

## Potilastiedot hyötykäyttöön perusterveydenhuollossa – tarvitaan kysymisen taitoa, dataa ja tiedonlouhinnan osaamista

Ilona Rönkkö TtM<sup>1</sup>, Ulla-Mari Kinnunen TtT<sup>2</sup>, Sami Kiviharju FM<sup>1,3</sup>, Risto Mäkinen LL<sup>4</sup>

<sup>1</sup> Hämeenlinnan kaupunki, terveyspalvelut, <sup>2</sup> Itä-Suomen yliopisto, Sosiaali- ja terveysjohtamisen laitos, <sup>3</sup> Tampereen yliopisto, Informaatiotieteiden yksikkö, <sup>4</sup> Helsingin kaupunki

**Ilona Rönkkö, Hämeenlinnan kaupunki, terveyspalvelut, Viipurintie 1-3, 13200 Hämeenlinna, FINLAND.  
Sähköposti: [ilona.ronkko@hameenlinna.fi](mailto:ilona.ronkko@hameenlinna.fi).**

### Tiivistelmä

Suomessa kansansairauksien hoidon tuloksissa on parannettavaa. Tätä haastetta lisää perusterveydenhuollon heikko hoitopääsy, joka näkyy pitkinä jonoina terveyskeskuslääkärille. Tarvitaan kokonaisvaltaista toiminnallista uudistusta. Chronic Care Model (CCM) jalkautettiin Suomeen perusterveydenhuollon kehittämisen viitekehystehtäväksi yli 60 kunnassa vuosina 2010-2012. Mallin kehittäminen perustui havaintoihin, joissa monisairaant ja useita terveysriskejä omaavat asiakkaat eivät saaneet riittävän kokonaisvaltaista hoitoa, vaan usein hoito keskittyi yksittäisten akuuttien ongelmien ratkaisuun. Tässä artikkelissa kuvattu tutkimus toteutettiin perusterveydenhuollon organisaatiossa, jossa oli tehty CCM -viitekehystehtävään perustuvia päätöksiä siitä, miten potilastietoja kirjataan ja miten niitä halutaan hyödyntää.

Tutkimuksen tarkoituksena oli kuvata, millaisia sairaus- ja terveysriskien muodostamia yhdistelmiä ilmenee asiakasryhmässä, jonka jäsenillä oli useita sairauksia ja terveysriskejä yhtäaikaaisesti. Tutkimusaineistona käytettiin tämän asiakasryhmän (n=769) rakenteisia potilasrekisteritietoja retrospektiivisesti. Aineistoa tutkittiin tiedonlouhinnan menetelmillä. Tavoitteena oli selvittää, millaisia terveysongelmien ilmenemisiä asiakasryhmässä esiintyy ja toisaalta kuvata, miten potilastietoja louhimalla voidaan tuottaa tällaista tietoa.

Tulokset osoittavat, että asiakasryhmän jäsenillä esiintyi runsaasti useiden eri sairauksien ja terveysriskien yhdistelmiä ja että asiakasryhmässä näiden yhdistelmien variaatiot olivat suuria. Monisairastavuus ja eri sairauksien terveysriskien liittyminen toisiinsa asettaa haasteita asiakkaan hoidolle. Yksittäisten terveysongelmien sijaan hoidon pitäisi kohdentua asiakkaan kokonaistilanteeseen.

Potilastietojen tiedonlouhinta on menetelmä tiedon tuottamiseksi organisaation palvelujen kehittämisen ja johtamisen tueksi. Organisaatioissa kannattaa päättää, mitä asiakkaista halutaan tietää, toimeenpanna sen edellyttämät kirjaamiskäytännöt ja varmistaa oma tiedonhallinnan osaaminen. Se luo pohjan voimavarojen kohdentamiseen haluttuihin asiakasryhmiin ja -tarpeisiin.

