

Amanda Veid

**MASSADATA TERVEYDENHUOLLOSSA: MAHDOL-
LISUUDET JA HAASTEET**



JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2021

TIIVISTELMÄ

Veid, Amanda

Massadata terveydenhuollossa: mahdollisuudet ja haasteet

Jyväskylä: Jyväskylän yliopisto, 2021, 46 s.

Tietojärjestelmätiede, kandidaatintutkielma

Ohjaaja: Halttunen, Veikko

Digitalisoituneessa maailmassa valtavat datamäärät kasvavat jatkuvasti ja massadatan hyödyntäminen on suuri trendi eri yritysten sekä organisaatioiden toiminnassa. Massadatalle on tunnistettu olevan merkittävää potentiaalia myös terveydenhuollon toimialalla, sillä tulevaisuuden terveydenhuolto on potilaskeskeinen ja digitaalisia ratkaisuja hyödyntävä. Tämän kandidaatintutkielman tarkoituksena oli kartoittaa massadatan hyödyntämiseen liittyviä mahdollisuuksia sekä haasteita terveydenhuollossa. Tutkielma toteutettiin kirjallisuuskatsauksena. Tutkimuksessa todettiin, että massadataa hyödyntävillä ratkaisulla voidaan parantaa hoidon laatua, optimoida terveydenhuollon resursseja sekä saavuttaa kustannussäästöjä. Hoidon laatua parantaviksi ratkaisuksi tunnistettiin personoidun terveydenhuollon ratkaisut, kuten potilasprofiilit sekä hoitovasteen ennustaminen. Lisäksi massadataa hyödyntävällä omatoimisella ja ennaltaehkäisevällä terveydenhuollolla todettiin potentiaalia terveyden edistämisessä sekä sairauksien ehkäisemisessä. Massadataa hyödyntämällä voidaan saavuttaa kansantaloudellisia säästöjä suorien ja epäsuorien säästömekanismien kautta. Säästömekanismeihin kuuluvat muun muassa hoidon ja tuottavuuden tehostuminen sekä kansantautien vähäisempi esiintyvyys. Massadatalle voidaan myös optimoida terveydenhuollon ajanhallintaa sekä edistää terveystaloustiedettä tieteenalana. Jotta hyötyjä voidaan saavuttaa pitkällä aikavälillä, on keskeistä myös tiedostaa massadatan hyödyntämiseen liittyviä haasteita. Merkittäviksi terveydenhuollon massadatan haasteiksi todettiin tietoturvallisuuden ja yksityisyyden haasteet, datan laatuun liittyvät ongelmat sekä datan keräämiseen ja omistajuuteen liittyvät haasteet. Tietoturvallisuuteen liittyvistä haasteista merkittäviksi tunnistettiin erityisesti henkilötietojen väärinkäyttö, potilasturvallisuuden vaarantuminen, taloudelliset vahingot, kansalaisten luottamuksen heikentyminen sekä negatiiviset ulkoisvaikutukset. Yksityisyyden haasteista merkittävimmiksi tunnistettiin datan epäeettinen väärinkäyttö sekä siitä koituvat mahdolliset seuraamukselliset tai deontologiset haasteet. Lisäksi datan laatuun on tärkeää kiinnittää erityistä huomiota, jotta personoidun terveydenhuollon ratkaisut toimivat tarkoituksenmukaisesti.

Asiasanat: massadata, terveydenhuolto, personoitu terveydenhuolto, terveystaloustiede, kyberturvallisuus, yksityisyys

ABSTRACT

Veid, Amanda

Big Data in healthcare: opportunities and challenges

Jyväskylä: University of Jyväskylä, 2021, 46 pp.

Information Systems Science, Bachelor's Thesis

Supervisor: Halttunen, Veikko

In the digitalized world, large amounts of data are constantly growing. The utilization of Big Data is a rapidly evolving trend in the operations of various companies and organizations. Big Data has also been identified as having significant potential in the healthcare industry, as in the future healthcare will be digitalized. The purpose of this study was to examine the opportunities and challenges related to the utilization of Big Data in healthcare. This Bachelor's Thesis was accomplished as a literature review. The study found that Big Data can improve the quality of health care and optimize health care resources. Additionally, cost savings can be achieved by utilizing Big Data. Personalized healthcare solutions, such as patient profiles and prediction of treatment response, were solutions that improve the quality of health care. Furthermore, individualized, and preventive health care solutions have the potential for health promotion and disease prevention. By utilizing Big Data, cost savings can be achieved through direct and indirect saving mechanisms. Saving mechanisms include improved treatment and productivity, as well as lower morbidity of public diseases. Big Data can also be used to optimize healthcare time management and to improve health economics as a discipline. To achieve long-term benefits, it is also essential to be aware of the challenges associated with exploiting Big Data. Significant challenges were identified as information security and privacy challenges. Furthermore, there are data quality issues, data collection, and ownership challenges. Information security challenges cause criminal use of personal data, patient safety challenges, and financial damage. Additionally, it can cause reduced confidence in the health care system and negative externalities. The biggest challenges to information privacy are the unethical misuse of data and the possible consequences. Furthermore, it is important to pay special attention to data quality for personalized health care solutions to work appropriately.

Keywords: big data, health care, personalized health care, health economics, security, privacy

KUVIOT

KUVIO 1 Massadatan V-malli	11
KUVIO 2 Potilaan sairastumisriskianalyysi	16
KUVIO 3 Potilasprofiilin algoritmin toiminta.....	17

TAULUKOT

TAULUKKO 1 Massadataan liittyvät mahdollisuudet.....	37
TAULUKKO 2 Massadataan liittyvät haasteet	38

SISÄLLYS

TIIVISTELMÄ	2
ABSTRACT	3
KUVIOT	4
TAULUKOT	4
SISÄLLYS.....	5
1 JOHDANTO.....	7
2 MASSADATA.....	9
2.1 Massadatan määritelmä.....	9
2.2 Massadatan ominaisuudet	10
2.3 Massadatan hallinta.....	11
2.4 Terveydenhuollon massadata.....	12
3 MASSADATAN HYÖDYNTÄMINEN TERVEYDENHUOLLOSSA	14
3.1 Personoitu terveydenhuolto ja potilasprofiilit	14
3.2 Hoitovasteen ennustaminen	18
3.3 MyData terveydenhuollossa	20
4 KANSANTALOUDELLINEN NÄKÖKULMA	22
4.1 Säästömekanismit	22
4.2 Prosessien tehostuminen	23
4.3 Eriarvoisuuden vähentäminen	24
4.4 Massadata ja terveystaloustiede	26
5 MASSADATAN HYÖDYNTÄMISEEN LIITTYVÄT HAASTEET	27
5.1 Tietoturvaongelmat	27
5.1.1 Tietoturvaongelmien aiheuttamat haasteet.....	27
5.1.2 Tietoturvan parantaminen	29
5.2 Yksityisyys	30
5.3 Massadatan laatuongelmat	32
5.4 Datan keräämiseen ja omistajuuteen liittyvät haasteet.....	33
6 JOHTOPÄÄTÖKSET	36
7 YHTEENVETO	39

LÄHTEET	41
---------------	----

1 JOHDANTO

Yleisesti massadatalla (engl. Big Data) tarkoitetaan jatkuvasti kasvavia ja vaikeasti hallittavia suuria datamassoja. Suuret datamäärät kasvavat digitalisoituneessa maailmassa jatkuvasti. Erityisesti eri toimialojen digitalisaatio, sosiaalinen media sekä IoT-laitteet ovat vauhdittaneet datan kasvua huomattavasti. Massadatan sekä massadata-analytiikan on todettu olevan merkittävä menestystekijä muun muassa yritysten liiketoiminnassa sekä kasvuprosesseissa (Walker, 2015). Antikaisen ym:n (2016) mukaan massadataa voidaan hyödyntää myös terveydenhuollon toimialalla. Massadata on digitalisaation tärkeä sovel-lusalue ja sen hyödyntämiseltä odotetaan kansantaloudellisia säästöjä sekä pa-rempaa terveydenhuollon laatua. (Antikainen ym., 2016.)

Terveydenhuollon toimiala on digitaalisessa murroksessa. Sosiaali- ja terveystieteiden ministeriön (2016) raportin mukaan Suomessa ikärakenne muuttuu, huol-tosuhde heikkenee ja valtion velka kasvaa – terveydenhuollon palveluilta vaa-ditaan siis tulevaisuudessa enemmän tuottavuutta aikaisempaa kustannuste-hokkaammin. Digitalisaation on tunnistettu olevan suuressa roolissa näihin haasteisiin vastaamisessa. (Sosiaali- ja terveystieteiden ministeriö, 2016.) Krumholzin (2014) mukaan terveydenhuollon massadatan avulla voidaan jalostaa arvokasta in-formaatiota edistyneen massadata-analytiikan avulla. Massadatalla sekä sen analytiikalla voidaan vastata niin potilaiden kuin lääkäreidenkin informaatio-tarpeisiin. Lisääntyneestä informaatiosta on hyötyä myös tutkijoiden sekä ter-veydenhuollon päättäjien työssä. (Krumholz, 2014.)

Massadatassa on siis tunnistettu olevan merkittävää potentiaalia terveydenhuollon toimialalla, mutta ratkaistavana on silti vielä myös useita haasteita. Terveydenhuollon toimialan on tunnistettu olevan erittäin altis tietoturvaon-gelmille, kuten tietomurroille (Karim, Abderrahim & Hayat, 2018). Lisäksi haas-teita aiheuttavat yksityisyyden ja anonymisoinnin haasteet sekä eettiset ja sosi-aaliset ongelmat (Antikainen ym., 2016). Haasteeksi on tunnistettu myös datan omistajuuteen liittyvät kysymykset (Agrawal & Prabakaran, 2020). Vuonna 2020 tapahtunut Psykoterapiakeskus Vastaamon tietomurto on varoittava esi-merkki siitä, mitä voi tapahtua, jos tietoturvasta ei huolehdita asianmukaisesti. Suur-Uskin ja Rydenfeltin (2020) mukaan tietomurron vuoksi kymmenien tu-

hansien ihmisten arkaluontoisia tietoja päätyi väärille tahoille. Henkilökohtaisten haittojen lisäksi vastaavan kaltaiset tapaukset voivat heikentää ihmisten luottamusta terveydenhuollon palveluihin sekä vähentää hoitoon hakeutumista. (Suur-Uski & Rydenfelt, 2020.)

Tässä kandidaatintutkielmassa perehdytään massadataan terveyden edistämisen sekä kansantaloudellisten säästöjen näkökulmasta. Lisäksi tarkoituksena on käsitellä aiheeseen liittyviä haasteita ja riskejä, koska pitkällä aikavälillä hyötyjä ei voida saavuttaa ilman riskien huomioimista. Tutkielman tavoitteena on kartoittaa mahdollisuuksia sekä haasteita liittyen massadatan hyödyntämiseen terveydenhuollossa. Tarkoituksena on luoda kattava yleiskatsaus, joka tarkastelee aihetta useasta eri näkökulmasta. Mikäli massadataa onnistutaan hyödyntämään tulevaisuuden terveydenhuollossa turvallisella ja järkevällä tavalla, voidaan saavuttaa merkittäviä hyötyjä sekä kansantaloudellisia säästöjä. Tutkimuksen motiivina on erityisesti aiheen ajankohtaisuus sekä yhteiskunnallinen merkittävyys. Tutkielman tutkimusongelma muodostuu seuraavista tutkimuskysymyksistä:

- Mitä hyötyjä massadatalta voidaan saavuttaa terveydenhuollossa?
- Mitä haasteita massadatan hyödyntämiseen terveydenhuollossa liittyy?

Tutkielma toteutettiin kirjallisuuskatsauksena. Tieteellisten JUFO-luokiteltujen artikkeleiden sekä julkaisujen lisäksi tutkielmassa hyödynnettiin muun muassa Valtioneuvoston sekä Sosiaali- ja terveysministeriön julkaisuja. Tiedonhankintaprosessissa käytettiin pääasiassa seuraavia tietokantoja: Google Scholar, IEEE sekä JYKDOKin kansainvälisten e-aineistojen haku. Hakusanoina käytettiin muun muassa seuraavia sekä niiden yhdistelmiä: *big data, health care, big data analytics, personalized healthcare, patient profile, health economics, privacy, data quality ja security*.

Johdannon ja yhteenvedon lisäksi tutkielma koostuu viidestä eri luvusta. Luvussa kaksi määritellään massadatan käsitettä, piirteitä sekä ominaisuuksia. Lisäksi luvussa käsitellään massadatan hallintaa sekä esitellään lähteitä, joista terveydenhuollon massadata koostuu. Kolmannessa luvussa tarkastellaan massadatan hyödyntämistä terveydenhuollossa terveyden ja hyvinvoinnin edistämisen näkökulmista. Tässä luvussa esitellään hoito- ja diagnosointiprosessien kannalta merkittäviä ratkaisuja.

Terveydenhuollon toimintaan ja organisointiin liittyy olennaisesti myös kansantaloudellinen näkökulma, johon siirrytään neljännessä luvussa. Neljännessä luvussa käsitellään massadatan avulla saavutettavia kansantaloudellisia säästöjä, prosessien tehostumista, eriarvoisuuden vähenemistä sekä massadataa ja terveystaloustiedettä. Viidennessä luvussa siirrytään massadatan hyödyntämiseen liittyviin haasteisiin, kuten tietoturvaongelmiin sekä yksityisyyden haasteisiin. Lisäksi käsitellään massadatan laatuongelmia sekä datan keräämiseen ja omistajuuteen liittyviä eettisiä ongelmia. Johtopäätökset-lukuun on koottu tutkielman tulokset, jonka jälkeen lopussa on lyhyt yhteenvedo tutkielmasta.

2 MASSADATA

Tässä luvussa esitellään massadatalle tyypilliset piirteet ja ominaisuudet. Ensimmäisessä alaluvussa määritellään massadatan käsite, jonka jälkeen toisessa alaluvussa massadatan ominaisuudet määritellään tunnetun V-mallin avulla. Kolmannessa alaluvussa käsitellään massadatan hallintaa yleisellä tasolla. Lopuksi esitellään lähteitä, joista terveydenhuollossa hyödynnettävää massadataa kerätään.

2.1 Massadatan määritelmä

Massadatan (engl. Big Data) käsite ei ole täysin vakiintunut ja se määritellään eri lähteissä hieman eri tavoin. Jotkut määritelmät keskittyvät siihen, mitä massadata on ja vastaavasti toiset pyrkivät vastaamaan siihen, mitä sillä voidaan tehdä (Gandomi & Haider, 2015). Collinsin (2016) mukaan massadata on yhteisnimitys tietojärjestelmien lisääntyneelle kapasiteetille, jonka avulla voidaan kerätä ja tallentaa suuria määriä strukturoimatonta dataa sekä tehdä siitä tilastollista analyysia.

Senthilkumar, Bharatendara, Amruta, Angappa ja Chandrakumarmangalam (2018) määrittelevät tutkimuksessaan massadatan käsitettä. Sen mukaan massadatalla tarkoitetaan suuria, nopeasti kasvavia, vaikeasti hallinnoitavia sekä jatkuvasti muuttuvia datamassoja. Lisäksi on tunnistettu, että massadatan hallinta vaatii edistynyttä teknologiaa ja analytiikkaa. (Senthilkumar ym., 2018.) Massadataa ei kuitenkaan tulisi määritellä vain koon ja nopeuden perusteella, koska se on niin monimuotoista (Lee & Sohn, 2016). Raghupathi ja Raghupathi (2014) esittävät tutkimuksessaan, että massadataan kuuluu strukturoidun datan lisäksi myös strukturoimatonta sekä puolistrukturoitua dataa. Strukturoitu data on jäsenneiltyä tietokantamuotoista dataa, jota on helppo tallentaa, analysoida, hakea ja käsitellä. Terveystieteiden kontekstissa strukturoitua dataa ovat muun muassa mittareiden mittaama data sekä potilas- ja lääketieteelliset tiedot. (Raghupathi & Raghupathi, 2014.) Käytännössä se on siis dataa, joka on tallen-

nettu vaikkapa relaatiotietokantaan tai taulukkolaskentaohjelmaan (Gandomi & Haider, 2015).

Datatyypit voidaan nähdä jatkumona vahvasti strukturoidusta datasta strukturoimattomaan (Balducci & Marinova, 2018). Lee ja Sohn (2016) määrittelevät strukturoimattoman datan olevan dataa, jota ei voida tallentaa perinteiseen tietokantamuotoon ja siihen lukeutuvat muun muassa tekstit, kuvat, videot sekä sijaintitiedot. Strukturoimatonta dataa on erittäin vaikea hallita, sillä sen sisällöt ovat niin monimuotoisia ja erilaisia keskenään. Monimuotoisuuden vuoksi tarvitaankin uusia käsittelymenetelmiä, jotta datasta voidaan jalostaa arvokasta informaatiota. (Lee & Sohn, 2016.) 90 % terveydenhuollon massadatatista on strukturoimatonta (Senthilkumar ym., 2018) ja siksi strukturoimattoman datan hallintaan tulisi kiinnittää erityistä huomiota.

