prosessien suunnittelussa silloin kun on olemassa vain vähän dataa nykyisestä tilanteesta (Greasley 2019). Deskriptiivinen analytiikka on ensimmäisenä analytiikan vaiheena tärkeä työkalu suunnittelutyössä (Gröhn ja Ranta 2022).

Prosessin toinen vaihe on diagnostinen analytiikka, jonka avulla saadaan vastaus kysymykseen ”Miksi jotain tapahtui?”. Diagnostisen analytiikan avulla menneisyyteen perustuvasta datasta etsitään syy-seuraussuhteita. Sen avulla voidaan selvittää esimerkiksi, miten jonkun tekijän muuttaminen vaikuttaa prosessin lopputulokseen. Tutkittavien syy-seuraus- eli kausaaliyhteyden selvittäminen voidaan toteuttaa satunnaistettujen kokeiden avulla. (Balali, Nouri, Nasiri & Zhao 2020)

Prediktiivinen eli ennakoiva data-analytiikka on datan, koneoppimisen ja tekoälyn hyödyntämistä tulevaisuuden ennustamiseksi (Jeyanthi 2022). Prediktiivinen data-analytiikka vastaa kysymykseen ”Mitä tulee tapahtumaan lähitulevaisuudessa?”. Tässä analytiikan vaiheessa aiemmin kerätty data käsitellään matemaattisten mallien avulla. Näiden mallien avulla pyritään löytämään ratkaisevia kaavoja ja trendejä, joiden avulla voidaan ennustaa tulevaa (Jeyanthi 2022).

Myös prediktiivistä analytiikkaa voidaan toteuttaa niin data- kuin mallilähtöisesti. Datalähtöinen prediktiivinen analytiikka on ylläkuvattua kaavojen ja riippuvuuksien tunnistamista historiatiedosta ja sen perusteella tehtäviä ennusteita (Greasley 2019). Mallilähtöiset lähestymistavat hyödyntävät toimialasta kerättyä tietoa rakentaakseen yksinkertaistetun mallin sen lähitulevaisuudesta (Greasley 2019). Ennakoiva analytiikka tarjoaa yrityksille ja organisaatioille apua toiminnan suunnitteluun, mutta se ei ole tae siitä, että ennustetut asiat tulevat todella tapahtumaan. Prediktiivinen analytiikka kertoo siis vain todennäköisyyksiä tapahtumille nojaten historiatietoon (Gröhn ja Ranta 2022).

Neljäs ja viimeinen analytiikkaprosessin vaiheista on ohjaileva analytiikka, joka rakentuu sekä kuvailevan että ennakoivan analytiikan pohjalta. Ohjailevaa analytiikkaa pidetään jo

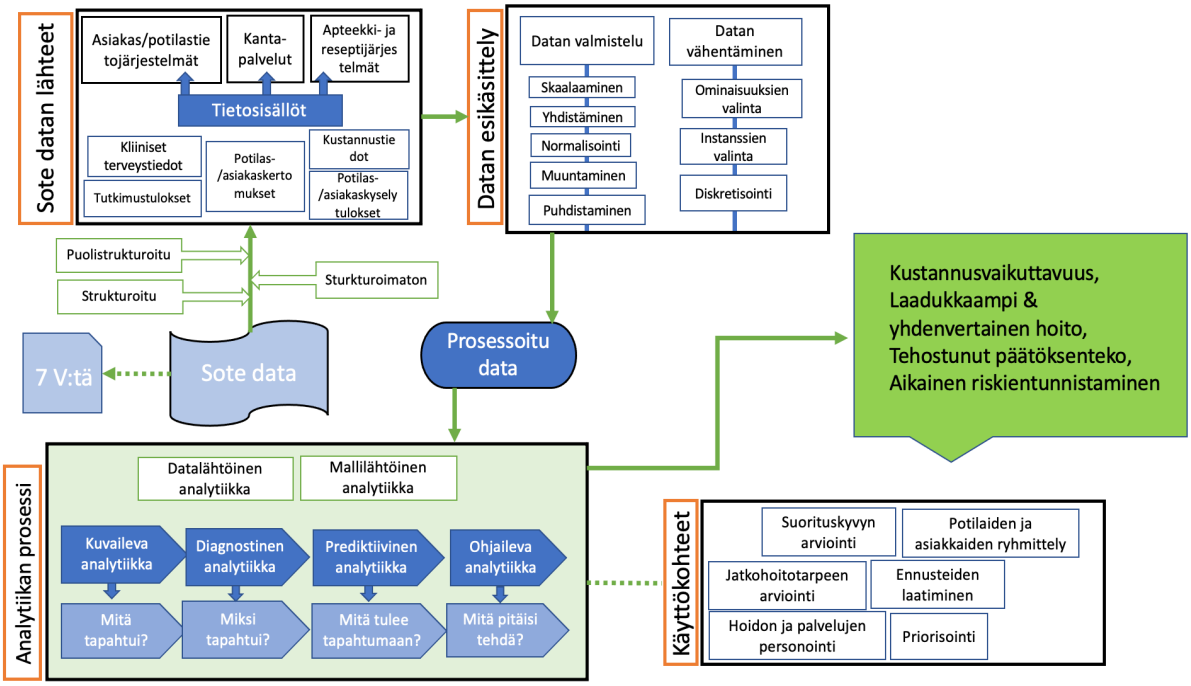
edistyksellisenä päätöksenteon tasona ja sen edellytyksenä on, että aikaisemmat analyysit on toteutettu riittävän kattavasti (Jeyanthi 2022). Ohjailevan analytiikan avulla voidaan luoda strategioita ja visioita historiatiedon ja ennustusten pohjalta mille tahansa aikahorisontille. Strategioiden pohjalla on kysymys siitä, miten jokin haluttu asia tai vaikutus saadaan toteutumaan (Jeyanthi 2022). Datalähtöinen ohjaileva analytiikka pyrkiikin löytämään konkreettisia keinoja hyödyn maksimoimiseksi tai haitan minimoimiseksi. Mallilähtöisessä lähestymistavassa voidaan puolestaan hyödyntää esimerkiksi erilaisia optimointia hyödyntäviä järjestelmiä (Greasley 2019).

4 Data-analytiikka kustannusvaikuttavuuden edistämisessä

Tässä luvussa esitellään viitekehys, jonka avulla pyritään kohti kustannusvaikuttavampaa sosiaali- ja terveydenhuoltoa data-analytiikkaa hyödyntäen. Yleiskatsauksen jälkeen käydään yksityiskohtaisemmin läpi viitekehyksen eri osa-alueita. Käsittely aloitetaan sosiaali- ja terveydenhuollon datan lähteiden löytämisellä sekä tunnistamalla haasteet, joita datan käsittelyyn liittyy. Sen jälkeen esitellään, miten teoriaosuudessa avattua data-analytiikan prosessia voidaan soveltaa sosiaali- ja terveydenhuollon arjessa ja järjestelmän kehittämisessä. Viimeiseksi esitellään neljä esimerkkiä, joissa kustannusvaikuttavuutta on saatu parannettua data-analytiikan prosessin avulla.