**Avainsanat:** potilastietojärjestelmä, tiedonlouhinta, perusterveydenhuolto

## Abstract

The patients with multimorbidity and several health risks are often inadequately treated and health care is uncoordinated and often responds to acute problems only. The Chronic Care Model offers a good framework for identifying what kind of information is needed for the improvement of the care of the patients with several health disorders. The data in the electronic health record is a valuable source of information for example of the health status of the population.

The purpose of this study was to use data mining and statistical methods to describe what kind of combinations of different chronic conditions and health risks can be found in the selected patient group on the basis of the EHR data. The research material was collected from the structured patient records of one chronic patients group at a single health care centre. The aim of this study was to increase information on the utilization possibilities of the data stored in EHR.

There is large variation in combinations of diseases and health risks in the selected patient group. Multimorbidity and health risk combinations are a challenge for the care of these patients: the focus should be on the patient's overall situation instead of separate health problems.

Data mining provides valuable information for the development of the organization's services and management. Organizations should decide what kind of information they need about the patients and then ensure that the data is recorded and that the organization has its own data management expertise.

**Keywords:** Data Mining, Primary Health Care, Electronic Health Records (MeSH)

## Johdanto

Suomalainen perusterveydenhuolto kamppailee hoidon saatavuuden, tuottavuuden ja vaikuttavuuden kanssa. Nämä haasteet ilmenevät esimerkiksi perusterveydenhuollon hoitopäässä [1], kansansairauksien hoitotuloksissa [2] sekä julkisen talouden kestävyysvajeessa. Valmisteilla olevan sosiaali- ja terveyspalveluiden (sote)-uudistuksen tavoitteena on kaventaa ihmisten hyvinvointi- ja terveyseroja sekä hallita kustannuksia [3]. Sote-tieto hyötykäyttöön -strategian 2020 yhtenä tavoitteena on saattaa tietoaineistot tukemaan reaaliaikaisesti palvelutuotannon johtamista ja yhteiskunnallista päätöksentekoa sen asukkaiden hyväksi. Toimenpiteinä mainitaan muun muassa sosiaali- ja terveyspalveluita kuvaavien luokitusten ja mittareiden yhdenmukaistaminen, väestötason tilasto- ja indikaattoripalveluiden kehittäminen, tiedon toissijaisen hyödyntämisen resursointi sekä siihen yhteisen infrastruktuurin luominen [4].

Terveystieteiden tutkimuskeskuksessa (1326/2010) [5] määritellään, että kunnan tulee *"seurata asukkaittensa terveyttä ja hyvinvointia sekä niihin vaikuttavia tekijöitä väestöryhmittäin sekä kunnan palveluissa toteutettuja toimenpiteitä, joilla vastataan kuntalaisten hyvinvointitarpeisiin"*. Kunnille on myös annettu tehtäväksi määrittellä ne hyvinvointi- ja terveysosoittimensa, jotka luovat perustan terveyden edistämisen tavoitteille ja toimenpiteille.

Chronic Care Model (CCM) tarjoaa kiinnostavan viitekehysten perusterveydenhuollon kehittämiseksi. Wagnerin ja Bodenheimerin kehittämä malli perustuu havaintoihin, joiden mukaan monisairaiden tai korkeassa terveysriskissä olevien henkilöiden hoito ei tuota haluttuja tuloksia. Taustalla on paine keskittyä akuutteihin ja yksittäisiin terveysongelmiin [6,7]. Malli jalkautui vuosien 2010–2012 aikana Suomeen erityisesti Sosiaali- ja terveysministeriön (STM) Kaste-rahoituksella tuettujen Potku 1 ja 2-hankkeiden avulla. Tuolloin yli 60 kuntaa oli mukana hankkeissa [8]. CCM -viitekehysten kulmakivet

perustuvat potilaan keskeiseen rooliin toimijana (omahoidon tuki), monipuoliseen palveluvalikoimaan, päätöksentuen ja potilastietojärjestelmän hyödyntämiseen, muiden yhteisöjen (kuten potilasjärjestöjen) kanssa tehtävään yhteistyöhön sekä johdon sitoutumiseen [9,10]. Mallin kulmakivet [11] haastavat terveydenhuollon organisaatioita kehittämään potilastietojärjestelmien hyödyntämistä, erilaisten asiakasryhmien ja palvelutarpeiden tunnistamista, palveluvalikoimaa sekä hoidon vaikuttavuuden arviointia.