Puolistrukturoitu data on jotain edellä määriteltyjen datatyyppejen välistä. Gandomin ja Haiderin (2015) mukaan puolistrukturoitu data ei täytä strukturoidun, eikä strukturoimattoman datan määritelmää: sillä on osittain strukturoidun datan piirteitä, mutta ei kuitenkaan tiukkoja standardeja. XML-tiedostot ovat tyypillinen esimerkki puolistrukturoidusta datasta. (Gandomi & Haider, 2015.) Tiivistetysti massadatala tarkoitetaan jatkuvasti ja nopeasti kasvavia suurina datamassoja, joita on vaikea hallita perinteisillä data-analytiikan keinoilla. Massadataan kuuluu strukturoitua, strukturoimatonta sekä puolistrukturoitua dataa.

2.2 Massadatan ominaisuudet

Massadatan ominaisuuksia havainnollistetaan usein V-mallin avulla. Massadataa voidaan tyypillisesti määritellä vaikkapa kolmen tai neljän V:n avulla, mutta Senthilkumarin ym:n (2018) mukaan terveydenhuollon massadatan ominaisuudet määritellään tyypillisesti kuuden V:n mukaan. Akhtar (2018) sekä Senthilkumar ym. (2018) määrittelevät V-mallin muodostuvan seuraavista ominaisuuksista:

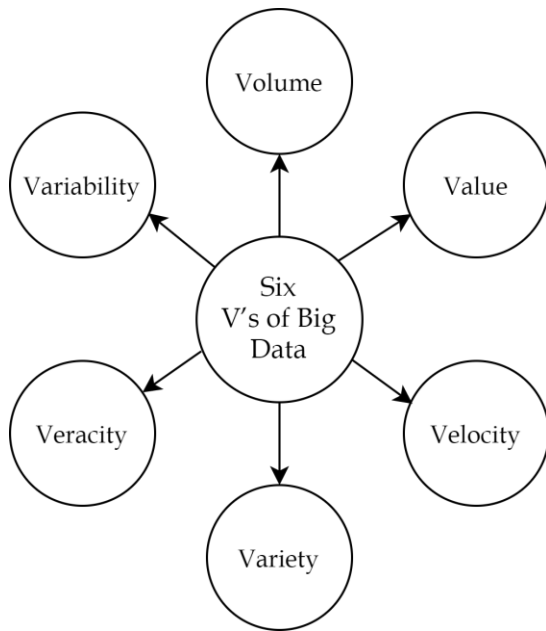
- **Määrä**, (engl. volume)
- **Arvo**, (engl. value)
- **Nopeus**, (engl. velocity)
- **Monimuotoisuus**, (engl. variety)
- **Tarkkuus ja todenmukaisuus**, (engl. veracity)
- **Vaihtelevuus**, (engl. variability)

Akhtarin (2018) määritelmän mukaan määrä-ominaisuudella viitataan valtaviin datamääriin, joita ei voida kerätä, hallita ja prosessoida perinteisillä data-analytiikan keinoilla. Lisäksi ominaisuus kuvaa sitä, että jo olemassa olevien valtaviin datamassojen lisäksi massadataa syntyy jatkuvasti suurina määriä lisää. (Akhtar, 2018.) Arvo-ominaisuudella kuvaillaan sitä, että massadatatista voidaan data-analytiikan avulla jalostaa arvokasta informaatiota (Akhtar, 2018). Käy-

tännössä informaatiota voidaan hyödyntää vaikkapa päätöksenteossa ja sillä voidaan saavuttaa merkittävää lisäarvoa (Senthilkumar ym., 2018).

Nopeus-ominaisuudella tarkoitetaan Akhtarin (2018) määritelmän mukaan sitä, että massadata kasvaa suurella frekvenssillä samanaikaisesti, kun jo olemassa olevaa dataa hallitaan. Lisäksi nopeus-ominaisuudella kuvaillaan datan käsittelyn nopeutta. (Akhtar, 2018.) Monimuotoisuus-ominaisuudella viitataan siihen, että massadataan kuuluu paljon ominaisuuksiltaan erimuotoista dataa, jota kerätään lukuisista eri lähteistä. Massadata koostuu strukturoidusta, strukturoimattomasta sekä puolistrukturoidusta datasta (Senthilkumar ym., 2018).

Akhtarin (2018) mukaan tarkkuuden ja totuudenmukaisuuden ominaisuudet ovat massadatassa alhaisia. Massadata ei ole täsmällistä ja aina totuudenmukaista dataa ja sen oikeellisuutta sekä tarkkuutta on vaikea vahvistaa. Massadatan hyödyntämisprosessissa nämä ominaisuudet saattavat aiheuttaa haasteita. (Akhtar, 2018.) Senthilkumarin ym:n (2018) määritelmän mukaan vaihtelevuus-ominaisuus viittaa datan vaihteluun sen käsittelyn ja elinkaaren aikana. Suuri vaihtelevuus lisää massadatan houkuttelevuutta sekä mahdollisuutta jalostaa arvokasta ja ajankohtaista informaatiota jatkuvasti. (Senthilkumar ym., 2018.) Alla olevaan kuvioon (kuvio 1) on koottu massadatan kuusi eri ominaisuutta V-mallin mukaisesti.



KUVIO 1 Massadatan V-malli (Senthilkumar ym., 2018)

2.3 Massadatan hallinta

Gandomin ja Haiderin (2015) mukaan yleisesti massadatan hallinta voidaan nähdä prosessina, johon kuuluu viisi vaihetta: (1) datan kerääminen ja tallen-

taminen, (2) datan puhdistaminen ja metadatan luominen, (3) integraatio, aggregaatio ja esittäminen, (4) mallintaminen ja analyysi sekä (5) datan tulkinta. Nämä viisi vaihetta muodostuvat kahdesta pääprosessista, jotka ovat tiedonhallintaa ja analytiikkaa. Tiedonhallintaprosessi sisältää vaiheet 1–3 ja siinä tavoitteena on hankkia, tallentaa ja valmistella dataa analytiikkavaihetta varten. Analytiikkavaiheessa dataa analysoidaan ja pyritään jalostamaan informaatioksi, jota voidaan myöhemmin tulkita ja hyödyntää käytännössä. (Gandomi & Haider, 2015.)

Raghupathi ja Raghupathi (2014) käsittelevät tutkimuksessaan massadatan hallintaa terveydenhuollon toimialalla. Sen mukaan massadatan hallinta nähdään samantyyppisenä prosessina kuin Gandomin ja Haiderin (2015) mallissa. Massadatan hallintaa havainnollistetaan nelivaiheisena prosessina, johon kuuluu (1) datan kerääminen, (2) transformointi, (3) prosessointi eri tekniikoiden avulla sekä (4) datan analysointi ja tulkinta. (Raghupathi & Raghupathi, 2014.)

Terveydenhuollon massadata sijaitsee lukuisissa sisäisissä ja ulkoisissa lähteissä sekä sovelluksissa. Raghupathi ja Raghupathi (2014) toteavat tutkimuksessaan, että massadatan hallintaprosessi alkaakin massadatan keräämisellä yhteen tietoaltaaseen. Seuraavassa vaiheessa on useita vaihtoehtoja: raakadata voidaan ottaa talteen, transformoida helpommin käsiteltävään muotoon ja siirtää eteenpäin massadatan hallintaan tarkoitettulle alustalle. Vaihtoehtoisesti ennen siirtovaihetta voidaan käyttää muita lähestymistapoja, kuten tietovarastointia. Tietovarastoinnissa eri lähteistä koottu data voidaan prosessoida, vaikka data ei olisikaan saatavilla reaaliajassa. (Raghupathi & Raghupathi, 2014.)

Raghupathin ja Raghupathin (2014) prosessimallin mukaan seuraavaksi tehdään päätöksiä tietojen syöttötavasta, hajautetusta suunnittelusta, työkalujen valinnasta sekä analytiikan eri vaihtoehtoista. Transformoitu data siirretään valituille alustoille ja työkaluille. Lopuksi hyödynnetään massadata-analytiikkaa. Tilastotieteen, tietojenkäsittelytieteen, sovelletun matematiikan sekä taloustieteen avulla on kehitetty lukuisia tekniikoita yhdistämään, käsittelemään, analysoimaan sekä visualisoimaan dataa. Terveydenhuollon massadatan hallinnassa tyypillisiä analytiikan vaihtoehtoja ovat kyselyt, raportit, OLAP-tekniikka sekä tiedonlouhinta. Kaikissa edellä mainituissa on tavoitteena visualisoida dataa sekä tuottaa informaatiota, jota voidaan käyttää hyödyksi vaikkapa päätöksenteossa. (Raghupathi & Raghupathi, 2014.)

2.4 Terveydenhuollon massadata

Raghupathin ja Raghupathin (2014) mukaan terveydenhuollon massadata kerättyä niin sisäisistä kuin ulkoisistakin lähteistä: terveydenhuollossa syntyvän datan lisäksi hyödynnetään siis myös terveydenhuoltojärjestelmän ulkopuolisia datalähteitä. Terveydenhuollon massadata koostuu sosiaalisen median datasta, laitteiden datasta, transaktiodatasta, biometrisistä tiedoista sekä ihmisten luomasta datasta. Sosiaalisen median dataan lukeutuu klikkaus- ja vuorovaikutus-

data eri sosiaalisen median kanavista sekä verkkosivuilta. (Raghupathi & Raghupathi, 2014.)

Raghupathin ja Raghupathin (2014) mukaan laitteiden datalla tarkoitetaan terveydenhuollossa käytettävien laitteiden dataa, eli vaikkapa sensoreiden, mittareiden ja elintoimintolaitteiden mittaamaa dataa. Transaktiodataan kuuluu terveydenhuoltoon liittyvät laskutustiedot ja biometrisiin tietoihin muun muassa geenitiedot, röntgen- ja muut lääketieteelliset kuvat sekä verenpaine- ja pulsuslukemat. Ihmisten luomalla datalla tarkoitetaan lääkäreiden muistiinpanoja, sähköisiä potilastietoja, sähköposteja sekä asiakirjoja. (Raghupathi & Raghupathi, 2014.)

Terveydenhuollossa hyödynnettävä massadata kertyy siis lukuisista eri lähteistä. Edellä mainittujen lisäksi sitä kerätään myös esimerkiksi potilastietojärjestelmistä sekä kliinisistä päätöksenteon tukijärjestelmistä (Senthilkumar ym., 2018). Andreu-Perez, Poon, Merrifield, Wong ja Yang (2015) esittävät tutkimuksessaan, että digitaaliset viestintävälineet ovat tuoneet uusia ulottuvuuksia yhteydenpitoon potilaiden ja lääkäreiden välillä. Lisäksi osa kroonista sairautta sairastavista potilaista jakaa myös kokemuksiaan vertaisten kanssa digitaalisilla alustoilla. Näiden viestintäkanavien on myös todettu olevan potentiaalisia lähteitä massadatan keräämiseen. (Andreu-Perez ym., 2015.)

Neittaanmäki, Lehto, Ruohonen, Kaasalainen ja Karla (2019) käsittelevät raportissaan Suomen terveystietoa ja sen hyödyntämistä. Raportin mukaan SOTE-tiedot Suomessa koostuvat asiakas- ja potilastiedoista, taloustiedoista sekä henkilö- ja resurssitiedoista. Asiakas- ja potilastiedot sisältävät dataa palvelujen käytöstä, terveydentilasta sekä elämäntilanteesta. Lisäksi Suomessa on käytössä kansallinen Kanta-järjestelmä, josta löytyy runsaasti dataa potilaista ja muun muassa eri lääkityspäätöksistä. Kanta-järjestelmää käytetään kaikissa sairaanhoitopiireissä sekä apteekeissa. Kanta-järjestelmän piirissä on lähes 200 perusterveydenhuollon yksikköä sekä arviolta 1000 julkisen terveydenhuollon palveluyksikköä. (Neittaanmäki ym., 2019.)

Terveydenhuollossa hyödynnettävä massadata koostuu siis potilastiedoista, sosioekonomisista tiedoista, muiden terveydenhuollon järjestelmien datasta, laboratoriotiedoista, biometrisistä tiedoista, lääketieteellisestä kuvantamisesta, laitteiden sensoreiden datasta sekä sosiaalisen median datasta. Näitä tietoja käytetään hyväksi suoraan hoitoprosessissa, mutta myös välillisesti esimerkiksi tietojohdamisessa ja terveydenhuollon kehittämisessä (Neittaanmäki ym., 2019). Seuraavassa luvussa käsitellään tarkemmin sitä, kuinka tätä eri lähteistä kertyvää massadataa voidaan hyödyntää terveydenhuollossa.

3 MASSADATAN HYÖDYNTÄMINEN TERVEYDENHUOLLOSSA

Tässä luvussa syvennyttään massadatan hyödyntämiseen terveydenhuollossa terveyden ja hyvinvoinnin edistämisen näkökulmasta. Antikainen ym. (2016) toteavat Suomessa olevan terveystietokantoja, jotka ovat maailmanlaajuisessa tarkastelussa poikkeuksellisen korkeatasoisia sekä kattavia. Terveystietokannoissa on muun muassa monipuolista dataa geenitietoihin liittyen ja lisäksi Suomessa on asiantuntijatasoinen osaamista bioinformatiikassa sekä molekyyli lääketieteessä. Tärkeimmiksi massadatan sovellusalueiksi terveyden edistämisen näkökulmasta tunnistetaan riskejä ennustavat terveydenhuoltopalvelut, itsehoiton palvelut, bioinformatiikka sekä molekyyli lääketiede. (Antikainen ym., 2016.) Seuraavissa alaluvuissa käsitellään personoitua terveydenhuoltoa, potilasprofiileja, hoitovasteen ennustamista sekä MyData-ratkaisua, eli omatoimista ja ennaltaehkäisevää terveydenhuoltoa. Myös bioinformatiikka ja molekyyli lääketiede on tunnistettu merkittäviksi terveydenhuoltoa edistäviksi osa-alueiksi, mutta aihepiirin laajuuden vuoksi edellä mainitut on rajattu tämän tutkielman ulkopuolelle.

3.1 Personoitu terveydenhuolto ja potilasprofiilit

Vähäkainun ja Neittaanmäen (2017) mukaan terveydenhuolto voidaan yleisesti jakaa reagoivaan ja ennakoivaan. Reagoivassa lähestymistavassa hoidetaan jo ilmenneitä oireita ja sairauksia, kun taas ennakoivassa lähestymistavassa puolestaan pyritään ennaltaehkäisemään sairauksien syntyminen. (Vähäkainu & Neittaanmäki, 2017.) Usein haasteeksi tunnistetaan potilaan kokonaisvaltaisen terveydentilan arviointi siten, että se sisältää sekä objektiivisiä että subjektiivisiä näkemyksiä (Hamad, Modrek, Kubo, Goldstein & Cullen, 2015). Chawlan ja Davisin (2013) mukaan terveydenhuolto kehittyy jatkuvasti kohti potilaskeskeisyyttä tautikeskeisyyden sijaan.

Chawla ja Davis (2013) toteavat tutkimuksessaan, että tautikeskeisessä terveydenhuollossa lääkäreiden päätöksenteko perustuu erilaisiin kokeisiin sekä kliiniseen asiantuntemukseen. Potilaskeskeisessä personoidussa terveydenhuollossa puolestaan potilaat osallistuvat omaan hoitoonsa ja palvelut ja hoitomuodot räätälöidään potilaiden tarpeiden ja mieltymysten mukaan. Terveydenhuolto kehittyy siis jatkuvasti ennakoivaan ja potilaskeskeisempään toteutustapaan ja siinä voidaan hyödyntää muun muassa potilasprofiileja, joiden avulla sairauksien ehkäiseminen ja hoitaminen on helpompaa. (Chawla & Davis, 2013.)

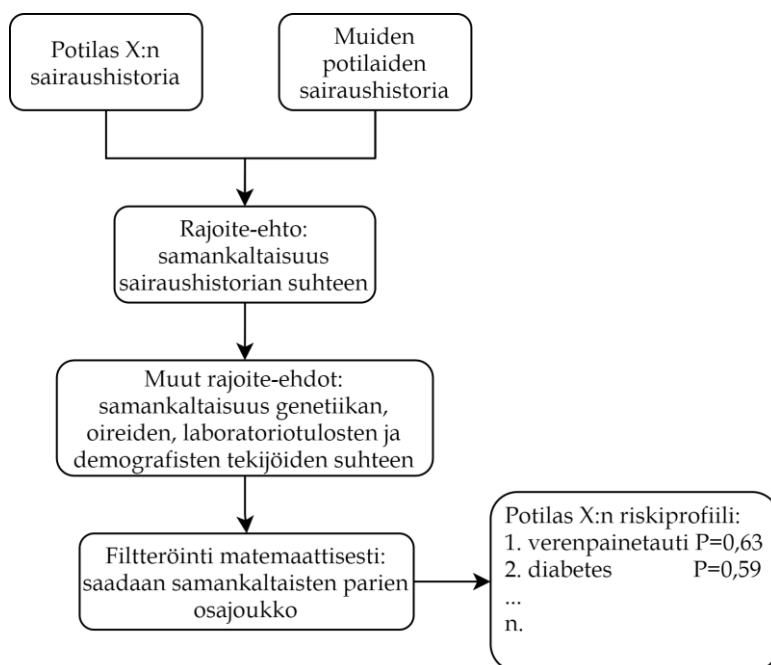
Chawla ja Davis (2013) esittävät tutkimuksessaan potilasprofiilin mallin, jossa otetaan huomioon potilaan riskiprofiili, hoitosuunnitelma sekä personoitu hyvinvointisuunnitelma. Mallissa otetaan erityisesti huomioon sairauksien ja ongelmien linkittyminen toisiinsa: sairauksien syntymiseen vaikuttaa geeniperimän lisäksi elintavat sekä ympäristötekijät. Potilailla, joilla edellä luetellut muuttujat ovat samankaltaisia, myös lopputulemat ovat usein samankaltaisia. (Chawla & Davis, 2013.) Tiivistettynä tähän ajatukseen perustuu personoitu terveydenhuolto: kun tiedetään potilaan genotyyppi, riskiprofiili, ympäristötekijät sekä sairaudet, voidaan ennakoida muiden samantyylisten profiilien perusteella vaikkapa sitä, mikä hoito toimisi kyseisellä potilaalla parhaiten.