4.1 Viitekehys kohti kustannusvaikuttavampaa sosiaali- ja terveydenhuoltoa

Nykypäiväiset datalähtöiset terveydenhuollon sovellukset vaativat monitieteellistä osaamista ja erityistä teknistä tietotaitoa, jotta suuresta määrästä dataa voidaan tunnistaa arvoa tuottavaa tietoa. Jotta data-analytiikan keinoja voidaan hyödyntää vaikuttavuusperusteisen sosiaali- ja terveydenhuollon luomiseksi, tarvitaan viitekehys, joka integroi analytiikan prosessin kustannusvaikuttavuuden edistämiseen (Naqishbandi & Ayyanathan 2019). Data-analytiikalla on tärkeä rooli siinä, että saadaan luotua järjestelmä, jossa interventioiden ja koko järjestelmän vaikuttavuutta voidaan arvioida ja ennustaa systemaattisesti. Naqishbandi ja Ayyanathan (2019) loivat tutkimuksessaan big data-analytiikan viitekehyksen, jota on sovellettu myös tässä työssä esitettyyn viitekehykseen. Tässä työssä esitetty viitekehys mukailee Naqishbandi ja Ayyanathan (2019) esittämän viitekehyksen rakennetta ja teoriaa, mutta sisältöä on päivitetty vastaamaan tämän tutkimuksen tuloksia ja suomalaista sosiaali-ja terveydenhuollon järjestelmää. Alla oleva kuva (Kuva 5) osoittaa sosiaali- ja terveydenhuollon datan reitin sen keräämisestä aina lopputuotokseen asti eli sen käyttökohteisiin ja vaikuttavuusperusteisen sosiaali- ja terveydenhuollon muodostumiseen.



Kuva 5: Viitekehys (mukaillen Naqishbandi & Ayyanathan 2019)

Prosessi etenee viitekehyksen nuolten osoittamaa reittiä ja lähtee liikkeelle datan keräämisestä. Mallissa esitetään kliinisen datan lähteet ja sen ominaisuuksia. Keräämisen jälkeen data käsitellään esikäsittelyn menetelmillä, jonka jälkeen se siirtyy data-analytiikan arvoa luovaan prosessiin. Kaikki data ei kulje jokaisen vaiheen läpi, vaan kulkureitti määräytyy datasta haettavan hyödyn perusteella. Datasta saatavalla tiedolla on lukuisia käyttökohteita, joissa sitä voidaan hyödyntää muun muassa riskien tunnistamiseen, ongelmien ehkäisyyn ja palvelutarpeiden ennakointiin, päätöksentekoon sekä kustannusvaikuttavuuden parantamiseen.

Viitekehyksen tavoite on luoda sosiaali- ja terveydenhuollon ammattilaisille ja sidosryhmille malli, joka esittää miten dataa voidaan hyödyntää toiminnan edistämisessä. Tarkoituksena on myös osoittaa, mitä hyötyä data-analytiikasta ja massadatasta voi olla arkikustannusvaikuttavuuden kannalta. Viitekehyksen yleinen tavoite on vähentää erilaisia riskejä, mikä tekee päätöksenteosta lopulta nopeampaa ja perustellumpaa. Seuraavaksi tutustutaan tarkemmin viitekehykseen ja sen sisältöön.

4.2 Sosiaali- ja terveydenhuollon data ja sen hyödyntämisen haasteet

Vaikuttavuustutkimukseen saadaan tietoa pääasiassa terveydenhuollon potilastietojärjestelmistä, sosiaalihuollon asiakastietojärjestelmistä sekä Kelan ylläpitämistä Kanta-palveluista (Neittaanmäki & Lehto 2018). Tietojärjestelmät sisältävät monenlaista dataa kuten kliinisiä terveystietoja, tutkimustuloksia, asiakas-/potilaskyselytuloksia, kustannustietoja ja muita tietoja, joita hoidoista tai asiakkaista/potilaista on relevanttia kerätä (Neittaanmäki & Lehto 2018). Asiakas- ja potilastietojärjestelmien on täytettävä niille annetut käyttötarkoituksenmukaiset vaatimukset ja esimerkiksi Valvira valvoo ja edistää näiden vaatimusten toteutumista (Valvira 2023). Vaatimukset liittyvät toiminnallisiin vaatimuksiin, yhteen toimivuuteen sekä tietosuojaan ja -turvaan (Valvira 2023). Kerätyn tiedon hyödyntäminen ei siis ole yksiselitteisen helppoa vaan siihen liittyy useita haasteita niin vaatimusten kuin myös datan määrän ja luonteen vuoksi.

Sosiaali- ja terveydenhuollon vaikutusten ja kulujen mittaaminen jokaisen asiakkaan/potilaan osalta on haaste, joka liittyy vahvasti big dataan. Terveydenhuollon monimutkainen dataekosysteemi, joka yhdistää sekä inhimillisiä komponentteja että vuorovaikutusta tietojärjestelmien kanssa edellyttää työskentelyä sosioteknisessä ympäristössä. Tämä tarkoittaa sitä, että tietojärjestelmien suunnittelua ja tehokkuuden arviointia määrittävät muun muassa organisaatioiden tavat, kulttuuri ja johtaminen sekä muut sosiaaliset näkökulmat. Lisäksi tulee osata huomioida big datan ominaisuudet ja sen 7V:tä. (Leung & van Merode 2018)

Ensinnäkin, jotta sosiaali- ja terveydenhuollon tietoa voidaan hyödyntää järjestelmää kehittävästi, tulee tietojärjestelmien tiedon olla laadukasta, ajantasaista sekä riittävän kattavaa. Sosiaali- ja terveydenhuollon data on pirstaloitunut lukuisiin eri järjestelmiin, jolloin tiedon löytäminen ja hyödyntäminen on integroituihin tietojärjestelmiin nähden huomattavasti haastavampaa. Tietojärjestelmien hajanaisuuden lisäksi haasteita syntyy, kun organisaation toiminnallisissa prosesseissa ja rakenteissa pääpaino on tiedon hankinnassa ja varastoinnissa. Tiedon jakamisen ja analysoinnin suunnittelu sekä tieto vaadittavista teknologioista tulisi pitää tiiviisti mukana prosesseissa ja johtamisessa, jotta riskit tiedon hyödyntämättä jäämisestä pienevät. (Salovaara, Leinonen ja Silén 2021)

Kolmantena tiedon hyödyntämiseen vaikuttaa massadatan strukturoimaton eli jäsentelemätön muoto. Sosiaali- ja terveydenhuollon data koostuu strukturoidun tiedon lisäksi suurilta

osin potilaskertomuksista, sähköposteista, viestiketjuista ja valokuvista, jotka luokitellaan strukturoimattomaksi dataksi (Müller, Junglas, Debortoli ja vom Brocke 2016). Strukturoimattoman datan haasteellisuus johtuu siitä, että se ei ole suoraan luettavissa tietokoneella, vaan sen hyödyntäminen vaatii ensin datan käsittelyä (Müller, Junglas, Debortoli ja vom Brocke 2016). Ei-rakenteisen tiedon hallintaa vaikeuttaa entisestään organisaation henkilöstön kognitiiviset puutteet toimia tiedon kanssa (Salovaara, Leinonen ja Silén 2021). Massadata luo datalajina jo omat haasteensa tiedon käsittelyyn, sillä suurien datamassojen hallitseminen ja prosessointi on perinteisin analytiikan keinoin haastavaa, vaikka tieto olisikin strukturoidussa tai puolistrukturoidussa muodossa (Garcia & Alvarez-Fernandez 2022).