CCM tarjoaa hoitomallin monisairaille ja terveysriskeissä oleville. Näiden asiakkaiden tunnistamiseksi tarvitaan indikaattoreita. Kansainvälisten tutkimusten perusteella monisairastuvuutta kuvaavia indikaattoreita löytyy jonkin verran. Esimerkiksi de Groot ym. [12] ovat artikkelikatsauksessaan arvioineet kolmeatoista erilaista monisairastuvuutta kuvaavaa mittaria. Osa mittareista käytti potilastietojärjestelmän tietoa, kuten diagnoositietoja ja laboratoriotuloksia. Osa perustui kliiniseen tutkimuksen tai haastattelun kautta saatuun tietoon. Yksi mittareista, "Disease Count", hyödynsi yhdistelmäinä potilaskertomusta, ICD-koodeja ja haastattelua ja perustui yksinkertaisesti sairauksien lukumäärän laskentaan. Tanskassa Roque ym. [13] louhivat kymmenen vuoden ajalta psykiatristen potilaiden (n=5543) narratiiviset potilaskertomustekstit ja yhdistivät tekstistä louhitut ICD10-koodit rakenteisesti kirjattujen ICD10-koodien kanssa. Yhdistettyjen koodien pohjalta he tarkastelivat diagnoosiparien komorbideettiä sekä ryhmittelivät potilaita diagnoosien samankaltaisuuden perusteella.

Potilastietojärjestelmät sisältävät runsaan asiakkaan terveydentilaa ja hoitoa koskevan kiinnostavan tietovarannon. Potilastietojärjestelmään tuotetulla tiedolla on todettu olevan yhteys potilaan hoidon tuloksiin [14]. Tiedonlouhinta tarjoaa menetelmiä olennaisen tiedon etsimiseen suurista tietomääristä [15]. Menetelmä sisältää tietojen valinnan, käsittelyn, yhdistämisen, analysoinnin ja raportoinnin vaiheita osana prosessiaan [16]. Tiedonlouhinta liittyy läheisesti tilastotieteeseen, tiedonhakuun, koneoppimiseen ja mallin tunnistukseen. Se käyttää hyväkseen muun muassa luokittelua ja ryhmittelyä [17]. Potilastietojärjestelmätietojen tiedonlouhinta on lisääntynyt vasta viime vuosien aikana [18].

Sen avulla voidaan tutkia muun muassa laajoja kroonisista sairauksista tallennettuja tietosisältöjä [19]. Trendiä kuvaa myös se, että marraskuussa 2015 tehdyn Scopus -tietokantahaun perusteella hakusanoilla "Data mining" and "EHR" löytyi vuosien 2006–2015 aikana yhteensä 151 artikkelia ja näistä artikkeleista 91 on tehty vuoden 2013 jälkeen [20].

Jos potilastietojärjestelmän tietoa halutaan hyödyntää organisaation palvelujen kehittämisessä, on varmistettava, että se sisältää tarvittavaa tietoa. Terveyden ja hyvinvoinnin laitos (THL) ohjaa kansallisesti potilastietojen kirjaamista koodistojen ja oppaiden avulla. Niissä ohjataan kirjaamaan muun muassa kansansairauksien kannalta tärkeitä rakenteisia tietosisältöjä, kuten elintapoihin liittyviä tupakointitilastietoja ja sairastuvuuteen liittyviä diagnoosikoodeja [21]. Monisairastuvuutta ei sellaisenaan ole ohjeissa eikä koodeissa huomioitu ja ne puuttuvat myös THL:n tilasto- ja indikaattoripankista [22]. Kuitenkin potilastiedon toisiokäytöllä on todettu olevan hyötyä potilaan hoitamisen lisäksi myös toiminnan ohjauksessa, raportoinnissa ja tilastoinnissa sekä päätöksenteontuessa [23].