Esitettyyn malliin liittyy kuitenkin myös haasteita. Chawla ja Davis (2013) esittelevät tutkimuksessaan tiedonlouhintamallin, jota hyödynnetään monissa personoiduissa palveluissa. Mallin perusteella ennustetaan käyttäjän mielipide tuotteesta tai palvelusta perustuen samanlaisten käyttäjäprofiilien mieltymyksiin. Oletetaan siis, että jos käyttäjien mieltymys on sama asioiden x ja y suhteen, on se sama myös asian z suhteen. Edellä kuvattua ideologiaa voitaisiin hyödyntää myös terveydenhuollon kontekstissa, mutta haasteeksi muodostuu kuitenkin se, että ei ole olemassa sairauksien ”luokitusjärjestelmää”, kuten vaikkapa suoratoistopalveluissa on elokuvien suhteen. Tiedossa on potilaan diagnoosit, mutta mikäli diagnoosia ei ole, se ei silti tarkoita, etteikö mahdollista sairautta olisi olemassa diagnosoimattomana. (Chawla & Davis, 2013.) Ongelmaa ei voida sivuuttaa ja siksi tulisikin ottaa huomioon, että potilasprofiilit voisivat toimia lääkärin apuna perinteisten menetelmien lisäksi. Mikäli potilaalla on oireita, tulisi tutkia mahdollisten diagnosoimattomien sairauksien olemassaolo ja päivittää potilaan diagnoosit ajan tasalle. Vasta sen jälkeen voidaan hyödyntää ennakoivaa massadata-analytiikkaa luotettavammin.

Potilasprofiileja voidaan mallintaa usealla eri tavalla ja useasta eri näkökulmasta. Tässä tutkielmassa esittelen kaksi eri viitekehystä, jotka eroavat hieman toisistaan. Chawla ja Davis (2013) esittävät tutkimuksessaan mallin, joka pyrkii ottamaan huomioon yleisesti demografiset tekijät, genetiikan sekä potilaan sairaushistorian. Malli toimii siten, että potilaan saapuessa vastaanotolle, hänen sairaushistoriaansa verrataan muiden potilaiden sairaushistoriaan. Massadatasta rajataan pois potilaat, jotka eivät ole profiililtaan riittävän samankaltaisia: samankaltaisuusrajoituksina toimivat sairaudet, oireet, genetiikka, laboriotulokset sekä demografiset tekijät. Matemaattisen filtoinnin jälkeen

saadaan tulokseksi luettelo sairauksista, jotka on järjestetty todennäköisyyden perusteella korkeimmasta pienimpään riskiin. (Chawla & Davis, 2013.)

Alla olevalla kuviolla (kuvio 2) on pyritty havainnollistamaan massadata-analytiikkaa hyödyntävää potilaan sairastumisriskianalyysia. Kuvio noudattaa pitkälti Chawlan ja Davisin (2013) mallia, mutta sitä on myös osittain muokattu. Mallissa asetetaan ensin rajoite-ehto sairaushistorioiden samankaltaisuudelle, jonka jälkeen asetetaan rajoite-ehtoja muiden muuttujien suhteen. Matemaattisen filteröinnin jälkeen saadaan samankaltaisten parien osajoukko, jonka perusteella syntyy potilaan riskiprofiili. Riskiprofiili esittää laskevassa järjestyksessä sairaudet, joihin potilaalla on riski sairastua. Lisäksi ilmoitetaan todennäköisyys sairastua. Todennäköisyyden ilmoittaminen riskiprofiilissa on merkittävää, koska muuten ei voida tietää, onko potilaan todennäköisyys sairastua esimerkiksi 0,05, 1 vai jotakin siltä väliltä.

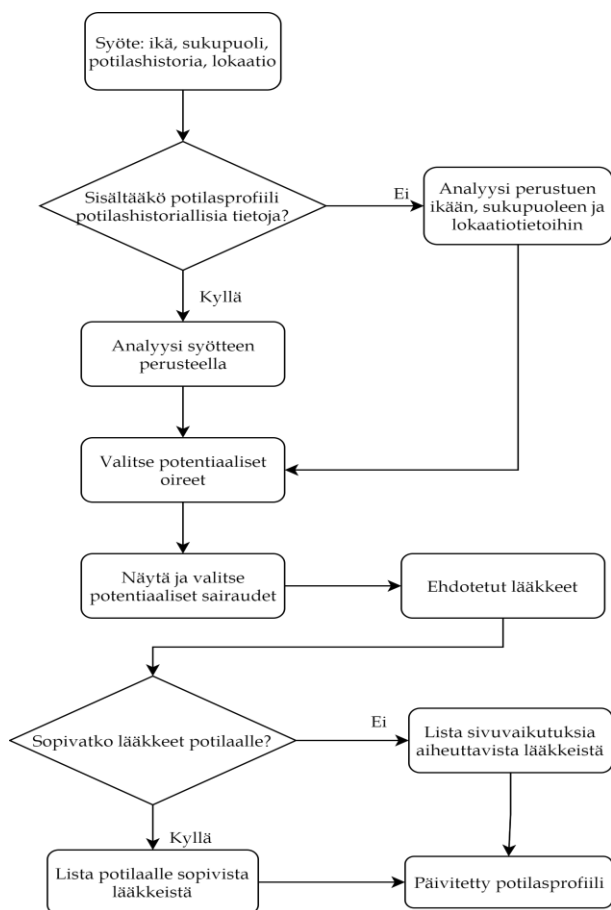


KUVIO 2 Potilaan sairastumisriskianalyysi (Chawla & Davis, 2013)

Myös Almobarak ja Jaziri (2019) esittävät tutkimuksessaan potilasprofiilin mallin, joka hyödyntää massadataa. Mallissa yhdistyvät personoidut ratkaisuehdotukset, tiedot ja palvelut kontekstiin liittyen sekä potilasprofiili. Potilas voi siis saada tietoa terveydestään usean eri lähestymistavan kautta. Potilasprofiili jakautuu kahteen osaan: potilaan ominaisuustietoihin ja potilashistoriaan. Potilaan ominaisuustietoihin kuuluvat muun muassa ikä, sukupuoli, kansalaisuus ja kaupunki. Lisäksi tietoihin kuuluvat sijaintitiedot, koska ympäristötekijöiden on todettu vaikuttavan terveyteen. Potilashistoriaan kuuluu terveystietojen historia, kuten aikaisemmat diagnoosit. (Almobarak & Jaziri, 2019.)

Almobarakin ja Jazirin (2019) esittelemän mallin (kuvio 3) toimintalogiikalla tavoitellaan samaa kuin aikaisemmin esitetyllä riskianalyysin mallilla: tarkempia diagnooseja sekä personoitua hoitoa. Lisäksi tavoitteena on kartoittaa

mahdollisia hoitomuotoja sekä lääkkeiden sivuvaikutuksia. Ensin järjestelmään syötetään tiedot, kuten potilashistoria, ikä, sukupuoli ja lokaatiotiedot. Järjestelmä analysoi potilaan tilannetta terveystietojen, demografisten tietojen sekä oireiden perusteella. Mikäli aikaisempia terveystietoja ei ole, analyysi perustuu vain muuhun dataan. Analyysin jälkeen syntyy luettelo mahdollisista sairauksista sekä potentiaalisista hoitomuodoista. Lisäksi järjestelmä tarkistaa, onko jokin ehdotettu lääke ristiriidassa potilashistorian kanssa, jotta mahdolliset sivuvaikutukset voidaan ottaa huomioon. (Almobarak & Jaziri, 2019.)



KUVIO 3 Potilasprofiilin algoritmin toiminta (Almobarak & Jaziri, 2019)

Edellä kuvatun mallin toimintalogiikassa on samankaltaisuuksia Chawlan ja Davisin (2013) esittelemän mallin kanssa, mutta tässä potilaan geenidataa ei analysoida. Erona on myös se, että Almobarakin ja Jazirin (2019) malli on pidemmälle jalostettu, sillä se ehdottaa myös potentiaalisia hoitomuotoja sekä analysoi lääkkeiden sopivuutta potilaalle.

Yleisesti potilasprofiilien tavoitteena on, että potilaalla olisi lääkärikäynnin jälkeen seuraavanlainen tilanne: aiempaa huomattavasti enemmän tietoa omasta terveydestä sekä personoituja elämäntapamuutosehdotuksia perustuen potilaan riskikartoitukseen. Lisäksi tavoitteena on, että potilaalla olisi myös mahdollisuus arvioida eri toimintamallien vaikutuksia omaan terveyteen. Lää-

käreiden näkökulmasta käytössä olisi apuväline, joka helpottaisi diagnoosien tekemistä sekä kokonaistilanteen arviointia muiden työvälineiden ohella.

Yllä esiteltiin kaksi erilaista lähestymistapaa potilasprofiileihin sekä massadatan avulla tapahtuvaan diagnosointiin. Toimintalogiikka useissa potilasprofiilien malleissa on samantapainen, mutta erot syntyvät erityisesti siitä, mitä massadatan osajoukkoa hyödynnetään. Jatkoa ajatellen edellä kuvatuista lähestymistavoista voisi jalostaa mallin, jossa yhdistyy edellä esiteltyjen mallien parhaat sekä toimivimmat ominaisuudet.

Riskikartoitus vaikuttaa toimivalta Chawlan ja Davisin (2013) mallissa, koska siinä otetaan huomioon laajalla skaalalla eri tekijöitä genetiikasta demografisiin tekijöihin. Vastaavasti Almobarakin ja Jazirin (2017) mallissa ei huomioida potilaan genetiikkaa, mikä on merkittävä puute, koska sen on todettu vaikuttavan olennaisesti sairauksien syntyyn. Chawlan ja Davisin (2013) mallista kuitenkin puuttuvat potentiaaliset hoitovaihtoehdot, joita Almobarakin ja Jazirin (2017) mallissa esitetään. Suurin ero näiden mallien välillä on siis se, mitä massadatan osajoukkoa hyödynnetään: voidaan painottaa joko geenidataa tai ottaa huomioon enemmän sosioekonomisia tekijöitä.

Parhaan lopputuloksen saavuttamiseksi olisi järkevää kartoittaa tarkemmin relevantteja sairauksien syntymiseen vaikuttavia muuttujia. Esimerkiksi Almobarakin ja Jazirin (2017) mallissa esitetyt lokaatiotiedot eivät välttämättä ole relevantteja tässä kontekstissa, vaan järkevämpää olisi analysoida geenitie-toja, kuten Chawlan ja Davisin (2013) mallissa. Myös Andreu-Perez ym. (2015) painottavat sairausmekanismien tutkimisessa erityisesti geenidatan tärkeyttä. Jokaisen potilaan geenidata tulisi yhdistää kliinisistä havainnoista, fysiologisista signaaleista sekä lääketieteellisistä kuvista saatuun fenotyyppiin. (Andreu-Perez ym., 2015.)

Mikäli edellä esitellyistä malleista yhdistettäisiin parhaat ominaisuudet, hyödynnettäisiin sitä massadatan osajoukkoa, jonka on eniten todettu vaikuttavan sairauksien syntyyn. Lisäksi diagnoosien perusteella voitaisiin ehdottaa potentiaalisia hoitovaihtoehtoja sekä niiden sopivuutta potilaalle. Hoitovastetta voidaan ennustaa tarkemmin massadatan avulla ja sitä käsitellään tarkemmin seuraavassa alaluvussa.

3.2 Hoitovasteen ennustaminen

Hoitovasteen ennustaminen on osa personoitua terveydenhuoltoa. Varadhanen, Segalen, Boyden, Wun ja Weissin (2013) mukaan potilaiden vaste hoitoon vaihtelee. Sama hoito tuottaa yhdelle huomattavaa hyötyä, toinen ei huomaa juuri mitään vaikutuksia ja kolmas saa haitallisia sivuvaikutuksia. Hoitovasteen heterogeenisyys voidaan käsittää ei-satunnaisena vaihteluna yksittäisten hoitovasteiden suunnassa ja laajuudessa. Mikäli hoitovasteen heterogeenisyyttä ymmärretään paremmin, voidaan hoidon vaikutusta ennustaa tapauskohtaisesti. (Varadhan ym., 2013.) Myös Collinsin (2016) mukaan massadatan avulla lääkkeiden vaikutuksia yksilöihin voidaan ennustaa paremmin. Käytännössä lääkäreil-

lä olisi siis enemmän tietoa, jonka pohjalta voitaisiin tehdä järkevämpiä hoitoon liittyviä päätöksiä. (Collins, 2016.)

Andreu-Perezin ym:n (2015) mukaan esimerkiksi syöpäsairauksien hoidossa ongelmana on ollut pitkään se, että useilla lääkkeillä on todettu olevan heterogeeninen hoitovaste jopa samantyyppiselle syövälle. Lisäksi jotkin lääkkeet toimivat vain hyvin pienellä määrällä potilaita ja erilaisen hoitovasteen on todettu johtuvan hyvin todennäköisesti eroista potilaiden geeneissä. (Andreu-Perez ym., 2015.) Heterogeeninen hoitovaste vaikeuttaa potilaiden hoitoa ja resursseja kuluu siihen, kun kokeilemalla yritetään etsiä sopivaa ratkaisua.

Bzdokin ja Meyer-Lindenbergin (2018) mukaan yleisesti perinteinen tutkimustavoite on ollut uudenlaisten hoitovaihtoehtojen löytäminen jonkin kliinisen ryhmän enemmistölle. Hoitovasteen ennustamisessa tavoitteena on puolestaan löytää optimaalinen hoitovaihtoehto yksittäiselle potilaalle. (Bzdok & Meyer-Lindenberg, 2018.) Erityisesti psykiatrian toimialalla yleisesti haasteeksi tunnustetaan epävarmuus siitä, miten jokin hoitomuoto x toimii henkilöllä y . Tenenbaumin ym:n (2019) mukaan mielenterveyden häiriöille on tyypillistä, että kahdella henkilöllä voi ilmetä sama diagnoosi, mutta erilaiset oireet. Tavoitteena on kehittää uusia luokitustapoja perustuen neurobiologisiin ja behavioristisiin dimensioihin: luokitustavassa otetaan huomioon geenit, mutaatiot, molekyylit, solut ja fysiologiset sekä behavioristiset tekijät. (Tenenbaum ym., 2019.)

Chenin ym:n (2018) mukaan sairauksien, diagnoosien sekä hoitovasteiden välisiä yhteyksiä voidaan analysoida massadata-analytiikalla. Kun tietämys psyykeen ja biologisten tekijöiden yhteyksistä kasvaa, voidaan saavuttaa parempia diagnooseja, interventioita sekä hoitotuloksia (Tenenbaum ym., 2019). Käytännössä tällaisessa lähestymistavassa hyödynnetään bioinformatiikkaa sekä koneoppimisen menetelmiä yhdistettynä massadataan. Andreu-Perezin ym:n (2015) mukaan bioinformatiikka on kasvava tieteenala, joka yhdistelee molekyylibiologiaa, biostatistiikkaa sekä tilastollisia geenitietoja kliinisen informatiikan kanssa. Bioinformatiikkaa voidaan hyödyntää muun muassa hoitovasteen ennustamisessa kompleksisillekin sairauksille. (Andreu-Perez ym., 2015.)

Tavanomaiset tilastolliset menetelmät eivät ole onnistuneet selittämään hoitovasteiden heterogeenisyyttä, mutta massadataa hyödyntävällä koneoppimisella sekä syväoppimisella on todettu olevan potentiaalia käsitellä tätä haastetta (Koppe, Meyer-Lindenberg & Durstewitz, 2021). Bzdokin ja Meyer-Lindenbergin (2018) mukaan koneoppiminen tarjoaa joukon työkaluja, jotka sopivat yksilötason kliinisten hoitovasteiden ennustamiseen. Edellä mainittujen lisäksi koneoppimisen menetelmistä hyödynnetään esimerkiksi ohjattua oppimista, ei-ohjattua oppimista, klusterointia sekä neuroverkkoja. (Bzdok & Meyer-Lindenberg, 2018.)