Sosiaali- ja terveysalalla käsiteltävä tieto on usein arkaluontoista ja sisältää asiakkaiden ja potilaiden henkilökohtaisia terveystietoja. Asiakas- ja potilastietojen hallintaa ohjaavat EU:n tietosuoja-asetus, tietosuojalaki sekä useat potilaan ja asiakkaan asemaa suojaavat asetukset sekä lait (STM 2023). Vaikuttavuustutkimuksen kannalta huomioitavaa on myös vuonna 2019 vahvistettu laki sosiaali- ja terveystietojen toissijaisesta käytöstä. Niin kutsutun toisiolain tavoitteena on mahdollistaa sosiaali- ja terveysalan tietojen tehokas ja tietoturvallinen hyödyntäminen toisiokäytössä (STM 2023). Toisiokäytöllä tarkoitetaan tiedon hyödyntämistä muussa kuin sen alkuperäisessä hyödyntämistarkoituksessa (STM 2023). Laissa määritellyt soveltamisalat ovat tilastointi, tieteellinen tutkimus, kehittämis- ja innovaatiotoiminta, opetus, tietojohtaminen, sosiaali- ja terveydenhuollon viranomaisohjaus ja -valvonta sekä viranomaisen suunnittelu- ja selvitystehtävät (Laki sosiaali- ja terveystietojen toissijaisesta käytöstä 552/2019) Lainsäädännöllä ja asetuksilla on siis pyritty tehostamaan tiedon hyödyntämistä, mutta samalla se edellyttää erityistä tarkkuutta tiedon hallinnassa ja käsittelyssä.

Sosiaali- ja terveydenhuollon kasvavat kustannukset ja vähäiset resurssit ovat esteenä tietojärjestelmien ja niiden hallinnan kehittämiselle. Hyvinvointialueille myönnettiin Valtiovarainministeriön toimesta vuonna 2022 155,7 miljoonan euron lisämääräraha hajanaisten tietojärjestelmien muutosbudjettiin, sillä alkuperäinen 440 miljoonan euron budjetti ei riittänyt kehittämään tietojärjestelmiä hyvinvointialueiden toiminnan edellyttämälle tasolle (VM 2022). Lisämääräraha tarkoituksena on helpottaa järjestelmien yhteensovittamista nyt kun hyvinvointialueet ovat tammikuussa 2023 aloittaneet toimintansa. Salovaara, Leinonen ja Silén (2021) totesivat tutkimuksensa nojalla resurssien olevan julkisessa terveydenhuollossa

suurin este tiedon hyödyntämiselle, kun taas yksityisellä sektorilla merkittävimmäksi haasteeksi nostettiin puutteet henkilöstön osaamisessa.

4.3 Datasta päätöksentekoon

Sosiaali- ja terveydenhuollon tietojärjestelmät koostuvat useista osatietojärjestelmistä, joihin on talletettu tietoja eri tarpeiden mukaan. Merkittävimmässä roolissa on asiakas- ja potilastietojärjestelmät, jonka vuoksi englanninkielisessä tutkimuksessa ja kirjallisuudessa terveydenhuollon tietojärjestelmistä käytetään yleisesti termiä EHR. EHR-järjestelmät (electronic health records) on suunniteltu tarjoamaan tapahtumatietojen tallennus- ja raportointimahdollisuuden, joka tukee terveydenhuoltoprosessia. Dokumentointi tietojärjestelmiin, eli esimerkiksi potilastietojen, käyntien, tunnistetietojen, diagnoosien ja lääkitysten kirjaaminen, tapahtuu pitkälti vielä manuaalisesti tai puolimanuaalisesti. Tiedot ovat siis alttiita inhimillisille virheille ja ne voivat olla puutteellisia tai ei-ajantasaisia. Lisäksi dokumentoidun tiedon ominaisuudet kuten muoto haittaavat sen hyödyntämistä analytiikan keinoin sellaisenaan. (Aguirre, Suarez, Fuentes & Sanchez-Gonzalez 2019)

Viitekehyksen mukaan ennen varsinaista analytiikan prosessia data tulee viedä esikäsittelyprosessin läpi, jolloin se saadaan paremmin analysoitavaan muotoon (Ferrão, Oliveira, Janela & Martins 2016). Lin ja Haug (2006) mukaan terveydenhuollon datan esikäsittely voidaan toteuttaa joko manuaalisesti tai se voidaan automatisoida. Heidän tutkimuksensa osoittaa, että esikäsittelyn automatisoinnilla saavutetaan parempia tuloksia verrattuna täysin manuaaliseen prosessiin. Suuri osa tietojärjestelmiin kerätyistä diagnooseista, hoidoista ja lääkemääräyksistä sisältävät epäselvyyksiä ja virheitä, jotka tulee korjata tai poistaa puhdistuksen, suodatuksen, täydentämisen tai standardoinnin avulla (Lin & Haug 2006).

Data-analytiikan prosessin ensimmäinen vaihe on kuvaileva eli deskriptiivinen analytiikka. Deskriptiivinen analytiikka on ollut tyypillisin analytiikan taso terveydenhuollossa, sillä sen ja tilastojen nojalla on voitu muodostaa käsitys siitä, mitä on tapahtunut ja mitä interventioilla on saavutettu (Ma 2020). Siinä datan karakterisoimisen, aggregoimisen ja luokittelun avulla muodostetaan terveydenhuollon ammattilaisille tietoa suorituskyvystä ja analyyseja päätöksenteon tueksi (Alghamdi, Alsubait, Baz & Alhakami 2021). Deskriptiivinen analytiikka auttaa kertomaan esimerkiksi, mikä on tietyn leikkauksen onnistumisprosentti ja mitkä

merkit tuloksissa viittaavat siihen, että potilas tarvitsee jatkohoitoa (Naqishbandi & Ayyanathan 2019).

Luokittelu on yksi tärkeistä kuvailevan analytiikan menetelmistä terveydenhuollossa. Se kuuluu koneoppimisen menetelmiin ja siinä terveydenhuollon data kuvataan ennalta määrättyihin luokkiin, jotka on helpompi ymmärtää sekä analysoida (Alghamdi, Alsubait, Baz & Alhakami 2021). Terveydenhuollossa luokittelu on tärkeää, sillä sen avulla voidaan vähentää päällekkäisyyksiä, jotka vaikeuttaisivat potilaisiin ja järjestelmiin liittyvien tapahtumien jäljittämistä (Alghamdi, Alsubait, Baz & Alhakami 2021). Deskriptiivinen analytiikka ei kuitenkaan yksinään riitä, sillä usein on tarve kohdistaa toimintaa pidemmälle aikaikkunalle ja tarvitaan muita analytiikan tasoja.