Hämeenlinnan terveystalot osallistui Potku 1 ja 2 -hankkeisiin vuosina 2010–2014. Hankkeen lähtötilanteessa Hämeenlinnan terveysasemille oli pitkät jonot, hoitoon pääsyä pidettiin vaikeana ja hoidon tuloksista ei ollut tietoa [24,25]. Hankkeessa juurrutettiin CCM-mallia kokonaisvaltaisesti ja puhuttiin systeemitason muutoksesta [24]. CCM-mallista kehitettiin Kanavamalli-sovellus, jossa suuren terveysaseman (väestöpohja noin 27 000) asiakkaat jaettiin kahteen ryhmään: yksittäisessä ja päivystysluontoisessa hoidon tarpeessa olevat asiakkaat kuuluivat episodikanavaan ja pitkäjänteisempää tukea tarvitsevat asiakkaat terveyshyötykanavaan. Palvelut haluttiin tuottaa siten, että tilapäistä hoitoa tarvitsevat asiakkaat saavat palvelunsa joustavasti, laadukkaasti ja turvallisesti. Korkeassa terveysriskissä olevien asiakkaiden ja pitkäaikaissairaiden palvelut ovat kokonaisvaltaisia, ennakoituja ja suunnitelmallisia. Lisäksi hoidon tuottavuuden ja vaikuttavuuden haluttiin paranevan. Tavoitteina oli myös asiakkaiden tyytyväisyys sekä henkilöstön työiihtyvyys [24,25]. Tavoitteiden määrittämisen jälkeen selvitettiin, mitä asiakastason ongelmia oltiin ratkaisemassa (tarpeen

määrittäminen) ja mitkä olisivat vaikuttavia interventi-  
oita (keinoja).

Tavoitteiden saavuttamiseksi ja arvioimiseksi tarvittiin osoittimia eli indikaattoreita. Hämeenlinnassa haluttiin hyödyntää erityisesti potilastietojärjestelmää hoitotulosten systemaattisessa seurannassa ja johtamisen tukemisessa. Potilastietojärjestelmän tiedon avulla haluttiin tunnistaa asiakkaiden kansansairaudet ja terveysriskit. Tarve kirjaamiskäytäntöjen muutokseen todettiin hankkeen alkuvaiheessa vuonna 2010, jolloin havaittiin, että terveysriskien dokumentointi potilastietojärjestelmään oli erittäin hajanaista ja suurelta osin narratiivisessa muodossa. Tuolloin päätettiin, mitkä indikaattorit (esimerkiksi verenpaine, painoindeksi ja tupakointistatus) ja pysyväisdiagnoosit kirjataan systemaattisesti ja rakenteisesti kaikista asiakkaista. Alettiin puhua strategisesta kirjaamisesta [26,27]. Vuonna 2012 organisaatioon palkattiin tiedonlouhinnan osaajia. Määrätietoisella johdon ja tiedonhallinnan ammattilaisten yhteistyöllä on nyt vuosia myöhemmin päästy arvioimaan toimintaa, tuloksia ja asettamaan uusia tavoitteita [24]. Tässä artikkelissa kuvataan, miten organisaation systemaattisen kehittämistyön tuloksena saatua strategisesti merkittävää rakenteista tietoa on voitu tiedonlouhinnan keinoin käyttää esimerkiksi asiakkaiden palvelutarpeen tunnistamisessa ja palvelujen kehittämisessä. Artikkelissa on käytetty Rönkön [28] pro gradu-tutkielman aineistoa, mutta tuloksia on analysoitu ja tarkasteltu eri tutkimuskysymysten kannalta. Tässä artikkelissa esitetyt tulokset ovat esimerkkejä analyysimahdollisuuksista.