Gillanin ja Whelanin (2017) mukaan usein potilaat joutuvat kokeilemaan useita eri hoitomuotoja ennen toimivan hoidon löytymistä. Useat potilaat joutuvat siis kärsimään pitkiä jaksoja oireista, koska hoito ei ole riittävän kehittyntä. Tavoitteena onkin löytää toimiva hoito heti hoitoprosessin alussa. (Gillan

& Whelan, 2017.) Käytännössä siis voitaisiin arvioida, miten jokin lääke tai muu hoitomuoto todennäköisesti toimii tietyllä genotyypillä ja näin ollen välttyä toimimattomalta hoidolta sekä mahdollisilta haitallisilta sivuvaikutuksilta. Lisäksi terveydenhuolto toimisi tehokkaammin, kun resursseja ei kuluisi tehottomien hoitovaihtoehtojen kokeiluun, vaan hoito kohdentuisi heti alussa tehokkaasti. Massadatan hyödyntämisellä hoitovasteen ennustamisessa voitaisiin siis saavuttaa parempaa hoitoa, resurssien tehokasta kohdentumista sekä säästöjä kustannuksissa.

3.3 MyData terveydenhuollossa

Poikolan, Kuikkaniemen, Kuittisen, Hongon ja Knuutilan (2018) mukaan MyDatalla tarkoitetaan mallia ja ilmiötä, jossa henkilötietojen hallinta muuttuu organisaatiokeskeisestä ihmiskeskeisemmäksi. Lisäksi se tarkoittaa, että henkilötiedoista sekä henkilökohtaisesta datasta tulee sellainen resurssi, jota ihmiset voivat hyödyntää elämässään. Tulevaisuudessa ihmisillä olisi siis mahdollisuus hallita, hyödyntää sekä halutessaan siirtää dataa kolmansille osapuolille. Toisaalta MyDatalla ei kuitenkaan tarkoiteta kaikkia henkilötietoja, vaan se on henkilötietojen osajoukko. Käytännössä siis kaikki MyData kuuluu henkilötietoihin, mutta kaikki henkilötiedot eivät ole osa MyDataa. MyDatalla ei tarkoiteta mitään yksittäistä teknologiaa, vaan se on viitekehys ja infrastruktuuri, jossa on useita eri kerroksia. (Poikola ym., 2018.)

Poikola ym. (2018) esittävät raportissaan, että terveyteen liittyvä data on suuri osa MyData-kokonaisuutta. Tulevaisuudessa MyData helpottaa ihmisten vuorovaikutusta terveydenhuollon palveluntarjoajien kanssa. Lisäksi MyData on tunnistettu tekijäksi, jonka avulla voidaan kehittää ennakoivaa terveydenhuoltoa sekä lisätä ihmisten tietoisuutta hyvinvoinnistaan. (Poikola ym., 2018.) Laajan terveyteen liittyvän informaation ennustetaan muuttavan kansalaisten käsitystä terveydestään ja itsestään (Antikainen ym., 2016). Myös Niinimäki, Pöyhönen, Äyrämö ja Neittaanmäki (2017) toteavat tutkimuksessaan, että yhä useampi kansalainen tunnistaa datan arvon ja yksittäisen henkilön merkitys terveystiedon hyödyntäjänä, tuottajana ja hallitsijana mullistuu lähitulevaisuudessa.

Antikaisen ym:n (2016) mukaan MyDataa hyödyntävässä terveydenhuollossa ajatuksena on jalostaa genomitiedoista sekä elintavoista koostuvasta massadatasta kansalaistasolla hyödynnettävää informaatiota. Elintapadataan kuuluu muun muassa ravitsemus- ja liikuntadata ja sen keräämisessä hyödynnetään erilaisia itsemittausteknologioita. (Antikainen ym., 2016.) Kansallisessa genomistrategiassa on linjattu tavoitteeksi, että tulevaisuudessa kansalaiset voisivat hyödyntää itseään koskevaa genomitietoa elämässään (Sosiaali- ja terveysministeriö, 2015). Haukkalan ym:n (2018) mukaan kansalaiset ovat kiinnostuneita perimään liittyvästä terveystiedosta ja uskovat, että genomitiedoista voidaan saada hyötyä terveyden edistämässä. Haasteeksi saattaa kuitenkin muodostua riittävän tietämyksen puute liittyen genetiikkaan. (Haukkala ym.,

2018.) Ratkaisussa tulisi siis erityisesti ottaa huomioon se, että informaatio esitetään riittävän kansantajuudessa muodossa.

MyData-ratkaisussa kansalaisilla on mahdollisuus kerätä itsestään terveys- sekä sosiaalitietoja. Sen lisäksi, että kansalaiset hyötyvät datasta itse, se voidaan yhdistää osaksi kansallisia KANTA- sekä KANSA-tietovarantoja (Niinimäki ym., 2017). Siten MyData-ratkaisu yhdistyisi myös osaksi aikaisemmin esiteltyjä personoituja potilasprofiileja, sillä MyData olisi yksi datan lähde, jota niissä hyödynnetään.

THL:n (2017) raportissa esitetään Omatietovaranto eli PHR (Personal Health Record), joka on Kelan kehittämä ja ylläpitämä kokonaisuus. Omatietovaranto mahdollistaa tietosisältöjen hyödyntämisen sekä jakamisen. Sovellukset ja omatietovaranto muodostavat yhdessä kansallisen PHR-kokonaisuuden. Omatietovarannon on tunnistettu edistävän kansalaisten hyvinvointia, tukevan sosiaali- ja terveydenhuollon toimintaprosesseja sekä vuorovaikutusta eri toimijoiden kesken. Toimijoita liittyen Omatietovarannon kokonaisuuteen ovat siis kansalaiset, sosiaali- ja terveysalan ammattilaiset sekä organisaatiot. (THL, 2017.)

Niinimäen ym:n (2017) mukaan Omatietovaranto eli PHR on terveys- ja hyvinvointitiedoille tarkoitettu tietokanta, jonka käyttäjäryhmänä ovat potilaat. Ideana on, että dataa voi syöttää käyttäjä itse tai data voidaan siirtää suoraan terveys- ja hyvinvointisovelluksista. Omakannan Omatietovarantoon soveltuvaan dataan kuuluvat muun muassa erilaiset fysiologiset mittaukset, kyselyt sekä riski- ja itsearviointitestit. Lisäksi tiedoissa on aina merkintä, millä laitteella ja miten data on kerätty. (Niinimäki ym., 2017.)

THL:n (2017) raportin mukaan Omatietovaranto mahdollistaa sen, että yksilön rooli hoitoprosesseissa sekä terveyden ja hyvinvoinnin edistämässä vahvistuu. Lisäksi Omatietovaranto voi helpottaa palveluihin ohjautumista ja yksilön kokonaisvaltainen kuva henkilökohtaisesta terveydestä laajenee. Omatietovaranto mahdollistaa myös sosiaali- ja terveydenhuollon ammattilaisille laajemman näkymän asiakkaan tilanteesta. Ajatuksena on, että sosiaali- ja terveydenhuollon ammattilaiset käyttävät tietojen hyödyntämiseen erillisiä sovelluksia, jotka on kehitetty eri palveluiden tarpeisiin. MyDataa voidaan hyödyntää myös tutkimus- ja kehitystyössä, mikäli siihen on saatu suostumus kansalaiselta. (THL, 2017.) MyData-ratkaisulla on siis vahva potentiaali vaikuttaa kansanterveyteen usealta eri tasolta. Yksityishenkilöt voivat hyödyntää terveysdataa omassa toiminnassaan ja tällä on mahdollisuus toimia osana ennaltaehkäisevää terveydenhuoltoa ja edistää siten kansanterveyttä. Lisäksi ratkaisu monipuolistaa mahdollisesti hoitohenkilökunnan käytössä olevaa tietoa sekä edistää terveyteen liittyvää tutkimus- ja kehitystyötä.

Massadatan avulla voidaan tehostaa hoito- ja diagnosointiprosesseja, mutta terveydenhuoltoon sekä sen toimintaan liittyy olennaisesti myös kansantaloudellinen näkökulma. On tärkeää, että terveydenhuolto toimii kustannustehokkaasti ja että resurssit kohdentuvat järkevästi. Seuraavassa luvussa perehdytään kansantaloudellisesta näkökulmasta hyötyihin, joita massadatan avulla voidaan saavuttaa.

4 KANSANTALOUDELLINEN NÄKÖKULMA

Hamoudin ja Sachsin (1999) mukaan väestön keskimääräisellä terveydentilalla on todettu olevan vahva yhteys kansantalouteen useamman mekanismin kautta. Terveyden ja taloudellisten tekijöiden välillä on vahva korrelaatio, joten terveys on myös taloudellisen menestyksen tekijä. Terveys vaikuttaa muun muassa tuloihin, työn tuottavuuteen, työvoiman määrään sekä säästö- ja sijoitusasteisiin. (Hamoudi & Sachs, 1999.) Edellisessä luvussa käsiteltiin terveyden edistämistä massadatan avulla ja se liittyy suoraan myös kansantaloudelliseen näkökulmaan: mitä terveempää kansa on, sitä enemmän on työkykyisiä, hyvinvoivia sekä kansantaloudellisesti tuottavia ihmisiä. Vastaavasti mitä huonommin yleisesti voidaan, sitä suuremmat ovat kansantaloudelliset kustannukset. Tässä luvussa käsitellään kansantaloudellisia säästöjä sekä kustannustehokkuutta, joita voidaan saavuttaa terveydenhuollon toimialalla massadatan avulla. Lisäksi käsitellään prosessien tehostamista sekä massadatan mahdollisuuksia ja haasteita terveystaloustieteessä.

4.1 Säästömekanismit

Massadatan hyödyntämisellä terveydenhuollossa voidaan saavuttaa kansantaloudellisia säästöjä, jotka kertyvät eri mekanismien kautta. Useiden eri asiantuntija-arvioiden sekä lähteiden perusteella on todettu, että informaatioteknologiaa hyödyntävillä ratkaisuilla on potentiaalia vähentää sosiaali- ja terveydenhuollon kustannuksia Suomessa 2,5–5,5 miljardilla eurolla vuosina 2019–2028 (Kaasalainen & Neittaanmäki, 2018). Neittaanmäen ym:n (2019) mukaan säästömekanismit voidaan nähdä suorina tai epäsuorina. Suoriin säästömekanismeihin kuuluvat esimerkiksi tuottavuuden sekä potilaan hoidon tehostuminen. Vastaavasti epäsuoriin säästömekanismeihin lukeutuvat muun muassa kansantautien vähäisempi esiintyvyys sekä sosiaali- ja terveyskulujen väheneminen. (Neittaanmäki ym., 2019.)

Neittaanmäen ym:n (2019) mukaan suoria säästöjä syntyy, kun tuottavuus ja prosessien tehokkuus teknologiaa hyödyntäen kasvavat. (Prosessien tehostumista tarkastellaan tarkemmin alaluvussa 4.2.) Lisäksi säästöjä syntyy, kun hoito- ja diagnosointiprosessit tehostuvat. Edistynyttä teknologiaa käyttämällä voidaan saavuttaa myös hoidon laadun sekä vaikuttavuuden parantumista. Kun hoitoprosessi toimii mahdollisimman tehokkaasti, se edistää myös esimerkiksi yksilön paluuta työelämään. (Neittaanmäki ym., 2019.) Edellisissä luvuissa esiteltiin personoidun terveydenhuollon ratkaisuja: potilasprofiileja ja hoitovasteen ennustamista. Potilasprofiilit sekä mahdollisuus hoitovasteen ennustamiseen tehostavat hoito- ja diagnosointiprosesseja ja terveyden edistämisen lisäksi toimivat myös osana säästömekanismia.

Neittaanmäen ym:n (2019) mukaan suoriin säästömekanismeihin kuuluvat myös työaikasäästöt. Työaikaa voidaan käyttää tehokkaammin, koska tekoälytuettujen ratkaisujen on todettu tehostavan järjestelmien käyttöä 10–20 %. Lisäksi massadataa hyödyntävät teknologiset ratkaisut mahdollistavat tilannekuvan hahmottamisen kansallisella, alueellisella sekä yksilötasolla. Tarkempi tilannekuva puolestaan helpottaa johtamista ja resurssien kohdentamista. (Neittaanmäki ym., 2019.)

Neittaanmäen ym:n (2019) mukaan epäsuoriin säästömekanismeihin luokituvat kansantautien vähäisempi esiintyvyys, työkyvyttömyyden sekä sosiaali- ja terveystalouden lasku ja psyykkisen hyvinvoinnin lisääntyminen. Suurimpia kansantaloudellisia kustannuksia aiheuttavat kansantaudit ovat mielen-terveyden haasteet, tuki- ja liikuntaelinsairaudet sekä sydän- ja verisuonitaudit (Neittaanmäki ym., 2019). Findikaattorin (2020) sekä Eläketurvakeskuksen (2020) mukaan Suomessa yleisin syy jäädä työkyvyttömyyseläkkeelle on mielen-terveyden haasteet, kuten masennus. Mielen-terveyden haasteet aiheuttavat valtavasti yksilöiden kärsimystä sekä kansantaloudellisia kustannuksia. Mahdollisimman varhaisen vaiheen interventioilla sekä diagnosoinnilla on siis laajasti vaikutuksia esimerkiksi työkyvyttömyyden ehkäisyssä sekä syrjäytymisriskin tunnistamisessa (Neittaanmäki ym., 2019).

Käytännössä epäsuorat säästöt ovat saavutettavissa ennakoivien ja oikea-aikaisten terveys- ja sosiaalipalveluiden kautta, koska silloin hyvinvointi lisääntyy ja sairastavuus vähenee (Neittaanmäki ym., 2019). Edellisessä luvussa esitellyt MyData-ratkaisu sekä potilasprofiilit ovat tähän soveltuvia ennaltaehkäisevän terveydenhuollon ratkaisuja.

4.2 Prosessien tehostuminen

Prosessien sekä hoidon tehostuminen tarkoittaa hoidon vaikuttavuuden ja laadun parantumisen lisäksi ajankäytön optimointia. Kun potilas saa oikeanlaista hoitoa mahdollisimman pian sekä minimoitujen välikäsien kautta, terveydenhuolto toimii kustannustehokkaasti. Mitä varhaisemmassa vaiheessa potilas saa oikeanlaista hoitoa, sitä parempi on myös hoitoennuste.

Erholan, Jormanaisen, Kovasinin, Rissasen ja Keskimäen (2020) mukaan Suomessa terveydenhuollon palveluita joutuu odottamaan kauemmin verrattuna muihin OECD-maihin. Muihin pohjoismaihin verrattuna ongelma on vielä suurempi. Mikäli tyytymättömyyttä hoitoon ilmenee, se johtuu tyypillisesti liian pitkästä odotusajasta. Kansallisella tasolla terveydenhuollon henkilöstövoimavarat jakautuvat epätasaisesti, mikä osaltaan eriarvoistaa ihmisiä. Ongelma on se, että niillä alueilla, joihin on kasaantunut terveystalveluiden tarvetta, palveluita on vähiten saatavilla. (Erhola ym., 2020.) Tähän haasteeseen vastaisemissa täytyy ottaa huomioon ensisijaisesti hoitohenkilökunnan riittävyys. Lisäksi teknologisten ratkaisujen avulla voidaan pyrkiä tehokkaampaan resurssien kohdentamiseen sekä tehokkaampaan ajankäyttöön.

Chen, Li, Tang, Bilal ja Li (2016) esittävät tutkimuksessaan, että yksi tehokkaasti toimivan terveydenhuollon haasteista on potilaiden hoitotoimenpiteisiin kuluvan ajan hallitsemattomuus. Hoitotoimenpiteisiin kuluu eri määrä aikaa riippuen potilaasta, mikä tekee aikaennusteiden ja -suositusten tekemisestä monimutkaista. Vaihteluun vaikuttavat esimerkiksi olosuhteet, eri ajanjaksot sekä potilaan kunto. Tyypillisesti esimerkiksi usein iäkkäämmällä potilaalla johonkin hoitotoimenpiteeseen kuluu enemmän aikaa kuin nuoremmalla potilaalla. Kun eri tekijät otetaan huomioon, ajan ennakoinnin avulla jonotusajat lyhenevät ja terveydenhuollon resursseja on mahdollisuus kohdentaa järkevämmiin. (Chen ym., 2016.)

Chen ym. (2016) esittävät tutkimuksessaan mallin potilaan hoitoajan ennustamiseen. Hoitoajan ennustaminen perustuu potilaan eri ominaisuuksien sekä muiden aikaan vaikuttavien tekijöiden laskentaan. Datamassasta poistetaan tarpeeton data, kuten potilaan nimi, osoite ja puhelinnumero. Mallissa hyödynnetään terveydenhuollon massadataa ja merkittäviä parametrejä ovat hoidon alkamis- ja lopetusaika, potilaan ikä sekä hoitoon liittyvä informaatio. Hoitotoimenpiteeseen kuluva aika ennustetaan matemaattisten menetelmien avulla. Mallin ideana on myös suositella tehokasta hoitosuunnitelmaa, jossa jonotusaika on minimoitu. Mallissa ennustetaan koko toimenpiteeseen kuluva aika, kuten vaikkapa 30 minuuttia. (Chen ym., 2016.) Vastaavanlaisella menetelmällä voitaisiin saavuttaa joitain hyötyjä ajankäytön optimoinnissa. Haasteena on kuitenkin myös edellä mainitut resurssien epätasainen kohdentuminen sekä eriarvoisuus.