Diagnostisesta analytiikasta on apua, kun terveydenhuollon organisaatio tavoittelee siirtymistä kokonaan datalähtöisyyteen ja se mahdollistaa kilpailullisen edun tavoittelun. Diagnostinen analytiikka perehtyy dataan yrittäen löytää syitä tapahtumille ja sen avulla voidaan saada vastaus esimerkiksi kysymykseen ”miksi potilaan tila huononi, vaikka hän sai parasta mahdollista hoitoa?”. Diagnostisessa analytiikassa prosessit liittyvät muun muassa piilotettujen rakenteiden löytämiseen, tietojen yhdistelyyn ja keräämiseen sekä analysointiin. Prosesseihin sisältyy myös potilas- ja asiakastietojen analysoiminen ja analyysin pohjalta muodostettava kuva potilas- tai asiakasryhmästä. (Alghamdi, Alsubait, Baz & Alhakami 2021)

Prediktiivinen analytiikka ja tulevaisuuden ennustaminen ovat yleistyneet nopeasti terveydenhuollossa. Ennakoivan analytiikan avulla voidaan ennustaa riskejä, laatia potilaiden ennusteita sekä arvioida ja ennakoida jatkohoidontarvetta (Naqishbandi & Ayyanathan 2019). Terveydenhuollossa prediktiivistä analytiikkaa hyödynnetään jo monipuolisemmin kuin sosiaalihuollossa. Iso-Britanniassa ja Pohjois-Irlannissa toteutetussa tutkimuksessa selvitettiin prediktiivisen analytiikan hyödyntämismahdollisuuksia lastensuojelussa, kun tutkittiin asiakkaan hyväksikäytön kohteeksi joutumisen riskien tunnistamista (Redden 2020). Tutkimuksessa nousi esille tarve lisäselvitykselle tiedon kirjaamisen ja luotettavuuden osalta ja siinä huomattiin muun muassa tietoja kirjaavien työntekijöiden asenteiden aiheuttavan pahimmillaan asiakkaiden syrjintää (Redden 2020). Prosessissa ja erityisesti asiakastietojen kirjaamisessa on vielä kehitettävää, mutta tulevaisuudessa prediktiivisestä analytiikasta saadaan yhä enemmän tukea myös sosiaalihuollon päätöksentekoon.

Ohjaileva analytiikka vastaa kysymykseen, mitä tulisi tehdä. Terveydenhuollossa havainnollistavia ohjailevan analytiikan kohteita olisivat esimerkiksi kysymykset siitä, mitä toimenpiteitä leikkauksen onnistumisprosentin kasvattamiseen tarvitaan ja mitä lääkkeitä määrätään, jotta toipuminen tehostuu (Naqishbandi & Ayyanathan 2019). Ohjaileva analytiikka hyödyntää dataa ja matemaattisia algoritmeja määritellyn ongelman ratkaisemiseen ja terveydenhuollossa se tarjoaa konkreettisia ratkaisuja, joiden avulla potilaan hyvinvointia saadaan edistettyä (Alghamdi, Alsubait, Baz & Alhakami 2021).

Data-analytiikan monivaiheisen prosessin avulla voidaan siis tehostaa prosesseja ja luoda varmuutta sosiaali- ja terveydenhuollon päätöksentekoon. Päätöksenteon varmistuessa terveyden- ja sosiaalihuollon ammattilaiset kykenevät tekemään ratkaisuja nopeammin ja olemaan varmempia päätöstensä oikeellisuudesta. Data-analytiikan keinoja voidaan päätöksentekoprosessissa hyödyntää myös priorisoinnin tukena ja niiden avulla voidaan alentaa hoidoista ja palveluista muodostuvia kokonaiskustannuksia (Naqishbandi & Ayyanathan 2019). Usein laajempien määrällisten kuten myös laadullisten tavoitteiden saavuttaminen vaatii useamman eri analytiikan tason soveltamista.

4.4 Esimerkkejä data-analytiikan avulla tuotetusta kustannusvaikuttavuudesta

Sosiaali- ja terveydenhuollon datan lähteiden, ominaisuuksien, analyyttisen prosessin sekä jalostetun tiedon hyödyntämiskohteiden jälkeen tutustutaan, miten viitekehyksen mukaista prosessia voidaan hyödyntää kustannusvaikuttavuuden edistämiseksi ja arvioinnin parantamiseksi. Alle on koottu esimerkkejä, joissa kustannusvaikuttavuutta on edistetty monipuolisesti eri menetelmin sekä eri käyttötarkoituksiin. Esimerkit on valittu niin, että ne tuovat esille data-analytiikan mahdollisuuksia eri vaikuttavuuden tasoilla.

Ensimmäinen esimerkki käsittelee toimenpiteen kustannusvaikuttavuuden edistämistä. Kyseessä on Yhdysvalloissa toteutettu tutkimus, jossa hyödynnettiin ennakoivaa ja ohjailevaa analytiikkaa, kun pyrittiin tuottamaan potilaille reaaliaikaista ja personoitua hoitoa. Tarkoituksena oli lyhentää sairaalassaoloaikaa leikkaushoidon jälkeen ja ehkäistä potilaan palaamista sairaalaan kyseisestä toimenpiteestä johtuvan syyn takia. Analytiikan metodit, joita

tutkimuksessa hyödynnettiin, perustuivat koneoppimiseen (ennakoiva) ja veren punasoluarvojen moduloimiseen (ohjaileva). Ennakoivaan analytiikkaan perustuva koneoppimismalli kehitettiin tunnistamaan merkittävimmät muuttujat, jotka osoittivat todennäköisyyttä sille, että potilas joutuu palaamaan takaisin sairaalaan. Punasolujen määrästä luotiin ennakoiva malli (predictive model), sillä hemoglobiinin liian alhaisen arvon on tutkittu johtavan kohonneeseen komplikaatioriskiin ja näin ollen hidastavan sairaalasta kotiutumista. Näiden ennakoivien mallien avulla voitiin personoida potilaan hoitoa ja konkreettisin teoin lyhentää sairaalassaoloaikaa. Tutkimuksessa tultiin siihen lopputulokseen, että siinä esitettyjen toimenpiteiden mukaan tuotettu personoitu hoito ja takaisinottoasteen pieneneminen voisivat vähentää terveydenhuollon kustannuksia Yhdysvalloissa jopa yli 20 miljoonaa dollaria vuosittain. Yksittäisen toimenpiteen kustannusvaikuttavuuden tehostuminen vaikuttaa siis myös koko järjestelmään. (Bertsimas, Li, Paschalidis & Wang, 2020).

Toinen esimerkki käsittelee vaikuttavuuden seuraavaa tasoa eli palveluketjun kustannusvaikuttavuutta. Sosiaalihuollon kohdalla prediktiivisestä analytiikasta saadaan apua ennaltaehkäisyyn. Kun verrataan lapsen kodin ulkopuolelle sijoittamisen kustannuksia ennaltaehkäiseviin toimiin, saadaan selville Lastensuojelun keskusliiton selvityksen mukaan, että lapsen sijoittaminen sijaisperheeseen maksaa 22 000 euroa vuodessa, ammatilliseen perhekotiin noin 60 000 euroa vuodessa ja laitoshoito yli 90 000 euroa vuodessa (Heinonen, Väisänen & Hipp 2012, s. 8). Ehkäisevän lastensuojelun toimet kattavat laajasti hyvinvointialueen perus- ja erityispalveluita lapsen kasvun, kehityksen ja hyvinvoinnin tukemiseksi (Heinonen, Väisänen & Hipp 2012). Tästä syystä ehkäisevän lastensuojelun kustannuksia ja kustannusvaikuttavuutta on haastavampi arvioida, mutta kustannukset jäävät reilusti lapsen sijoittamista pienemmiksi (Heinonen, Väisänen & Hipp 2012).

Analytiikan avulla voidaan tunnistaa asiakas ja huomata riskit jo varhaisessa vaiheessa sekä ohjata asiakas ennaltaehkäisevien palvelujen pariin (Redden 2020). Riskien tunnistaminen ja ennustaminen pohjautuvat prediktiivisen analytiikan keinoihin, joista esimerkkinä on Los Angelesin lasten- ja perhepalvelujen yksikön sekä ohjelmistoyritys SAS:in pilotoima malli (Church & Fairchild 2017). Pilottihankkeessa luotu algoritmi arvioi jokaisen asiakkaan tilanteen numerolla nollan ja 1000 välillä, joista riskiarvion 900 ylittävä arvo indikoi vakavinta hyväksikäytön riksiä. (Church & Fairchild 2017) Algoritmin heikkous on kuitenkin siinä, että se arvioi parhaiten vain epätodennäköisimpiä tilanteita ja yksittäisen numeroarvion on katsottu olevan liian suppea arvioinnin keino. (Church & Fairchild 2017) Eli kuten

Reddenin (2020) artikkelissa todettiin, tarvitaan sosiaalihuollon datan hyödyntämiseen vielä lisää osaamista ja tutkimusta.

Kolmas esimerkki huomioi jo laajemman kokonaisuuden ja se keskittyy monialaisia palveluja tarvitsevien tunnistamiseen ja ennakoiviin malleihin. Tällä hetkellä pieni osuus suomalaisista käyttää suurimman osan sosiaali- ja terveydenhuollon palveluista. Palvelutarvetta ja eniten kustannuksia aiheuttavat diabetes, sydän- ja verisuonitaudit, usean kroonisen sairauden sairastaminen sekä päihde- ja mielenterveysongelmat. Sosiaali- ja terveydenhuollon palvelujen yhteensovittamisella uskotaan olevan positiivisia vaikutuksia paljon palveluja tarvitsevien aikaisessa tunnistamisessa ja oikeanlaiseen hoitoon ohjaamisessa. Varhainen tunnistaminen ja ennakointi vähentävät tarvetta palveluille tulevaisuudessa ja näin ollen syntyy säästöjä kustannuksissa. Julkaisussa tutkittiin jo käytössä olevia ja potentiaalisia monialaisia palveluja tarvitsevien tunnistamisen ja ennakoinnin malleja. Selvityksen mukaan useimmat malleista olivat hyvin geneerisiä ja näin ollen hyödynnettävissä myös Suomessa. (Kivipelto, Suhonen, Koivisto, Tiirinki & Miikki, 2019)

Tunnistamiseen ja ennakointiin hyödynnettiin kattavasti eri analytiikan vaiheita ja keinoja. Henkilön tunnistamisen keinot kattavat deskriptiivisen ja diagnostisen analytiikan keinoja ja sisältävät muun muassa asiakas- ja potilastietojen analysointia, kyselypohjaisia mittareita ja niiden tulosten analysointia, asiakasprofilointia ja tiedonlouhintaa. Ennakoinnissa hyödynnettiin pitkälti vielä tilastollisia menetelmiä, mutta selvityksessä nostettiin esille yksi prediktiivisen analytiikan keinoja hyödyntävä systeemi. Kroatiassa kehitetyn GeroS-systeemin yhtenä tehtävänä on laatia tulevaisuuden ennusteita ikääntyvien palvelujen tarpeesta. Palvelu sisältää tiedon kaikista ikääntyneistä eikä vain esimerkiksi kotihoidon piirissä olevista, mikä voidaan nähdä sekä haasteena että etuna. Sillä palvelu sisältää paljon tietoa suuresta joukosta ihmisiä, on se raskas käyttää ja vie paljon resursseja. Palvelun käyttömahdollisuudet ovat kuitenkin kattavat ja sen laajan tietokannan avulla voidaan edistää niin yksittäisten henkilöiden kuin koko kansan terveyttä. (Kivipelto, Suhonen, Koivisto, Tiirinki & Miikki 2019)

5 Johtopäätökset

Tämän kandidaatintyön tavoitteena oli tutkia data-analytiikan hyödyntämismahdollisuuksia sosiaali- ja terveyspalvelujen kustannusvaikuttavuuden edistämisessä ja esitellä viitekehys, jonka avulla voidaan havainnollistaa kustannusvaikuttavuuden muodostumista. Työssä vastattiin tavoitteen mukaan määriteltyyn päätutkimuskysymykseen, joka oli:

- *Miten sosiaali- ja terveyspalvelujen kustannusvaikuttavuutta voidaan edistää data-analytiikan avulla?*

Työn tuloksena huomattiin, että data-analytiikan mahdollisuudet ovat hyvin moninaiset ja eri menetelmiä voidaan hyödyntää vaikuttavuuden eri tasoilla. Työssä esitettiin kolme esimerkkiä, joissa kustannusvaikuttavuutta oli edistetty datan ja analytiikan menetelmin. Data-analytiikan keinoista ennakoiva mallintaminen (predictive modeling) toistui jokaisessa esimerkissä. Sosiaalityön ja lastensuojelun kohdalla tunnistettiin tarve lisätutkimuksille analytiikan hyödyntämisestä. Työssä esitetyt esimerkit on tiivistetty alla olevaan taulukkoon (Taulukko 3).

Taulukko 3: Tiivistelmä työn esimerkeistä

Esimerkki	**Sairaalassaoloajan lyhentäminen leikkaushoidon jälkeen ja personoitu hoito**	**Lastensuojelun kustannusvaikuttavuus, lapsen kodin ulkopuolelle sijoittaminen**	**Moninaisen palveluntarpeen tunnistaminen ja ennakoiminen, GerosS-systeemi**
Vaikuttavuuden taso	Toimenpiteen kustannusvaikuttavuus	Palveluketjun kustannusvaikuttavuus	Järjestelmän kustannusvaikuttavuus
Data-analytiikan tasot	Prediktiivinen ja ohjaileva analytiikka	Prediktiivinen analytiikka	Kuvaileva, diagnostinen ja prediktiivinen analytiikka
Analytiikan menetelmät	Koneoppiminen ja ennakoiva malli veren punasolujen määrän moduloimiseen	Ennakoivat mallit ja algoritmit, riskiarviot	Asiakasprofilointi, tiedonlouhinta, kyselypohjaisten mittareiden tulosten arviointi, ennakoivat mallit
Vaikutukset kustannusvaikuttavuuteen	Reaaliaikainen ja personoitu hoito vähentävät sairaalassaoloaikaa leikkaushoidon jälkeen ja ehkäisevät tarvetta palata takaisin sairaalaan. Potilaan lyhyempi sairaalassa vietetty aika säästää kustannuksia.	Keskitytään ehkäiseviin palveluihin ja ehkäistään tarvetta sijoittaa lapsi kodin ulkopuolelle. Kodin ulkopuolelle sijoittaminen on kallista ja ehkäisevät palvelut palvelevat usein paremmin lapsen etua.	Paljon palveluja tarvitsevien tunnistaminen ja oikeanlaiseen hoitoon ohjaaminen aikaisessa vaiheessa, jolloin syntyy säästöjä kustannuksissa ja hoito on vaikuttavampaa.