4.3 Eriarvoisuuden vähentäminen

Sosioekonomisella asemalla on todettu olevan vahva yhteys terveyteen ja hyvinvointiin (Voigt, 2017). Erholan ym:n (2020) mukaan Suomessa tuloerot ovat melko pienet, köyhyys on vähäistä ja sosiaaliturvajärjestelmä on toimiva. Lisäksi elinajanodote on pitkä ja lapsikuolleisuus on vähäistä. Asiat ovat siis melko hyvin, mutta terveys- ja hyvinvointierot perustuen sosioekonomiseen asemaan ovat kuitenkin selkeästi havaittavissa. Yleisesti yhteiskunnassa heikoimmassa asemassa olevilla myös terveydentila on heikompi verrattuna paremmassa so-

sioekonomisessa asemassa oleviin henkilöihin. (Erhola ym., 2020.) Sosiaaliset sekä taloudelliset tekijät vaikuttavat terveyteen ja terveydenhuollon tulisi osallttaa pystyä vastaamaan tähän haasteeseen ja pyrkiä vähentämään eriarvoisuutta. Eriarvoisuutta vähentämällä voitaisiin myös pitkällä aikavälillä vähentää sosiaali- ja terveydenhuollon kustannuksia. Potilaskeskeisen ja personoidun terveydenhuollon ratkaisut ovat myös painottaneet sosiaalisten tekijöiden vaikutusta terveyteen ja terveydenhuoltojärjestelmään (Garg, Jack & Zuckerman, 2013). Massadatalta on tunnistettu olevan potentiaalia tähän haasteeseen vastaamisessa (Murdoch & Detsky, 2013).

Voigt (2017) esittää tutkimuksessaan lähestymistavan eriarvoisuuden vähentämiseen. Ajatuksena on, että massadatan avulla saadaan informaatiota potilaan ympäristöstä sekä sosiaalisista terveyteen vaikuttavista tekijöistä. Toisaalta kuitenkin täytyy ottaa huomioon, että massadataa hyödyntävät ratkaisut saattavat lisätä eriarvoisuutta entisestään, koska heikossa asemassa olevat saattavat olla aliedustettuja datassa. (Voigt, 2017.)

Edellä todettiin, että sosiaaliset tekijät vaikuttavat merkittävästi terveyteen. Bazemoren ym:n (2016) mukaan lääkäreiltä puuttuu kuitenkin pääsy tämänkaltaiseen dataan, vaikka siitä voisi olla hyötyä potilaiden kohtaamisessa sekä hoidossa. Mikäli dataa olisi saatavilla myös esimerkiksi asuinpaikasta, ympäristöstä sekä sosiaalisesta pääomasta, voitaisiin terveydentilaa sekä esimerkiksi sairastumisriskejä arvioida paremmin. (Bazemore ym., 2016.) Perinteisesti hoitohenkilökunta saa informaatiota ainoastaan vuorovaikutuksessa potilaan kanssa tai mikäli vaikkapa lääkäriellä on sattumoisin tietoa potilaan elinympäristöstä sekä sen vaikutuksista (Voigt, 2017).

Voigtin (2017) mukaan tämänkaltaisen sosioekonomisiin tekijöihin liittyvä data on sensitiivistä ja esimerkiksi köyhyys tai heikot elinolosuhteet saattavat aiheuttaa häpeää. Lisäksi sosiaaliin tekijöihin liittyvä data saattaa aiheuttaa huolta siitä, että se vaikuttaa negatiivisesti ammattilaisten suhtautumiseen. Pahimmassa tapauksessa siis massadata tässä kontekstissa saattaisi lisätä epätasaa entisestään. Toisaalta kuitenkin saattaa olla haitallista, mikäli sosioekonomisesta asemasta kertova data jätetään hyödyntämättä. Datan avulla heikossa asemassa olevien potilaiden ei tarvitse käydä mahdollisesti häpeälliseksi koettuja asioita läpi lääkärin kanssa, koska lääkäri on jo datan perusteella tietoinen potilaan tilanteesta. (Voigt, 2017.)

Lisäksi Voigtin (2017) mukaan datan hyödyntämättä jättämisessä on se riski, että hoitohenkilökunta tekee yksinkertaistettuja oletuksia potilaiden taustoista. Tässä saattaa tapahtua virheellisiä oletuksia ja tästä syystä potilaiden tilanne voi olla heikompi verrattuna tilanteeseen, jossa lääkäreillä olisi mahdollisuus hyödyntää oikeaa dataa potilaiden ympäristöstä tai muista tekijöistä, jotka vaikuttavat terveyteen. (Voigt, 2017.) Eriarvoisuutta ei voida tietenkään vähentää ainoastaan massadatan avulla, mutta se voi toimia apuvälineenä, joka lisää hoitohenkilökunnan informaatiota ja siten potilaiden taustat voidaan huomioida paremmin hoitoprosessissa.

4.4 Massadata ja terveystaloustiede

Terveystaloustieteen tutkimus pyrkii arvioimaan terveydenhuoltojärjestelmän muutosten ja toimenpiteiden arvoa vertaamalla tuloksia sekä kustannuksia. Perinteisesti tutkimuksessa on keskitytty kustannustehokkuuteen tai esimerkiksi uuden lääkkeen budjettivaikutuksiin (Chen ym., 2016). Massadatalta on todettu potentiaalia terveystaloustieteen tieteenalan edistämiseksi (Collins, 2016). Terveystaloustiede tieteenalana on puolestaan tärkeä, kun halutaan kehittää mahdollisimman toimivaa terveydenhuoltoa.

Collins (2016) nimeää tutkimuksessaan useita hyötyjä ja haasteita liittyen massadataan terveystaloustieteessä. Hyödyksi on tunnistettu mahdollisuus mallintaa vakaampia pitkän aikavälin terveystaloustieteellisiä ennusteita, kun hyödynnetään avointa massadataa. Lisäksi lääkkeiden kustannustehokkuutta voidaan arvioida paremmin, kun massadatan avulla saadaan tietoa siitä, kuinka lääkkeet toimivat käytännössä. Lääkehankkeet ovat usein hintavia, mutta massadatan avulla saadaan tietoa tyypillisistä mekanismeista ja sivuvaikutuksista: tämän pitäisi vähentää uusien epäonnistuneiden lääkehankkeiden määrää. (Collins, 2016.)

Collins (2016) mainitsee tutkimuksessaan myös personoidun terveydenhuollon, jota käsiteltiin tutkielman aikaisemmassa alaluvussa. Terveystaloustieteen näkökulmasta personoitu terveydenhuolto sekä personoidut lääkkeet tehostavat toimintaa sekä vähentävät kustannuksia. Lisäksi massadatan avulla lääkkeiden vaikutuksia yksilöihin voidaan ennustaa paremmin: lääkäreillä on siis enemmän tietoa, jonka pohjalta voidaan tehdä järkevämpiä päätöksiä. Tämä puolestaan tarkoittaa sitä, että terveydenhuolto toimii kustannustehokkaammin. Hyödyksi tunnistetaan myös suurten tietoaaineistojen parempi kommunikointi analytiikan avulla. Tämä puolestaan helpottaa huomattavasti monimutkaisten analyysien tekemistä. (Collins, 2016.)

Collinsin (2016) mukaan täytyy kuitenkin myös huomioida, että suurten tietomäärien tallentaminen ja käsittely voi olla kallista ja massadata-analytiikassa voi syntyä myös virheellisiä assosiaatioita. Massadata-analyysi ei myöskään täytä tarvittavaa tieteellistä tarkkuutta satunnaistetuissa ja kontrolloiduissa kokeissa, eikä siten korvaa niitä. Lisäksi ekonomistit tarvitsevat uusia taitoja massadatan kanssa työskentelyyn, mikä aiheuttaa lisäkustannuksia. (Collins, 2016.) Toisaalta kuitenkin täytyy huomioida, että vaikka massadata-analyysi ei täytä tarvittavia tieteellisiä vaatimuksia, sitä voidaan silti käyttää hyödyllisenä apuvälineenä. Lisäksi ekonomistien lisäkoulutus voidaan nähdä myös investointina, joka myöhemmin maksaa itsensä takaisin.

5 MASSADATAN HYÖDYNTÄMISEEN LIITTYVÄT HAASTEET

Edellisissä luvuissa tunnistettiin useita massadataan liittyviä hyötyjä, mutta ratkaistavana on silti vielä myös useita haasteita. Tässä luvussa käsitellään terveydenhuollon massadataan liittyviä tietoturvaan, yksityisyyteen sekä datan laatuongelmiin liittyviä haasteita. Lisäksi käsitellään datan keräämiseen ja omistajuuteen liittyviä eettisiä ongelmia.

5.1 Tietoturvaongelmat

Kyberturvallisuuskeskuksen (2020) määritelmän mukaan tietoturvallisuudella tarkoitetaan tiedon luottamuksellisuuden, eheyden sekä käytettävyyden varmistamista. Luottamuksellisuuden periaatteella tarkoitetaan sitä, että tietoja hyödyntävät vain tahot, jotka ovat siihen oikeutettuja. Eheydellä tarkoitetaan sitä, että tietoja voivat muokata vain siihen oikeuden saaneet. Käytettävyyden periaatteella varmistetaan, että tiedot ovat käyttöön oikeutettujen käytettävissä. Tietoturvallisuus varmistetaan hallinnollisten sekä teknisten toimien avulla. (Kyberturvallisuuskeskus, 2020.) Vastaavasti tietoturvaongelmat ovat puutteita edellä mainituissa tietoturvallisuuden periaatteissa. Karimin ym:n (2018) mukaan massadatan kontekstissa tietoturvallisuudella tarkoitetaan datan suojaamista väärinkäytöltä sekä tietomurroilta. Terveydenhuollon massadatalla voidaan saavuttaa merkittäviä hyötyjä, mutta haittapuolina ovat tietoturvallisuuden ja riittävän teknisen tuen puute. Terveydenhuollon datan on tunnistettu olevan erittäin altis tietomurroille. (Karim ym., 2018.)

5.1.1 Tietoturvaongelmien aiheuttamat haasteet

Tietoturvallisuuden haasteita tarkastellessa on järkevää erotella yksilöihin, organisaatioihin sekä kansantalouteen kohdistuvat vaikutukset (Lehtilä, Nyström,

Ronikonmäki & Sirviö, 2021). Lehto, Pöyhönen ja Lehto (2019) toteavat tutkimuksessaan, että terveydenhuollon toimialalla tietoturvallisuuden kohdistuu erityisiä vaatimuksia, koska haavoittuvuudet tietoturvallisuudessa voivat johtaa esimerkiksi potilasturvallisuuden vaarantumiseen. Turvallisen hoidon kannalta on välttämätöntä, että potilastietojen saatavuus ja eheys varmistetaan. Lisäksi tietoturvallisuutta täytyy korostaa erityisesti henkilötietojen rikollisen käytön torjumiseksi sekä yksityisyyden vuoksi. (Lehto, Pöyhönen & Lehto, 2019.)

Lehdon ym:n (2017) mukaan potilastiedot sisältävät tyypillisesti esimerkiksi luottokorttinumeroita, työnantajatietoja sekä sairaushistoriatietoja. Vastaavanlainen data on usein voimassa vuosia ja rikolliset toimijat voivat käyttää tällaista dataa identiteettivarkauksissa, petoksissa sekä tietojenkalasteluhyökkäyksissä. (Lehto ym., 2017.) Hyökkääjät käyttävät toiminnassaan tiedonlouhintatekniikoita ja -menetelmiä arkaluontoisen datan löytämiseksi (Karim ym., 2018).

Vuonna 2020 tapahtunut Psykoterapiakeskus Vastaamon tietomurto on varoittava esimerkki siitä, mitä voi tapahtua, jos tietoturvallisuudesta ei huolehdi asianmukaisesti. Suur-Uskin ja Rydenfeltin (2020) mukaan tietomurron vuoksi kymmenien tuhansien ihmisten arkaluontoisia tietoja päätyi väärille tahoille. Vastaavan kaltaiset tapaukset voivat myös heikentää ihmisten luottamusta terveydenhuollon palveluihin sekä vähentää hoitoon hakeutumista. (Suur-Uski & Rydenfelt, 2020.) Myös Lehtilän ym:n (2021) mukaan tietoturvallisuuden liittyvät ongelmat voivat heikentää palveluiden toimintaa sekä niihin kohdistuvaa luottamusta. Lisäksi tietoturvaongelmat voivat aiheuttaa organisaatioille haitallista mainetta, joka puolestaan mahdollisesti voi vaikuttaa vaikkapa asiakkaiden määrään. (Lehtilä ym., 2021.)

Lehtilän ym:n (2021) mukaan tietoturvaongelmat aiheuttavat myös kansantaloudellisia kustannuksia. Riittävän tietoturvallisuuden ratkaisut voidaan nähdä kannustinongelmana, koska yksittäisillä toimijoilla ei välttämättä ole riittävää kannustinta investoida tietoturvaan. Tämän haasteen on tunnistettu johtuvan siitä, että tietoturvallisuuden liittyvissä ongelmatilanteissa osa seurauksista siirtyy yhteiskunnan kannatettavaksi. Tietoturvaongelmissa on siis negatiivisia ulkoisvaikutuksia, koska vastuu ei kasaudu täysin yksittäisille toimijoille. Psykoterapiakeskus Vastaamon tapaus muistuttaa edellä kuvailtua, koska siinä tietomurron seuraukset jäivät yksittäisten henkilöiden ja yhteiskunnan kannateltavaksi. Toisaalta edellä kuvattuun haasteeseen liittyy myös yksityisyysparadoksi (Lehtilä ym., 2021). Yksityisyysparadoksilla viitataan tapauksiin, joissa yksittäiset toimijat väittävät olevansa kiinnostuneita tietoturvallisuudesta, mutta toiminnassa riittävä tietoturvallisuus jää kuitenkin toteutumatta (Kokolakis, 2017). Lehtilän ym:n (2021) mukaan Psykoterapiakeskus Vastaamon tapauksessa yksittäiset henkilöt joutuivat pelkäämään tietojen väärinkäyttöä ja toisaalta yhteiskunnan tehtäväksi jäi tukipalveluiden sekä neuvonnan tarjoaminen tietomurron uhreille. Tapauksessa tunnistettiin yksityisyysparadoksin sekä kannustinongelman aiheuttamia haasteita. (Lehtilä ym., 2021.)

Yhteenvedona siis tietoturvaongelmat voivat aiheuttaa yksilöille haittaa identiteettivarkauksien, petoksien sekä tietojenkalastelun muodossa. Lisäksi tietoturvaongelmat aiheuttavat erilaisia yksityisyyteen liittyviä haasteita, joita tarkastellaan tarkemmin myöhemmässä alaluvussa. Tietoturvaongelmat voivat myös heikentää potilasturvallisuutta, vähentää terveydenhuoltojärjestelmään kohdistuvaa luottamusta sekä aiheuttaa taloudellista vahinkoa. Lisäksi tietoturvaongelmien on tunnistettu aiheuttavan kansantaloudellisia kustannuksia sekä negatiivisia ulkoisvaikutuksia.

5.1.2 Tietoturvan parantaminen

Tietoturvallisuuteen vaikuttavien toimenpiteiden prosessi on monimuotoinen. Lehtilän ym:n (2021) mukaan tietoturvallisuuden tulisi olla erottamaton ja sisäänrakennettu ominaisuus erilaisissa teknologisissa ratkaisuissa. Tietoturvallisuus täytyy siis ottaa huomioon teknologisen ratkaisun koko elinkaaren aikana. (Lehtilä ym., 2021.) Karimin ym:n (2018) mukaan terveydenhuollon organisaatioilta vaaditaan laitteistoja ja ohjelmistoja turvaamaan massadatan turvallisuutta. Lisäksi tarvitaan kliinistä sekä hallinnollista tietoa sisäisistä ja ulkoisista riskeistä. (Karim ym., 2018.) Edellä mainittujen tekijöiden lisäksi tietoturvallisuus muodostuu tietoturvalaeista ja -asetuksista, toimintakulttuurista, tilanteen ja toimintaympäristön ymmärtämisestä sekä relevantista yhteistyöstä eri tahojen kanssa (Lehtilä ym., 2021).