Päätutkimuskysymyksen lisäksi työssä esitettiin kolme apututkimuskysymystä, jotka auttoivat päätutkimuskysymykseen vastaamisessa ja työn rakenteen jäsentelyssä. Ensimmäinen apututkimuskysymys oli:

- *Miten kustannusvaikuttavuus muodostuu sosiaali- ja terveyspalveluissa?*

Vaikuttavuus kuvaa muutosta potilaan tai asiakkaan terveydentilassa tai hyvinvoinnissa, joka on saatu aikaan tuotetuilla palveluilla. Vaikuttavuuden voidaan katsoa muodostuvan neliportaisen vaikuttavuusketjun kautta, joka alkaa käytetystä panoksesta kuten resursseista. Toinen askelma on panoksen avulla saavutettu tuotos ja kolmantena tuotoksen vaikutus. Kolme edellä mainittua askelmaa muodostavat yhdessä vaikuttavuuden. Kustannusvaikuttavuus sisällyttää vaikuttavuuden tarkasteluun terveystaloustieteellisen näkökulman ja kustannusten tarkastelun. Kustannusvaikuttavuus saadaan, kun vaikuttavuus jaetaan sen aikaansaamiseen käytetyillä kustannuksilla. Vaikuttavuutta voidaan tutkia joko kokeellisessa ympäristössä satunnaistetussa kontrolloidussa kokeessa tai havainnoivan vertaiskontrolloidun tutkimuksen avulla. Vertaiskontrolloitua tutkimusta käytetään, kun tutkimusryhmien satunnaistaminen ei ole eettisesti kannattavaa.

Vaikuttavuuden mittaamisen pohjana on laaja joukko eri mittareita. Tässä tutkielmassa mittarit ryhmiteltiin tiedon lähteen mukaan. Vaikuttavuuden mittaamisessa ja mittaristossa on siirrytty huomioimaan yhä enemmän asiakkaiden ja potilaiden itse raportoimia tuloksia eli PROM-mittareilla saatuja tuloksia. Kustannuksia voidaan tutkia tarkemmin kustannusvaikuttavuusanalyysin, kustannus-utitileettianalyysin ja kustannushyötyanalyysin avulla. Kustannusvaikuttavuusanalyysissä merkittävä tunnusluku on inkrementaalinen kustannusvaikuttavuus, joka kuvaa vertailutoimenpiteen tuomaa kustannusvaikuttavuuden kasvua verrattuna indeksitoimenpiteeseen. Kustannus-utiliteettianalyysissä kustannusvaikuttavuutta kuvataan elämänlaatua kuvaavien mittareiden avulla. Kustannushyötyanalyysi perustuu terveyshyötyjen arvottamiseen rahallisesti, mikä voidaan nähdä osittain ongelmallisena. Työn toinen apututkimuskysymys oli:

- *Millainen on yksinkertaistettu data-analytiikan prosessi?*

Datan ominaisuudet ja käyttökohteet ovat hyvin laajoja, minkä vuoksi tässä työssä pyrittiin yksinkertaistamaan monimutkaista prosessia helpommin ymmärrettävään muotoon. Dataa, joka on määrältään suurta ja rakenteeltaan haasteellista sekä nopeasti muuttuvaa kutsutaan big dataksi. Big datan jalostamiseen vaikuttavat sen seitsemän ominaisuutta eli 7 V:tä, jotka

ovat määrä, moninaisuus, nopeus, arvo, todenmukaisuus, vaihtelevuus ja visualisointi. Datan jalostamisen ensimmäinen ja samalla eniten resursseja vievä vaihe on esikäsittely. Esikäsittelyn avulla dataa siivotaan, muokataan ja järjestellään niin, että sitä on helpompi käsitellä analytiikan prosessin mukaan. Prosessin ensimmäinen vaihe eli toisin sanoen ensimmäinen analytiikan taso on deskriptiivinen analytiikka. Se vastaa kysymykseen siitä, mitä tapahtui. Toinen taso eli diagnostinen analytiikka pyrkii etsimään syy-seuraussuhteita ja löytämään vastauksen siihen, että miksi jotain tapahtui. Prediktiivisen analytiikan avulla voidaan luoda ennusteita siitä, mitä tulee tapahtumaan. Ohjaileva analytiikka on jo edistynyt analytiikan taso ja sen avulla etsitään konkreettisia ratkaisuja sille, miten jokin saadaan tapahtumaan. Työn kolmas ja viimeinen apututkimuskysymys oli:

- *Mitä data-analytiikan menetelmiä voidaan hyödyntää kustannusvaikuttavuuden edistämisessä?*

Analytiikan menetelmiä valitessa tulee ottaa huomioon sosiaali- ja terveydenhuollon datan hyödyntämiseen liittyvät haasteet. Merkittävimpiä haasteita ovat hajanaiset tietojärjestelmät, tiedon määrä ja muoto, henkilöstön osaaminen sekä puutteelliset resurssit. Sosiaali- ja terveydenhuollon datan käsittelyä ohjaavat myös useat lait ja asetukset. Lait, kuten toisiolaki, pyrkivät asiakkaiden ja potilaiden oikeuksien turvaamisen lisäksi helpottamaan tiedon monipuolisempaa hyödyntämistä.

Sosiaali- ja terveydenhuollon data tulee esikäsitellä, jotta se saadaan helpommin analysoitavaan muotoon. Esikäsittely voidaan tehdä joko manuaalisesti tai hyödyntäen automaatiota. Käsittelyn automatisoinnin on todettu olevan tehokkaampaa ja sen avulla voidaan saavuttaa parempia tuloksia verrattuna täysin manuaaliseen käsittelyyn. Esikäsittelyn jälkeen ensimmäinen analytiikan prosessin vaihe on kuvaileva analytiikka. Kuvaileva analytiikka on ollut terveydenhuollossa pitkään tyypillisin analytiikan taso. Sen avulla on voitu selvittää mitä hoidossa on tapahtunut ja mitä interventioilla on saatu aikaan.

Diagnostisen analytiikan avulla voidaan löytää syy-seuraussuhteita hoidon ja sen vaikutusten välillä. Potilaiden ja asiakkaiden luokittelu tapahtuu pitkälti diagnostisen analytiikan keinoin. Kolmas analytiikan taso on prediktiivinen analytiikka, joka on noussut nopeasti tärkeään rooliin. Sen merkittävimmät hyödyntämiskeinot ovat ennusteiden laatiminen, riskien tunnistaminen sekä jatkohoidontarpeen arviointi. Ohjailevan analytiikan keinoista terveydenhuollossa hyödynnetään muun muassa matemaattisia algoritmeja, joiden avulla etsitään

konkreettisia ratkaisuja ongelmiin. Ohjailevaa analytiikkaa voidaan hyödyntää esimerkiksi, kun halutaan selvittää millä toimenpiteillä jonkin intervention onnistumisprosentti saadaan kasvamaan.

Tämän tutkimuksen nojalla voidaan todeta, että vaikuttavuuden ja kustannusvaikuttavuuden arviointi on vahvasti sidoksissa sosiaali- ja terveydenhuollon dataan. Tietojärjestelmät ovat täynnä tietoa, jota voidaan jalostaa vaikuttavuuden tutkimiseen ja edistämiseen. Nykytilanteessa on kuitenkin vielä lukuisia haasteita, jotka heikentävät kerätyn tiedon hyödyntämistä. Merkittävimmät haasteet ovat analytiikan ja datan käsittelyn osaamisessa, tietorakenteissa sekä mittaamisessa. Tätä työtä varten tehdyn selvityksen perusteella mittareiden käyttö ei ole yhtenäistä. Työssä esitellyissä tutkimuksissa sosiaali- ja terveydenhuollossa on hyödynnetty innovatiivisesti data-analytiikkaa, jonka avulla on saatu parannettua arkikustannusvaikuttavuutta. Tutkimuksissa pääpaino on analytiikassa eikä niissä tutkimuksen asettelun tai tulosten perusta ole yleisissä mittareissa.

Tulevaisuudessa kustannusvaikuttavuuden arvioinnissa tulisi siis entistä enemmän yhdistää sekä yhtenäisiä mittareita että analytiikkaa. Valtakunnallisten vaikuttavuustavoitteiden täyttyminen vaatii, että keräämisen lisäksi dataa osataan myös hyödyntää uusilla keinoilla. Tulevaisuuden tutkimuskohteena voidaan myös mainita sosiaalihuollon kustannusvaikuttavuuden arviointi. Tutkimuksen tavoitteena voisi olla vaikuttavamman rakenteellisen sosiaalityön kehittäminen entistä monipuolisemman ja tehokkaamman analytiikan avulla.

Lähteet

Agrawal, R., Kadadi, A. 2022. Data Integration. In: Schintler, L.A., McNeely, C.L. (eds) Encyclopedia of Big Data. Springer, Cham.

Aguirre, R., Suarez, O., Fuentes, M., Sanchez-Gonzalez, M. 2019. Electronic Health Record Implementation: A Review of Resources and Tools. *Cureus*, *11*(9), e5649.

Aistrich, M. 2014. Kannattaako vaikuttavuutta yrittää mitata? Sitra. Viitattu: 17.3.2023. Saatavilla: https://www.sitra.fi/artikkelit/kannattaako-vaikuttavuutta-yrittaa-mitata/

Alghamdi, A., Alsubait, T., Baz, A. and Alhakami, H. 2021. Healthcare Analytics: A Comprehensive Review, *Engineering, Technology & Applied Science Research*. Greece, 11(1), s. 6650–6655.

Balali, F., Nouri, J., Nasiri, A., Zhao, T. 2020. Data Analytics. In: Data Intensive Industrial Asset Management. Cham: Springer.

Bang, H., & Zhao, H. 2012. Median-Based Incremental Cost-Effectiveness Ratio (ICER). Journal of statistical theory and practice, 6(3), s. 428–442.

Barnes, T. 2013. Big data, little history. *Dialogues in Human Geography*, *3*(3), s. 297-302.

Bertsimas, D., Li, M. L., Paschalidis, I. C., & Wang, T. 2020. Prescriptive analytics for reducing 30-day hospital readmissions after general surgery. *PloS one*, *15*(9), e0238118.

Bhadani, A., Jothimani, D. 2016. Big data: Challenges, opportunities and realities. In Singh, M.K., & Kumar, D.G. (Eds.), Effective Big Data Management and Opportunities for Implementation. Pennsylvania, USA, s.1-21.

Black, N. 2013. Patient reported outcome measures could help trans-form healthcare. BMJ (Clinical research ed), s. 346 . ISSN 0959-8138

Booth, N., Aronen, P., Mäkelä, M. 2017. Selvitys kustannusvaikuttavuuden käyttämisestä yhtenä terveydenhuollon palveluvalikoiman määrittelykriteerinä. Sosiaali- ja terveysministeriön raportteja ja muistioita 2017:30.

Church, Christopher E., & Fairchild, Amanda J. 2017. In search of a silver bullet: Child welfare's embrace of predictive analytics. *Juvenile & Family Court Journal*, *68*(1), s. 67–81.

Digitaalinen Helsinki. 2023. Käsitteistö. Datastrategian lähtökohdat. Viitattu 2.4.2023. Saatavilla: https://digi.hel.fi/esittely/helsinki-datastrategia/helsinki-datastrategia-tiivistelma/11-käsitteistö/

Doyle, C., Lennox, L., Bell, D. 2013. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ

Drummond, MF., Sculpher, MJ., Torrance, GW., O'Brien, BJ., Stoddart, GL. 2005. Methods for the economic evaluation of health care programme. Third edition. Oxford University Press.

Eberendu, A. 2016. Unstructured Data: an overview of the data of Big Data. International Journal of Computer Trends and Technology, 38(1), s. 46-50.

Ferrão, J. C., Oliveira, M. D., Janela, F., & Martins, H. M. 2016. Preprocessing structured clinical data for predictive modeling and decision support. A roadmap to tackle the challenges. *Applied clinical informatics*, *7*(4), s. 1135–1153.

Garcia, S., Ramírez-Gallego, S., Luengo, J., Benítez, J., Herrera, F. 2016. Big data preprocessing: methods and prospects.

González, C., Álvarez-Fernández, E. 2022. What Is (Not) Big Data Based on Its 7Vs Challenges: A Survey. Big Data and Cognitive Computing. 6. s. 158.

Greasley, A. 2019. Simulating Business Processes for Descriptive, Predictive, and Prescriptive Analytics.

Heinonen, H., Väisänen, A., & Hipp, T. 2012. Miten lastensuojelun kustannukset kertyvät?. Terveyden ja hyvinvoinnin laitos. ISBN 978-951-9424-92-7

Gröhn, A., Ranta, J. 2022. Syrjäyttääkö ohjaava analytiikka ohjaajan tulevaisuudessa? Vartu 2.0 -hanke. Saatavilla: https://www.karelia.fi/2022/06/syrjayttaako-ohjaava-analytiikka-ohjaajan-tulevaisuudessa/

Heiskanen, J., Rannanheimo, P., Härkönen, U. 2018. Terveyteen liittyvä elämänlaatu – mitä oikeasti mittaamme ja miten hyödynnämme kerättyä tietoa?. Fimea.

Heliskoski, J., Humala, H., Kopola, R., Tonteri, A., Tykkyläinen, S. 2018. Vaikuttavuuden askelmerkit. Työkaluja ja esimerkkejä palveluntuottajille. Sitran selvityksiä 130.

Hinkka-Yli-Salomäki, S. 2021. Vaikuttavuustutkimus: Satunnaistettu vertailukoe tulee suunnitella huolella. Kasvun tuki. Viitattu 7.3.2023 Saatavilla: https://kasvuntuki.fi/aikakauslehti/julkaisut/kasvun-tuki-aikakauslehti-22021/hinkka-yli-salomaki/

Ikonen, T., 2019. Sosiaali- ja terveysministeriön raportteja ja muistioita 2019:70. Asiakasryhmäkohtainen tieto laadusta ja vaikuttavuudesta sosiaali- ja terveydenhuollon tiedolla johtamisessa ja ohjauksessa – Laaturekisterien asema palvelujärjestelmässä. ISSN pdf: 2242–0037

Javatpoint. 2023. Semi-Structured Data. Viitattu 3.4.2023. Saatavilla: https://www.javatpoint.com/semi-structured-data

Jeyanthi, P. 2022. Decision intelligence analytics and the implementation of strategic business management. Cham, Switzerland: Springer.

Jonsson, P. M., Pikkujämsä. S., Heiliö, P. L (toim.). 2019. Kansalliset laaturekisterit sosiaali- ja terveydenhuollossa. Toimintamalli, organisointi ja rahoitus. Terveyden ja hyvinvoinnin laitos

Katal, A., Wazid, M., Goudar, R. 2013. Big data: Issues, challenges, tools and Good practices. Sixth International Conference on Contemporary Computing (IC3), Noida, India. s. 404-409.