Lehtilä ym. (2021) esittävät tutkimuksessaan, että tietoturvallisuuden toteutumiseen voidaan vaikuttaa sääntelemällä mekanismeja ja luomalla toimijoille kannusteet tietoturvaliselle toiminnalle. Kannustinjärjestelmä voi sisältää sekä ennaltaehkäiseviä että jälkikäteisiä toimia. Ennaltaehkäiseviin toimiin kuuluvat esimerkiksi tietosuojalait ja -asetukset sekä vaaditun tietosuojan tason aiheuttamat kannustimet investoida tietoturvaratkaisuihin. Investoinnit puolestaan pienentävät tietomurtojen riskiä, mutta aiheuttavat myös kustannuksia. Toisaalta kuitenkin ennakointi tulee todennäköisesti edullisemmaksi kuin se, että tietoturvariskeihin ei varauduttaisi etukäteen. (Lehtilä ym., 2021.)

Lehtilän ym:n (2021) mukaan jälkikäteiset toimet pitävät sisällään vastuunkantamista, eli esimerkiksi haittojen korvaamista tietomurron uhreille. Tällainen ratkaisu kannustaa toimijoita investoimaan tietoturvaratkaisuihin sekä sisäistämään haittoja, joita tietomurto voi aiheuttaa yksilöille. (Lehtilä ym., 2021.) Tietoturvallinen massadatan hyödyntäminen terveydenhuollossa lähtee siis mekanismeista ja kannustimista. Lisäksi organisaatiossa tulee ottaa huomioon massadatan koko elinkaarenaikainen tietoturvallisuus.

Karim ym. (2018) esittävät tutkimuksessaan mallin, jonka avulla voidaan torjua tietoturvauhkia massadatan elinkaaren eri vaiheissa. Massadatan elinkaari terveydenhuollossa jakautuu tässä mallissa neljään eri vaiheeseen, jotka ovat: (1) datan keräämisvaihe, (2) transformaatiovaihe, (3) mallintamisvaihe sekä (4) tietämyksen luomisen vaihe. Jokaisessa vaiheessa tietoturvallisuus tulee ottaa huomioon. (Karim ym., 2018.)

Karimin ym:n (2018) mukaan datan keräämisvaiheessa dataa tulisi kerätä ainoastaan luotettavista lähteistä ja lisäksi täytyy huomioida, että potilaan yksityisyys säilyy. Täytyy siis varmistaa, että datamassasta ei voida tunnistaa ketään yksittäistä henkilöä. Transformaatiovaiheessa data transformoidaan helpommin käsiteltävään muotoon. Tietoturvallisuuden näkökulmasta täytyy ottaa huomioon, että kerätty data voi sisältää arkaluontoista tietoa. Turvallisuuden takaamiseksi täytyy ylläpitää käyttöoikeustasoa sekä pääsynhallintaa. (Karim ym., 2018.)

Karimin ym:n (2018) esittelemän mallin mukaan datan tietoturvallisen keräämisen, muuntamisen ja tallentamisen jälkeen siirrytään mallinnusvaiheeseen. Mallinnusvaiheessa datasta tehdään mallinnuksia ja ennusteita erilaisten tiedonlouhintatekniikoiden avulla. Tietoturallinen datankäsittely-ympäristö on ensiarvoisen tärkeää. Tiedonlouhinnassa käytetään algoritmeja, jotka voivat poimia arkaluontoisia tietoja. Tiedonlouhinnan prosessi ja verkkokomponentit täytyy konfiguroida sekä suojata tiedonlouhintaan perustuvilta hyökkäyksiltä ja kaikilta tietoturvaloukkauksilta. Prosessin aikana täytyy myös huomioida, että datan kanssa työskentelee ainoastaan valtuutettu ja luotettava henkilökunta. (Karim ym., 2018.)

Lopuksi Karimin ym:n (2018) mallissa siirrytään tietämyksen luomisen vaiheeseen. Tässä vaiheessa saadaan uutta ja arvostettua tietoa, joka on kuitenkin arkaluontoista etenkin kilpailuympäristössä. Näin ollen turvallisuuden noudattaminen on ensisijainen tavoite tässäkin vaiheessa. (Karim ym., 2018.)

Tietoturvallisuuden takaamisessa tulisi siis ottaa huomioon taloudelliset mekanismit sekä kannusteet. Lisäksi täytyy huomioida datan anonymisointi, luotettavat datalähteet sekä tietoturvalliset datankäsittely-ympäristöt, joihin on pääsy ainoastaan rajatulla henkilöstöllä. Tietoturvallisuudesta huolehtiminen on olennainen osa myös ehkäpä suurimmaksi tunnistettua massadataan liittyvää haastetta, eli yksityisyyttä.

5.2 Yksityisyys

TerveYTEEN liittyvä data on henkilökohtaista ja arkaluontoista. Mikäli datan keräämiseen ja hyödyntämiseen on suostumus, täytyy silti huomioida yksityisyyteen liittyviä seikkoja. Karimin ym:n (2018) määritelmän mukaan yksityisyydellä tarkoitetaan massadatan kontekstissa arkaluontoisen datan suojelua siten, että yksityishenkilöä ei voida tunnistaa datamassasta. Lisäksi yksityisyydellä tarkoitetaan henkilöiden mahdollisuutta päättää, kuinka heitä koskevaa dataa hyödynnetään. Esimerkiksi dataa ei voida luovuttaa kolmansille osapuolille ilman käyttäjän suostumusta. Yksityisyyden turvaamisessa keskitytään erityisesti käytäntöjen sekä lupavaatimusten kehittämiseen, jotta voidaan varmistaa, että henkilötietoja kerätään, jaetaan ja käytetään oikein. (Karim, 2018.) Pricen ja Cohenin (2019) määritelmän mukaan yksityisyydensuojan rikkominen tapahtuu silloin, kun käytäntöjä ja lupavaatimuksia ei noudateta oikein. Yksityisyydensuojan rikkominen ei tarkoita välttämättä tietomurtoa, vaan vaikkapa sitä,

että väärällä toimijalla on pääsy dataan. (Price & Cohen, 2019.) Käytännössä siis yksityisyydensuoja ei toteudu, jos vaikkapa organisaation sisällä jollakin on pääsy tietoihin ilman relevanttia tarkoitusta.

Pricen ja Cohenin (2019) tutkimuksen mukaan yksityisyydensuojan rikkomukset aiheuttavat niin seuraamuksellisia kuin deontologisiakin haasteita. Seuraamuksellisilla haasteilla tarkoitetaan niitä ongelmia, joita yksityisyydensuojan rikkomuksesta aiheutuu yksilölle. Mikäli vaikkapa henkilön arkaluontoiset terveystiedot päätyvät väärin tahojen tietoon ja leviävät, se voi aiheuttaa konkreettisia haasteita, kuten syrjintää tai sosiaalista stigmaa. (Price & Cohen, 2019.)

Myös Dochertyn (2014) mukaan terveystietojen väärinkäyttöön perustuva syrjintä on tunnistettu yhdeksi merkittäväksi haasteeksi. Terveystietojen hyödyntämisellä on tarkoitus parantaa hoitoa, mutta silti tietoja voidaan käyttää väärin aiheuttaen potilaalle syrjintää ja stigmaa. (Docherty, 2014.) Mikäli terveystietoja hyödynnetään epäeettisesti, se saattaa Pricen ja Cohenin (2019) mukaan vaikuttaa esimerkiksi yksilön työllistymiseen. Mikäli vaikkapa työnantaja saa tietoonsa, että työnhakijalla on jokin mahdollisesti kustannuksia aiheuttava sairaus, saattaa muuten pätevä työntekijä jäädä palkkaamatta. (Price & Cohen, 2019.) Lisäksi yksityisten tietojen leviäminen voi aiheuttaa yksilölle merkittävää häpeää, vainoharhaisuutta tai muita mielenterveyden haasteita (Ford & Price, 2016). Ahdistusta ja distressiä voi aiheuttaa myös pelkästään tieto siitä, että yksityiset terveystiedot ovat jossakin muiden saatavilla, ennen kuin mahdollista tietojen väärinkäyttöä on edes tapahtunut (Price & Cohen, 2019). Edellä kuvatuissa tapauksissa massadatan hyödyntäminen ei siis edistäisi terveyttä ja hyvinvointia, vaan saattaisi jopa aiheuttaa terveydellisiä ongelmia.

Pricen ja Cohenin (2019) mukaan deontologiset huolenaiheet eivät ole riippuvaisia kielteisten seurausten kokemisesta, vaan haaste on periaatteellinen. Toisin sanoen siis ongelma ilmenee, vaikka henkilö ei edes tietäisi tietomurrosta. Esimerkiksi oletetaan, että jokin taho tietomurron tehtyään pyrkii hyötymään häikäilemättömästi arkaluontoisesta datasta. Mikäli tämä tapahtuu esimerkiksi ulkomailla ja yksityishenkilö ei saa ikinä tietää tästä, seuraamuksellisia ongelmia ei todennäköisesti ilmene. Vastaavanlainen datan väärinkäyttö on kuitenkin eettinen ongelma, vaikka varsinaista vahinkoa ei tapahtuisikaan. (Price & Cohen, 2019.)

Massadatan hyödyntämiseen liittyy siis useita yksityisyyteen liittyviä haasteita. Yksityisyyttä suojelevat erilaiset tietosuojalait sekä -asetukset, jotka poikkeavat toisistaan eri valtioissa (Karim ym., 2018). Tietosuojalait sekä -asetukset ovat tärkeä ja erittäin laaja osa tietoturvallisuuden ja yksityisyyden turvaamista, mutta niiden tarkempi käsittely on rajattu tämän tutkielman ulkopuolelle. Täytyy kuitenkin myös huomioda, että vaikka toimittaisiin tietosuojalain ja -asetusten puitteissa, saattaa silti ilmetä yksityisyyden haasteita.

Edellä kuvailtuja haasteita voidaan eliminoida, mikäli hyödynnettävästä massadatasta pystytään poistamaan henkilötiedot (Karim, 2018). Yleisesti anonymisoinnin (engl. de-identification) ideana on saada hyötykäyttöön merkittävä data siten, että henkilöä ei voida identifioida datan perusteella. Datan ano-

nymisoitiin on olemassa useita eri tapoja ja tekniikoita (Sedayao, Bhardwaj & Gorade, 2014). Article 29 Data Protection Working Party (2014) mukaan datan anonymisointia voidaan lähestyä satunnaistamisen tai yleistämisen näkökulmasta. Satunnaistamiseen perustuvilla tekniikoilla pyritään poistamaan yhteys muun datan ja tunnistettavan datan välillä. Yleistämiseen perustuvilla tekniikoilla pyritään yleistämään ja epätarkentamaan dataa esimerkiksi muuttamalla asteikkoa tai suuruusluokkaa. Käytännössä tämä voisi tarkoittaa sitä, että datasta ilmenee vaikkapa yksilön asuinpaikkaan liittyen maakunta kaupungin sijaan. Lähestymistapoja ja niihin perustuvia tekniikoita voidaan myös yhdistää keskenään. Lisäksi molempiin lähestymistapoihin liittyy hyötyjä ja haasteita ja on aina kontekstisidonnaista, mitä tekniikkaa tai niiden yhdistelmää tulisi käyttää. (Article 29 Data Protection Working Party, 2014.)

Karimin ym:n (2018) mukaan anonymisointi ei kuitenkaan välttämättä ole toimiva ratkaisu, mikäli dataa väärinkäyttävällä taholla on mahdollisuus mitätöidä anonymisointi ja päästä käsiksi yksityiseen dataan. Tällaisessa tapauksessa on kyse reidentifikaatiosta. (Karim, 2018.) Anonymisointi tulisi siis tehdä mahdollisimman kontekstiin sopivalla menetelmällä siten, että anonymisointia ei voida jälkikäteen purkaa.

Massadatan hallinnassa täytyy siis ottaa huomioon myös yksityisyyden haasteisiin kehitettyihin ratkaisuihin liittyvät puutteet. Lisäksi on todettu, että datan anonymisoinnin tekniikat saattavat aiheuttaa informaation epätarkkuutta (Aggarwal & Yu, 2008). Seuraavassa alaluvussa käsiteltävät massadatan laatuongelmat sekä siitä seuraava informaation mahdollinen epätarkkuus ovatkin yksi terveydenhuollon massadataan liittyvistä haasteista.

5.3 Massadatan laatuongelmat

Sukumarin, Natarajanin ja Ferrellin (2015) mukaan terveydenhuollon massadataa hyödyntäessä tulisi huomioida datan virheistä sekä epätarkkuuksista johtuvat ongelmat. Laatuun liittyviä ongelmia voi syntyä massadatan elinkaaren jokaisessa vaiheessa: datan keräämisen, integroinnin, muuntamisen tai päätelmien vaiheessa. Laadunarvioinnissa on huomioitava datassa esiintyvät virheet ja epätarkkuudet sekä datan lähteet ja taustat. Lisäksi täytyy ottaa huomioon, että tiedonkeruun taustalla olevat tarkoitukset saattavat vaikuttaa analytiikkavaiheeseen. (Sukumar ym., 2015.) Datan hyödyntämiseen yleisesti liittyy laatuongelmia, mutta Clarcken (2016) mukaan massadatan kontekstissa riski haasteisiin kasvaa.

Massadatan ominaisuuksista erityisesti sen monimuotoisuuden sekä alhaisen todenmukaisuuden on todettu aiheuttavan laatuongelmia. (Rao, Gudivada & Raghavan, 2015). Hershin ym:n (2013) mukaan laatuongelmia ovat datan epätäydellisyys ja virheet sekä transformoinnissa heikentynyt informaation merkitys. Lisäksi ongelmaksi on tunnistettu datan tuntematon alkuperä, tarkkuustason riittämättömyys sekä datan keräämiseen liittyvät ongelmat. Haasteena on se, että dataa ei ole kerätty tutkimusprotokollien mukaisesti ja

tekstimuotoinen data on vaikeasti palautettavissa toissijaisia hyödyntämistarkoituksia varten. (Hersh ym., 2013.) Edellä mainittujen haasteiden lisäksi Hoffmanin ja Podgurskin (2013) mukaan laatuongelmia ovat datan hajanaisuus sekä tarkoitushakuisten tai virheellisten tutkimustulosten riski.

Massadatan laatuongelmilla tarkoitetaan siis käytännössä vaikkapa tietojen syöttövirheitä, puuttuvaa dataa tietokentissä sekä analytiikkavaiheessa syntyviä virheitä. Laatuongelmia aiheuttavat niin ihmisten toiminta kuin teknologiatkin. Mikäli hyödynnettävässä datassa on virheitä, on ilmeistä, ettei esimerkiksi aikaisemmassa luvussa esitelty potilasprofiilin malli toimi tarkoituksenmukaisesti. Täydellisesti toimivasta algoritmista ei ole hyötyä, mikäli hyödynnettävässä datassa on virheitä, koska lopputulos on virheellinen. Pahimmassa tapauksessa datan laatuongelmat saattavat johtaa esimerkiksi virheellisiin käsitelmiin potilaan terveysriskeistä tai jopa hoitovirheisiin.

Sukumarin ym:n (2015) mukaan on tyypillistä, että terveydenhuollossa esimerkiksi potilastiedot tallennetaan järjestelmiin käsin. Tämän vuoksi terveydenhuollon massadata on erityisen altista inhimillisille virheille, verrattuna massadataan yleisesti. (Sukumar ym., 2015.) Potilaskertomusten on todettu olevan keskeinen osa terveydenhuollon massadataa, myös toissijaisissa hyödyntämistarkoituksissa (Galetsi, Katsaliaki & Kumar, 2020). Potilaskertomusten kirjaamiseen tulisi kiinnittää erityistä huomiota, jotta niistä voidaan hyötyä myös toissijaisissa käyttötarkoituksissa. Lisäksi Sukumarin ym:n (2015) mukaan ongelmaksi on tunnistettu se, että käytettävä data on alun perin kerätty jotakin muuta tarkoitusta varten. Esimerkiksi osa terveydenhuollon massadatasta on kerätty taloudellisia hyödyntämistarkoituksia varten ja siten ei välttämättä täytä kliinisen tai epidemiologisen tutkimuksen laatuvaatimuksia. (Sukumar ym., 2015.)

Datan laatuun on tärkeää kiinnittää huomiota, koska muuten tavoiteltujen hyötyjen sijaan saattaa aiheutua merkittäviä haittoja sekä kustannuksia. Datan laatuongelmat voivat aiheuttaa terveysriskejä potilaille, mikäli päätöksiä tehdään virheellisen datan perusteella. Vastaavasti mikäli vaikkapa kustannustehokkuuteen tai terveydenhuoltojärjestelmään liittyviä päätöksiä tehdään virheellisen tai puutteellisen datan perusteella, lopputuloksena ei ole optimaalisesti toimiva systeemi. Yksittäisille terveydenhuollon toimijoille massadatan laatuongelmat saattavat aiheuttaa kilpailukyvyn alenemista. Koska massadatan laatuongelmia voi syntyä missä tahansa elinkaaren vaiheessa, täytyy vastaavasti laatuongelmien minimointiin kiinnittää huomiota jokaisen vaiheen aikana. Laadunvarmistamisessa täytyy ottaa huomioon, että siihen vaikuttaa niin ihmisten kuin teknologioidenkin toiminta.