Kivipelto, M., Suhonen, M., Koivisto, J., Tiirinki, H., Miikki, R. 2019. Monialaisia palveluja tarvitsevien tunnistamisen ja ennakoinnin mallit: kartoittava kansainvälinen katsaus. Terveyden ja hyvinvoinnin laitos. Helsinki.

Laine, J. 2021. Kustannusvaikuttavuus, arvoperusteinen terveydenhuolto ja arvon arviointi. Suomen Farmasialiitto. Dosis 37. s. 430–443.

Laki potilaan asemasta ja oikeuksista. 17.8.1992/785. Finlex. Viitattu 22.4.2023. Saatavilla: https://finlex.fi/fi/laki/ajantasa/1992/19920785

Laki sosiaali- ja terveystietojen toissijaisesta käytöstä. 26.4.2019/552. Finlex. Viitattu Saatavilla: 17.3.2023. https://www.finlex.fi/fi/laki/alkup/2019/20190552

Laki sosiaalihuollon asiakkaan asemasta ja oikeuksista. 22.9.299/812. Finlex. Viitattu 22.4.2023. https://finlex.fi/fi/laki/ajantasa/2000/20000812

Leung, T. I., & van Merode, G. G. 2018. Value-Based Health Care Supported by Data Science. In P. Kubben (Eds.) et. al., *Fundamentals of Clinical Data Science*. s. 193–212. Springer.

Lin, J. H., & Haug, P. J. 2006. Data preparation framework for preprocessing clinical data in data mining. Annual Symposium proceedings. AMIA Symposium, 2006, s. 489–493.

Liu, J., Sun, J., Wang, S. 2006. Pattern Recognition: An Overview. IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.6.

Ma A. C. 2020. Making Data Reports Useful: From Descriptive to Predictive. Cureus, 12(10), e10920.

Malmivaara, A. 2022. Vaikuttavuus sosiaali- ja terveydenhuollossa. 1. painos. Helsinki: Duodecim.

Miller, J. 2017. Big Data Visualization. 1. painos. Birmingham, England: Packt Publishing.

Müller, O., Junglas, I., vom Brocke, J.,Debortoli, S. 2016. Utilizing big data analytics for information systems research: challenges, promises and guidelines. European Journal of Information Systems. 25:4, s. 289-302.

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. 2014. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC health services research.

Naqishbandhi, T., Ayynathan, N. 2019. Clinical Big Data Predictive Analytics Transforming Healthcare: - An Integrated Framework for Promise Towards Value Based Healthcare.

Nilsson, E., Orwelius, L., & Kristenson, M. 2016. Patient-reported outcomes in the Swedish National Quality Registers. Journal of internal medicine, 279(2), s. 141–153.

Neittaanmäki, P., Lehto, M. 2018. Suomen kansalliset SOTE-tiedonlähteet ja tietojen hyödyntäminen. Infromaatioteknologian tiedekunnan julkaisuja. ISSN 2323-5004.

NetPromoter. 2023. What Is Net Promoter?. Viitattu 20.3.2023. Saatavilla: https://www.netpromoter.com/know/

Pitkänen, L., Haavisto, I., Vähäviita, P., Torkki, P., Leskelä, R., Komssi, V. 2018. Vaikutus SOTE:ssa. Suoritteista tuloksiin. Nordic Healthcare Group. Saatavilla: https://nhg.fi/wp-content/uploads/2018/11/Vaikuttavuus-sotessa-suoritteista-tuloksiin.pdf

Pitkänen, L., Torkki, P., Tolkki, H, Valtakari, M & Leskelä, R. 2020. Reittiopas vaikuttavuuteen – Vaikuttavuusperusteinen ohjaus sote- ja töllisyyspalveluissa. Valtioneuvoston selvitys- ja tutkimustoiminta. Valtioneuvoston kanslia, Helsinki.

Peura, P., Turunen, J., Purmonen, T., Happonen, P., Martikainen, J. 2011. Mitä lääkehoitojen kustannusvaikuttavuus tarkoittaa?. Fimea. Saatavilla: https://urn.fi/URN:NBN:fi-fe2014120350623

Redden, J. 2020. Predictive Analytics and Child Welfare: Toward Data Justice. Canadian Journal of Communication. Volume 45 Issue 1.

Porter, M., Teisberg, E. 2006. Redefining Health Care: Creating Value-Based Competition on Results. Boston: Harvard Business School Press.

Sagiroglu, S., Sinanc, D. 2013. Big data: A review, 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 2013, s. 42-47.

Salovaara, S., Leinonen, J., Silén, M. 2021. Barriers to the utilisation of data gathered through information systems in the management of social welfare organisations. Vol. 13 No. 4.

Sintonen, H., Roine, R., Ryynänen, O., Blom, M., Hytönen, M. 2022. Terveyshyödyn mittarit valintojen tueksi. Lääkärilehti. Viitattu 22.4.2023. Saatavilla: https://www.laakarilehti.fi/terveydenhuolto/terveyshyodyn-mittarit-valintojen-tueksi/

Sosiaali- ja terveysministeriö. 2022. Sosiaali- ja terveysministeriön julkaisuja 2022:18. Sosiaali- ja terveydenhuollon valtakunnalliset tavoitteet vuosille 2023–2026. ISSN pdf: 1797–9854

STM. 2023. Toisiolaki mahdollistaa sosiaali- ja terveystietojen tietoturvallisen käytön. Viitattu 20.3.2023. Saatavilla: https://stm.fi/sote-tiedon-hyodyntaminen

Terveyskirjasto. 2016. Lääketieteen sanasto. Viitattu 5.3.2023. Saatavilla: https://www.terveyskirjasto.fi/ltt01376

Torkki, P. 2019. Kustannusvaikuttavuus. Helsingin yliopisto. Viitattu 20.3.2023. Saatavilla: https://blogs.helsinki.fi/klaara-network/files/2019/03/Paulus-Torkki-2019_03_15.pdf

Valtonen, H. 2011. Ehdotukset sosiaali- ja terveydenhuollon rahoituksen kehittämiseksi. Valtioneuvoston kanslian julkaisusarja. 19/2011. ISSN 1799-7828

VM. 2022. Hyvinvointialueiden valmisteluun on myönnetty lisämäärärahaa vuodelle 2022. Valtiovarainministeriön tiedote. Viitattu 17.3.2023. Saatavilla: https://vm.fi/-/hyvinvointialueiden-valmisteluun-on-myonnetty-lisamaararahaa-vuodelle-2022

Vuorela, P. 2017. Virtsankarkailun taloudellinen arviointi. Käypä hoito. Viitattu 27.3.2023. Saatavilla: https://www.kaypahoito.fi/nix01799

Vuorela, T., 2018. Mikä parantaisi terveyttäsi? – Edistynyt analytiikka valottaa vaihtoehtoja. Sitra. Viitattu. 12.4.2023. Saatavilla: https://www.sitra.fi/blogit/mika-parantaisi-terveyttasi-edistynyt-analytiikka-valottaa-vaihtoehtoja/

Waller, M., Fawcett, S. 2013. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. J Bus Logist, 34. s. 77–84.