5.4 Datan keräämiseen ja omistajuuteen liittyvät haasteet

Datan keräämiseen ja hyödyntämiseen tulisi olla aina suostumus, joten luvatta tapahtuva datan kerääminen tai hyödyntäminen on jo lähtökohtaisesti epäeettistä. Custersin (2016) mukaan pohdittaessa suostumukseen liittyviä seikkoja,

oletetaan usein, että vain tietoinen suostumus on pätevä. Toisin sanoen siis henkilön tulee olla täysin tietoinen siitä, mihin häntä koskevaa dataa hyödynnetään ja mitä mahdollisia seurauksia sillä voi olla. Jokaisella henkilöllä tulisi olla oikeus päättää, miten, milloin ja mihin häntä koskevaa dataa hyödynnetään. (Custers, 2016.)

Yleisesti lupa datan keräämiseen voidaan saada kahdella eri tavalla. Quinnin (2014) mukaan henkilöltä voidaan erikseen kysyä, suostuuko hän datan keräämiseen ja hyödyntämiseen. Toisaalta yleisemmin käytössä on tapa, jossa dataa kerätään ja hyödynnetään silloin, kun sitä ei erikseen kielletä. (Quinn, 2014.) Terveystietojen kontekstissa tulisi aina käyttää ensimmäisenä mainittua eettisempää tapaa, koska terveyteen liittyvät tiedot ovat niin sensitiivisiä ja arkaluontoisia.

Lisäksi tulisi huolehtia siitä, että tietosuojakäytännöt ovat ymmärrettäviä ja selkeitä. Soloven (2013) mukaan tietosuojakäytäntöjen hyväksymiseen liittyy yleisesti ongelma, että ihmiset eivät lue tietosuojakäytäntöjä tai ymmärrä niitä. Mikäli henkilö lukee ja ymmärtää tietosuojakäytännön, ongelmaksi saattaa puolestaan muodostua riittämätön taustatieto, jota vaadittaisiin tietoisesta päätöksestä tekemiseen. (Solov, 2013.) Vaikka potilaat suostuisivat kliinisten tietojensa kaikkiin toissijaisiin käyttötarkoituksiin, heillä ei todennäköisesti ole riittävää kykyä arvioida riittävän tarkasti datan jakamiseen liittyviä näkökulmia (Ballantyne, 2020).

Custers (2016) pohtii tutkimuksessaan, kuinka kauan dataa kerääväällä taholla on oikeus kerätä dataa tietosuojakäytännön hyväksymisen jälkeen. Terveystietojen kontekstissa tulisi pohtia samaa: mikäli suostumus datan hyödyntämiseen saadaan, kuinka kauan se on voimassa? Ratkaisuna voisi toimia se, että potilas suostuu joksikin määritellyksi ajaksi datan keräämiseen, jonka jälkeen hänellä on mahdollisuus pohtia asiaa uudestaan.

Lisäksi tulisi kiinnittää huomiota datan keräämiseen sekä yhdenvertaisuuteen liittyviin seikkoihin, sillä Lermanin (2013) mukaan osa ihmisistä jää datan keräämisen ulkopuolelle. Heikommissa sosioekonomisissa asemassa olevien dataa ei kerätä, koska elintavat eroavat paremmassa sosioekonomisessa asemassa olevien elintavoista. Tämän seurauksena huono-osaiset ihmiset saattavat olla aliedustettuna datassa. Mikäli päätöksiä tehdään perustuen massadataan, saattavat päätökset suosia paremmassa asemassa olevia ja vastaavasti heikentää entisestään vähäosaisten asemaa. (Lerman, 2013.)

Cullen (2003) esittää tutkimuksessaan, että digitalisaatioon liittyy eriarvoistumista, josta käytetään nimitystä digitaalinen kahtiajako. Digitaalisella kahtiajaolla tarkoitetaan ihmisten jakautumista heihin, joilla on mahdollisuus hyödyntää teknologiaa ja vastaavasti heihin, joilla tätä mahdollisuutta ei ole. Digitaaliseen kahtiajakoon liittyvät erot sosioekonomisessa asemassa, mutta myös toisaalta erot esimerkiksi teknologian käyttötaidoissa. (Cullen, 2003.) Myös terveydenhuollon massadataan liittyvissä ratkaisuissa tulisi pohtia sitä, kenelle hyödyt loppupeleissä kasaantuvat. Esimerkiksi MyData-ratkaisua tarkastellessa tulisi pohtia, kuinka se olisi saavutettavissa kaikille halukkaille. Heikommissa sosioekonomisissa asemassa olevilla ei välttämättä ole mahdolli-

suuksia hyödyntää itsenäisen ja ennaltaehkäisevän terveydenhuollon digitaalisia ratkaisuja, mikäli ei esimerkiksi ole mahdollista hankkia riittävän uutta teknologiaa. Toisaalta voidaan pohtia myös sitä, hyötyvätkö loppupeleissä kaikki ne ihmiset, joiden dataa hyödynnetään, vai kasautuuko hyöty parempiosaisille. Digitaalisten ratkaisujen elinkaaren alussa yksityisellä terveydenhuollolla on todennäköisesti paremmin mahdollisuuksia investoida uuteen teknologiaan. Tässä tapauksessa hyöty kasautuu parempituloisille, jotka hyödyntävät yksityistä terveydenhuoltoa.

Massadataan liittyen usein keskustellaan myös siitä, kuka datan lopulta omistaa. Omistaako datan potilas vai dataa hyödyntävä taho? Ballantynen (2020) mukaan vallitseva käsitys on, että potilaat omistavat oman terveystatansa. Potilaiden lisäksi terveystatasta ovat kiinnostuneita dataa hyödyntävät tahot sekä he, jotka ovat muuten kiinnostuneita informaatiosta. (Ballantyne, 2020.) Useat eri tahot ovat siis kiinnostuneita myös datan omistajuudesta.

Ballantynen (2020) mukaan usein keskustelussa siirrytään suoraan siihen oletukseen, että data kuuluu potilaalle ja potilas päättää, miten dataa käytetään. Tämä lähestymistapa on kuitenkin tunnustettu kapeaksi ja ongelmalliseksi. Tietoturvallisuuteen tähtäävät näkemykset sekä lähestymistapa, jonka mukaan data kuuluu potilaalle, ovat ristiriidassa keskenään. Potilastietojen omistaminen esitetään usein ratkaisuna massadataan liittyviin ongelmiin, kuten yksityisyyden haasteisiin. Usein potilaat pitävät pääsyään kliiniseen dataan merkittävänä tekijänä yksityisyyden säilyttämisessä, mutta potilaat saattavat kuitenkin yliarvioida omaa kykyään suojella dataansa. (Ballantyne, 2020.) Tämä saattaa puolestaan lisätä tietoturvariskien todennäköisyyttä.

Lisäksi on esitetty näkökulma, jonka mukaan potilaalla itsellään tulisi halutessaan olla mahdollisuus myydä dataansa kolmansille osapuolille (Ballantyne, 2020.) Tässä tilanteessa datan omistajuus olisi ainakin henkilöllä itsellään. Toisaalta kuitenkin Ballantynen (2020) mukaan tähän liittyy se oletus, että potilaat eivät saa riittävää hyötyä datansa luovuttamisesta toissijaiseen käyttöön. Toisaalta kuitenkin potilaat hyötyvät paremmasta hoidosta, tehokkaammasta terveydenhuollosta sekä lisääntyneestä informaatiosta liittyen terveyteen. Lisäksi Ballantynen (2020) mukaan tähän näkökulmaan liittyy oletus, että potilaat haluavat hyötyä taloudellisesti terveystatasta. Empiirisissä tutkimuksissa on kuitenkin todettu, että potilaat eivät innostu kliinisen datan kaupallistamisesta. (Ballantyne, 2020.)

Massadatan keräämiseen ja omistajuuteen liittyy lukuisia eettisiä ongelmia, joihin on esitetty erilaisia näkökulmia. Toisaalta jotkut näkökulmat, joita on esitetty ratkaisuihin haasteisiin, saattavat olla huomattavassa ristiriidassa keskenään. Tässä voidaan huomata se, että massadatan hyödyntäminen on vielä suhteellisen uusi ilmiö, minkä vuoksi näkemykset ja toimintatavat eivät ole vielä riittävän kehittyneitä täysin vastuulliseen massadatan hyödyntämiseen.

6 JOHTOPÄÄTÖKSET

Tässä luvussa esitellään tutkimuksen tulokset. Tutkimusongelma muodostui kahdesta tutkimuskysymyksestä, jotka olivat ”Mitä hyötyjä massadatalta voidaan saavuttaa terveydenhuollossa?” ja ”Mitä haasteita massadatan hyödyntämiseen terveydenhuollossa liittyy?”. Ensin käsitellään massadataan liittyviä hyötyjä, jonka jälkeen siirrytään haasteisiin.

Tutkielmassa hyödynnetyn aineiston perusteella merkittäviksi terveydenhuoltoa edistäviksi ratkaisuiksi tunnistettiin personoituun terveydenhuoltoon kuuluvat potilasprofiilit sekä hoitovasteen ennustaminen. Personoidun terveydenhuollon ratkaisut parantavat ja tehostavat hoitoprosesseja ja lisäksi lisääntyneen informaation ansiosta resursseja pystytään kohdentamaan järkevämmiin. Tiivistetysti potilasprofiilien ideana on massadata-analytiikan avulla ennustaa sairastumisriskiä ja vastaavasti hoitovasteen ennustamisessa pyritään löytämään yksittäiselle potilaalle optimaalinen hoitomuoto.

Kansalaisten omaa käsitystä terveydestä ja hyvinvoinnista voidaan edistää MyData-ratkaisun avulla. Ajatuksena on, että tulevaisuudessa terveystietoa olisi enemmän myös kansalaisten hyödynnettävissä. MyData-ratkaisu vahvistaa yksilön roolia hoitoprosessissa sekä sairauksien ennaltaehkäisyssä. Edellä mainituilla ratkaisuilla voidaan parantaa ja tehostaa hoitoprosesseja, optimoida terveydenhuollon resursseja, saavuttaa kansantaloudellisia säästöjä sekä kehittää terveydenhuollon toimintaa ennaltaehkäisevämpään ja potilaskeskeisempään suuntaan.

Kansantaloudelliset säästöt ja niukkojen resurssien tehokkaampi allokointi ovat myös saavutettavissa massadatan avulla. Kansantaloudelliset säästöt syntyvät suorien ja epäsuorien säästömekanismien kautta. Suoria säästöjä syntyy, kun massadataa hyödyntävien ratkaisujen avulla tuottavuus sekä hoito- ja diagnosointiprosessien tehokkuus kasvavat. Lisäksi voidaan saavuttaa työaikäsäästöjä, koska tekoälypohjaiset ratkaisut tehostavat terveydenhuollon tietojärjestelmien käyttöä. Epäsuorat säästöt kertyvät työkyvyttömyyden vähenemisen, kansantautien vähäisemmän esiintyvyyden, sosiaali- ja terveystieteiden tutkimuksen sekä psyykkisen hyvinvoinnin lisääntymisen kautta.

Terveydenhuollossa haasteeksi on tunnistettu liian pitkät odotusajat sekä useista tekijöistä johtuva potilaiden hoitotoimenpiteisiin kuluvan ajan hallitsemattomuus. Massadatan avulla voidaan ennustaa hoitotoimenpiteisiin kuluva aikaa ja optimoida terveydenhuollon ajanhallintaa. Massadatasta voi olla hyötyä myös eriarvoisuuden vähentämisessä, kun hoitoprosesseissa on käytössä enemmän informaatiota potilaiden taustoista. Lisäksi massadata voi edistää terveystaloustiedettä tieteenalana, joka puolestaan johtaa parempiin terveydenhuollon organisoimisiin päätöksiin.

Tiivistetysti siis massadataa hyödyntävillä ratkaisulla voidaan parantaa ja tehostaa hoitoprosesseja, optimoida terveydenhuollon resursseja, saavuttaa kansantaloudellisia säästöjä sekä kehittää terveydenhuoltoa ennaltaehkäisevämmäksi ja potilaskeskeisemmäksi. Merkittäviksi tunnistettuja mahdollisuuksia on koottu alla olevaan taulukkoon (taulukko 1).

TAULUKKO 1 Massadatan mahdollisuudet terveydenhuollossa

Osa-alue	Hyödyt	Lähteet
Potilasprofiilit	Sairastumisriskin ennustaminen, personoitu ja potilaskeskeinen terveydenhuolto	(Chawla & Davis, 2013), (Almobarak & Jaziri, 2019)
Hoitovasteen ennustaminen	Tehokkaampi hoitoprosessi, personoitu ja potilaskeskeinen terveydenhuolto	(Bzdok & Meyer-Lindenberg, 2018), (Chen ym., 2018), (Gillan & Whelan, 2017)
MyData-ratkaisu	Ennaltaehkäisevä ja kansalaisia aktivoiva lähestymistapa	(Antikainen ym., 2016), (Poikola ym., 2018), (THL, 2017)
Terveydenhuollon talous Prosessien tehostuminen	Kansantaloudelliset säästöt Prosessien tehostuminen, hoitoajan ennustaminen, ajankäytön optimointi	(Neittaanmäki ym., 2019) (Chen ym., 2016)
Eriarvoisuus	Paremman informaation avulla eriarvoisuuden huomioiminen	(Bazemore ym., 2016), (Voigt, 2017)
Terveystaloustieteen tutkimus	Vakaampien terveystaloustieteellisten ennusteiden mallinnus, lääkkeiden kustannustehokkuuden arviointi	(Collins, 2016)

Massadatan hyödyntämiseen liittyviksi haasteiksi tutkielmassa tunnistettiin tietoturvaongelmat, yksityisyyden ongelmat, datan keräämiseen ja omistajuuteen liittyvät haasteet sekä datan laatuongelmat. Ongelmat ovat merkittäviä ja vaikka haasteisiin on esitetty myös ratkaisuehdotuksia, riskejä on silti olemassa. Massadataan liittyvät tietoturvaongelmat aiheuttavat potilasturvallisuuden vaarantumista, yksityishenkilöiden henkilötietojen väärinkäyttöä sekä taloudellisia vahinkoja organisaatioille. Lisäksi tietoturvaongelmien todettiin heikentä-

vän ihmisten luottamusta terveydenhuollon palveluihin, aiheuttavan kansantaloudellisia kustannuksia sekä negatiivisia ulkoisvaikutuksia.

Tietoturvaongelmat aiheuttavat myös yksityisyyden haasteita. Yksityisyyden haasteet saattavat aiheuttaa yksityishenkilöille seuraamuksellisia tai deontologisia ongelmia. Näihin kuuluvat esimerkiksi syrjintä ja sosiaalinen stigma. Haasteeksi tunnistettiin myös massadataan liittyvät laatuongelmat: mikäli datassa esiintyy virheitä, puutteita tai epätarkkuuksia, saattaa tavoiteltujen hyötyjen sijaan aiheutua haasteita. Datan laatuongelmat voivat aiheuttaa terveysriskejä potilaille sekä systeemin epäoptimaalista toimivuutta, koska massadataa hyödyntävät ratkaisut eivät tällöin toimi tarkoituksenmukaisesti.

Datan keräämiseen ja omistajuuteen liittyy lukuisia eettisiä haasteita ja kysymyksiä, kuten kysymyksiä datan keräämisen epätasa-arvosta sekä datan omistajuudesta. Lisäksi haasteita aiheuttavat monimutkaiset tietosuojakäytännöt ja puutteet ihmisten tietämyksessä. Haasteena on myös se, että kun esitetään ratkaisuja eri haasteisiin, ratkaisut ovat usein ristiriidassa keskenään. Merkittäviksi tunnistettuja massadatan hyödyntämiseen liittyviä haasteita on koottu alla olevaan taulukkoon (taulukko 2).

TAULUKKO 2 Massadatan haasteet terveydenhuollossa

Osa-alue	Haitat	Lähteet
Tietoturvaongelmat	Potilasturvallisuuden vaarantuminen, henkilötietojen väärinkäyttö, taloudelliset vahingot, ihmisten luottamuksen heikentyminen, kansantaloudelliset kustannukset ja negatiiviset ulkoisvaikutukset	(Lehto ym., 2017), (Lehto ym., 2019), (Lehtilä ym., 2021)
Yksityisyyden haasteet	Seuraamukselliset ja deontologiset haasteet, syrjintä ja sosiaalinen stigma	(Docherty, 2014), (Price & Cohen, 2019)
Massadatan laatuongelmat	Massadataa hyödyntävien ratkaisujen virheellinen toiminta	(Sukumar ym., 2015)
Datan kerääminen	Epäselvät tietosuojakäytännöt, voimassaoloaika liittyen suostumukselliseen datan keräämiseen, datan keräämiseen liittyvä epätasa-arvo	(Custers, 2016), (Lerman, 2013), (Solov, 2013)
Datan omistajuus	Epäselvyys liittyen datan omistajuuteen	(Ballantyne, 2020)

7 YHTEENVETO

Tutkielman tavoitteena oli muodostaa kattava yleiskatsaus massadatan terveydenhuollossa – tutkielmassa käsiteltiin merkittäviksi tunnistettuja mahdollisuuksia sekä haasteita. Tutkielmassa tarkasteltiin massadataa terveydenhuollon toiminnan tehostamisen, terveyden ja hyvinvoinnin edistämisen sekä kansantalouden näkökulmasta. Lisäksi käsiteltiin haasteita ja ongelmakohtia, jotka tulee ottaa huomioon massadatan hyödyntämisessä.

Massadatan terveydenhuollossa liittyy tutkimuksen perusteella useita mahdollisuuksia sekä haasteita. Johtopäätökset-luvussa esiteltiin tutkimustulokset, mutta aiheen laajuuden vuoksi tämän tutkimuksen esitetyt tulokset eivät sisällä kaikkia mahdollisia aiheeseen liittyviä mahdollisuuksia ja haasteita. Tässä kandidaatintutkielmassa ei olisi ollut mahdollista tai järkevää käydä läpi jokaista osa-aluetta yksityiskohtaisesti. Esimerkiksi molekyyli lääketiede ja bioinformatiikka ovat massadataa hyödyntäviä ja tärkeitä terveydenhuollon kehittämiseen liittyviä osa-alueita: nämä osa-alueet on kuitenkin aihepiirin laajuuden vuoksi rajattu tämän tutkielman ulkopuolelle. Tutkielman aiheita voitaisiin tarkastella tarkemmin niin tietojärjestelmätieteen, taloustieteen, tietotekniikan, terveystieteen kuin lääketieteenkin näkökulmista. Tutkielman rajoitteena on siis laaja aihepiiri, jota on ollut pakko rajata ja näin ollen osa olennaisista massadatan sovellusalueista on jäänyt tutkielman ulkopuolelle. Kuitenkin tutkielmassa käsitellyt osa-alueet muodostavat kattavan yleiskatsauksen aiheesta.

Koska yhteiskunta on kompleksinen, ei terveyteen ja terveydenhuoltoon liittyviä ongelmia voida ratkaista yksittäisillä ratkaisuilla tai tekijöillä. Haasteet ovat syvällä ja esimerkiksi terveydenhuollon tuottavuutta voidaan tehostaa, mutta ihmisten terveyteen vaikuttavat lukuisat muutkin eri asiat. Teknologiset ratkaisut voivat toimia kuitenkin osana terveydenhuollon ja yhteiskunnan kehittymistä ja sairastumisen sekä vaikkapa syrjäytymisen riskiä voidaan ennakoida aiempaa paremmin. Toisaalta tulee myös pohtia sitä, kuinka paljon on järkevää investoida teknologisiin ratkaisuihin ja kuinka paljon muihin terveyteen vaikuttaviin tekijöihin. Teknologisiin investointeihin liittyvät vaihtoehtois-kustannukset eivät saa kasvaa liian suuriksi. Lisäksi riskinä saattaa olla myös

kansantaloudellinen hyvinvointitappio, mikäli palveluiden tuottamisen lisäkustannukset ylittävät potilaan saaman lisähyödyn arvon.

Tutkimusta massadatan hyödyntämisestä terveydenhuollossa on tarpeen jatkaa tulevaisuudessa. Aihe on erittäin laaja ja eri osa-alueita pitäisi tutkia spesifisti, mikäli teknologisista ratkaisuista halutaan mahdollisimman hyvin toimivia. Erityisen tärkeä tutkimusaihe on hoitovasteen ennustaminen, sillä se voisi toimia mullistavana tekijänä terveydenhuollossa niin potilaiden kuin systeemin toiminnankin näkökulmasta. Potilasprofiilien malleissa erot syntyvät tyypillisesti siitä, mitä massadatan osajoukkoa hyödynnetään. Jatkossa olisi hyödyllistä tutkia, mikä malli toimii parhaiten ja mitä massadatan osajoukkoa on hyödyllistä analysoida sairastumisriskin kartoituksessa tai hoitovasteen ennustamisessa.

Massadataan liittyvät haasteet ovat tutkielmassa hyödynnetyn aineiston perusteella selkeät. Jatkossa tulisi tutkia erityisesti haasteiden ratkaisemiseen tunnistettuja vaihtoehtoja sekä niiden toimivuutta. Lisäksi tulisi innovoida uusia ratkaisuja ja tehdä tutkimusta niiden toimivuudesta. Myös terveydenhuollon digitalisaatioon sekä siihen liittyviin eriarvoisuuden riskeihin liittyen tulisi tehdä tutkimusta. Tutkimusta näihin edellä lueteltuihin voitaisiin tehdä poikkitieteellisesti tai vastaavasti useilla eri tieteenaloilla: siten saataisiin monitieteellisiä näkökulmia siihen, kuinka terveydenhuollon digitalisaation sekä massadata-analytiikan kanssa kannattaa edetä.

LÄHTEET

- Almobarak, A.S. & Jaziri, W.A. (2019). Considering the context and patient's profile to support ubiquitous healthcare systems. (s. 147-152) doi:10.1109/DeSE.2019.00036
- Abouelmehdi, K. (2018). Big healthcare data: Preserving security and privacy. *Journal of Big Data*, 5(1), 1-18. doi:10.1186/s40537-017-0110-7
- Aggarwal, C. C. & Philip, S. Y. (2008). A general survey of privacy-preserving data mining models and algorithms. *Privacy-preserving data mining* (s. 11-52) Springer.
- Agrawal, R. & Prabhakaran, S. (2020). Big data in digital healthcare: Lessons learnt and recommendations for general practice. *Heredity*, 124(4), 525-534.
- Akhtar, S. M. F. (2018). *Big data architect's handbook: A guide to building proficiency in tools and systems used by leading big data experts*. Packt Publishing Ltd.
- Andreu-Perez, J., Poon, C. C. Y., Merrifield, R. D., Wong, S. T. C. & Yang, G. (2015). Big data for health. *IEEE Journal of Biomedical and Health Informatics*, 19(4), 1193-1208. doi:10.1109/JBHI.2015.2450362
- Antikainen, J., Eskelinen, J., Koski, H., Niemi, T., Pajarinen, M., Pyykkönen, S. & de Vries, M. (2016). *Massadatasta liiketoimintaa ja tehokkaita julkisia palveluja*. Valtioneuvoston kanslia.
- ARTICLE 29 DATA PROTECTION WORKING PARTY. (2014). Opinion 05/2014 on anonymisation techniques.
- Balducci, B. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science*, 46(4), 557-590. doi:10.1007/s11747-018-0581-x
- Ballantyne, A. (2020). How should we think about clinical data ownership? *Journal of Medical Ethics*, 46(5), 289-294.
- Bazemore, A. W., Cottrell, E. K., Gold, R., Hughes, L. S., Phillips, R. L., Angier, H., . . . DeVoe, J. E. (2016). "Community vital signs": Incorporating geocoded social determinants into electronic records to promote patient and population health. *Journal of the American Medical Informatics Association*, 23(2), 407-412.
- Bzdok, D. & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, 3(3), 223. doi:10.1016/j.bpsc.2017.11.007

- Chawla, N. V. & Davis, D. A. (2013). Bringing big data to personalized healthcare: A patient-centered framework. *Journal of General Internal Medicine*, 28(3), 660-665.
- Chen, J., Li, K., Tang, Z., Bilal, K. & Li, K. (2016). A parallel patient treatment time prediction algorithm and its applications in hospital queuing-recommendation in a big data environment. *IEEE Access*, 4, 1767-1783.
- Chen, Y., Guzauskas, G. F., Gu, C., Wang, B., Furnback, W. E., Xie, G., . . . Garrison, L. P. (2016). Precision health economics and outcomes research to support precision medicine: Big data meets patient heterogeneity on the road to value. *Journal of Personalized Medicine*, 6(4), 20.
- Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, 26(1), 77-90.
- Collins, B. (2016). Big data and health economics: Strengths, weaknesses, opportunities and threats. *PharmacoEconomics*, 34(2), 101-106.
- Cullen, R. (2003). The digital divide: A global and national call to action. *The Electronic Library*, 21(3), 247-257. doi:10.1108/02640470310480506
- Custers, B. (2016). Click here to consent forever: Expiry dates for informed consent. *Big Data & Society*, 3(1) doi:10.1177/2053951715624935
- Daniel J. Solove. (2013). Privacy self-management and the consent dilemma. *Harvard Law Review*, 126, 1879-2479.
- Docherty, A. (2014). Big data – ethical perspectives. *Anaesthesia*, 69(4), 390-391. doi:10.1111/anae.12656
- Eläketurvakeskus. (2020, 8. huhtikuuta). Mielenterveyden sairaudet yleisin työkyvyttömyyseläkkeelle siirtymisen syy. Haettu osoitteesta <https://www.etk.fi/ajankohtaista/mielenterveyden-sairaudet-yleisin-tyokyvyttomyyselakkeelle-siirtymisen-syy/>
- Erhola, M., Jormanainen, V., Kovanen, M., Rissanen, P. & Keskimäki, I. (2020). Suomen terveydenhuolto muuttuvassa toimintaympäristössä. *Yhteiskuntapolitiikka* : 1.
- Findikaattori. (1.4.2020). Työkyvyttömyyseläkettä saaneet. Haettu osoitteesta <https://findikaattori.fi/fi/76>
- Ford, R. A., Price, W. & Nicholson, I. I. (2016). Privacy and accountability in black-box medicine. *Mich.Telecomm.& Tech.L.Rev.*, 23, 1.
- Galetsis, P., Katsaliaki, K. & Kumar, S. (2020). Big data analytics in health sector: Theoretical framework, techniques and prospects. *International Journal of Information Management*, 50, 206-216. doi:10.1016/j.ijinfomgt.2019.05.003

- Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. doi:10.1016/j.ijinfomgt.2014.10.007
- Garg, A., Jack, B. & Zuckerman, B. (2013). Addressing the social determinants of health within the patient-centered medical home: Lessons from pediatrics. *Jama*, 309(19), 2001-2002.
- Gillan, C. M. & Whelan, R. (2017). What big data can do for treatment in psychiatry. *Current Opinion in Behavioral Sciences*, 18, 34-42. doi:10.1016/j.cobeha.2017.07.003
- Hamad, R., Modrek, S., Kubo, J., Goldstein, B. A. & Cullen, M. R. (2015). Using "big data" to capture overall health status: Properties and predictive value of a claims-based health risk score. *PLoS One*, 10(5), e0126054.
- Hamoudi, A. A. & Sachs, J. D. (1999). Economic consequences of health status: A review of the evidence. *CID Working Paper Series*.
- Haukkala, A., Vornanen, M., Halmesvaara, O., Konttinen, H., Perola, M., Kääriäinen, H., . . . Aktan-Collan, K. (2018). Suomalaisten geenitietämys ja suhtautuminen perimästä saatavaan terveystietoon.
- Hersh, W. R., Weiner, M. G., Embi, P. J., Logan, J. R., Payne, P. R., Bernstam, E. V., . . . Cimino, J. J. (2013). Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical Care*, 51(8 0 3), S30.
- Hoffman, S. & Podgurski, A. (2013). The use and misuse of biomedical data: Is bigger really better? *American Journal of Law & Medicine*, 39(4), 497-538.
- Karim, A., Abderrahim, B. & Hayat, K. (2018). Big healthcare data: preserving security and privacy. *Journal of Big Data*.
- Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 64, 122.
- Koppe, G., Meyer-Lindenberg, A. & Durstewitz, D. (2021). Deep learning for small and big data in psychiatry. *Neuropsychopharmacology (New York, N.Y.)*, 46(1), 176-190. doi:10.1038/s41386-020-0767-z
- Krumholz, H. M. (2014). Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system. *Health Affairs*, 33(7), 1163-1170.

- Kyberturvallisuuskeskus. (09.07.2020). Tietoturva. Haettu osoitteesta <https://www.kyberturvallisuuskeskus.fi/fi/toimintamme/saantely-ja-valvonta/tietoturva>
- Lee, H. & Sohn, I. (2016). *Fundamentals of big data : Network analysis for research and industry* (First edition). Chichester, UK: John Wiley & Sons.
- Lehtilä, O., Nyström, P., Ronikonmäki, N. & Sirviö, T. (2021). Tietoturvan ja tietosuojan parantaminen yhteiskunnan kriittisillä toimialoilla: Työryhmän loppuraportti. *Liikenne- ja viestintäministeriön julkaisuja*.
- Lehto, M. & Linné, J. (2017). Suomen kyberturvallisuuden nykytila, tavoitetila ja tarvittavat toimenpiteet tavoitetilan saavuttamiseksi. *Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisu*.
- Lehto, M., Pöyhönen, J. & Lehto, M. (2019). *Kyberturvallisuus sosiaali- ja terveydenhuollossa*. Value From Public Health Data With Cognitive Computing (VFH) ja Watson Health Cloud Finland (WHC) -hankkeiden (2016–2019) loppuraportti, Vol. 2. Jyväskylän yliopisto.
- Lerman, J. (2013). Big data and its exclusions. *Stan.L.Rev.Online*, 66, 55.
- Murdoch, T. B. & Detsky, A. S. (2013). The inevitable application of big data to health care. *Jama*, 309(13), 1351-1352.
- Neittaanmäki, P. & Kaasalainen, K. (2018). *SOTE-toimintojen tehostaminen IT:N avulla - kehittämispotentiaali ja toimenpideohjelma*. Informaatioteknologian tiedekunnan julkaisuja No. 51/2018. Jyväskylän yliopisto.
- Neittaanmäki, P., Lehto, M., Ruohonen, T., Kaasalainen, K. & Karla, T. (2019). *Suomen terveystiedon ja sen hyödyntäminen*. Value From Public Health Data With Cognitive Computing (VFH) ja Watson Health Cloud Finland (WHC) -hankkeiden (2016–2019) loppuraportti, Vol. 4. Jyväskylän yliopisto.
- Niinimäki, E., Pöyhönen, J., Äyrämö, S. & Neittaanmäki, P. (2017). *Omadata terveydenhuollon tietointensiivisessä rakenteessa*. Informaatioteknologian tiedekunnan julkaisuja, Jyväskylän Yliopisto.
- Parempaa terveyttä genomitiedon avulla. kansallinen genomistrategia. työryhmän ehdotus. (2015). *Sosiaali ja terveysministeriön julkaisuja*.
- Poikola, A., Kuikkaniemi, K., Kuittinen, O. & Honko, H. (2018). MyData – johdatus ihmiskeskeiseen henkilötiedon hyödyntämiseen. *Liikenne- ja viestintäministeriön julkaisuja*.
- Price, W. N. & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1), 37-43.

- Quinn, M. J. (2014). *Ethics for the information age*. Pearson.
- Raghupathi, W. & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, 2(1), 1-10.
- Rao, D., Gudivada, V. & Raghavan, V. (2015). Data quality issues in big data. 2015 *IEEE International Conference on Big Data*. doi:10.1109/BigData.2015.7364065
- Sedayao, J., Bhardwaj, R. & Gorade, N. (2014). Making big data, privacy, and anonymization work together in the enterprise: Experiences and issues. 2014 *IEEE International Congress on Big Data*. doi:10.1109/BigData.Congress.2014.92
- Senthilkumar, S. A., Rai, B. K., Meshram, A. A., Gunasekaran, A. & Chandrakumarmangalam, S. (2018). Big data in healthcare management: A review of literature. *American Journal of Theoretical and Applied Business*, 4(2), 57-69.
- Sosiaali- ja terveystieteiden ministeriö. (2016). Digitalisaatio terveyden ja hyvinvoinnin tukena - sosiaali ja terveystieteiden ministeriön digitalisaatiolinjaukset 2025. *Sosiaali- ja Terveystieteiden ministeriön Julkaisuja 2016*.
- Sukumar, S. R. (2015). Quality of big data in health care. *International Journal of Health Care Quality Assurance*, 28(6), 621. doi:10.1108/IJHCQA-07-2014-0080
- Suur-Uski, J. & Rydenfelt, H. (2020). Oikeus informaation yksityisyyteen terveydenhuollossa: Tapaus Vastaamo ja sen seuraukset. *Nuori Lääkäri*, 2020(5).
- Tenenbaum, J. D., Bhuvaneshwar, K., Gagliardi, J. P., Fultz Hollis, K., Jia, P., Ma, L., . . . Visweswaran, S. (2019). Translational bioinformatics in mental health: Open access data sources and computational biomarker discovery. *Briefings in Bioinformatics*, 20(3), 842-856.
- THL (2017). *Valtakunnallinen omatietovaranto ja hyvinvointitietojen hallinta*. Sosiaali- ja terveydenhuollon tietohallinnon operatiivisen ohjauksen yksikkö OPER.
- Varadhan, R., Segal, J. B., Boyd, C. M., Wu, A. W. & Weiss, C. O. (2013). A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *Journal of Clinical Epidemiology*, 66(8), 818-825.
- Voigt, K. (2019). Social justice, equality and primary care: (How) can 'Big Data' Help? *Philosophy & Technology*, 32(1), 57-68.

Vähäkainu, P. & Neittaanmäki, P. (2017). *Digitaalinen terveys ja älykäs terveydenhuollon teknologia*. Tekes-Hanke: Value from Public Health Data with Cognitive Computing. Informaatioteknologian tiedekunnan julkaisuja, Jyväskylän Yliopisto.

Walker, R. (2015). *From big data to big profits: Success with data and analytics*. Oxford University Press.