### Tutkimuksen tarkoitus ja tavoite

Tutkimuksen tarkoituksena oli kuvata tietojärjestelmään rakenteisesti kirjattua potilastietoa hyödyntäen, millaisia sairaus- ja terveysriskien muodostamia yhdistelmiä ja korrelaatioita ilmenee asiakasryhmässä, jonka jäsenillä oli useita sairauksia ja terveysriskejä yhtäaika-

sesti. Tavoitteena oli selvittää asiakasryhmän terveysongelmien ilmentymistä ja mahdollistaa sen perusteella organisaation palvelutoiminnan kehittäminen ja kohden-

taminen. Toisaalta tavoitteena on myös kuvata, miten potilastietoja louhimalla voidaan tuottaa tietoa asiakkaista. Tiedonlouhinnan prosessin, menetelmien ja tulosten kuvaus voi tarjota myös muille organisaatioille keinoja arvioida mahdollisuuksiaan kerätä ja analysoida potilastietojärjestelmästä tietoa oman toiminnan kehittämiseksi.

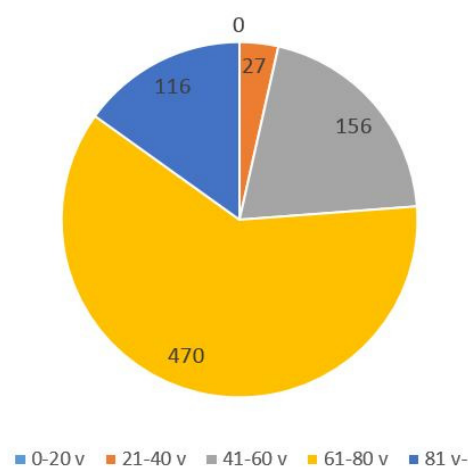
Tutkimuskysymysten avulla etsitään vastauksia asiakasryhmän sairauksien ja terveysriskien ilmentymiin:

1. Millaisia sairaus- ja terveysriskiyhdistelmiä asiakasryhmässä esiintyy?
2. Miten asiakasryhmän eri terveysriskit ja sairaudet liittyvät toisiinsa?

### Aineisto ja menetelmät

Tutkimusaineistona käytettiin yhden terveysaseman terveyshyötyasiakasryhmän jäsenten (n=769) rakenteisia potilasasiakirjamerkintöjä. Asiakkaat kirjataan potilastietojärjestelmässä asiakasryhmään, joten jäsenet saatiin poimittua ryhmämerkinnän mukaisesti. Tähän asiakasryhmään ohjataan pitkäaikaissairaat, paljon palveluita käyttävät ja erilaiset riskiryhmät [24]. Valintakriteereistä johtuen asiakasryhmän keski-ikä on 69 vuotta. Ikäjakauma esitetään kuviossa 1.

Terveyshyötysasiakkaat (n=769)



Kuvio 1. Terveyshyötyasiakkaiden ikäjakauma.

Aineisto poimittiin potilastietojärjestelmän raportointitietokannasta SQL-kyselyin, joiden avulla saatiin tietoja taulukoissa 1 ja 2 esitettyjen tietosisältöjen ja rajausten mukaan. Asiakasryhmä tunnistettiin potilastietojärjestelmän asiakasryhmähallinnan kautta.

Terveyshyötyasiakkaiden sairaustiedot haettiin taulukossa 1 esitettyjen ICD10 ja ICPC2 -koodien perusteella ilman aikarajausta. Koodien valinnan ja sairausryhmitteilyn teki tutkimuskohteena olevan organisaation lääketieteellinen johto sen mukaisesti, minkä se katsoi olevan asiakasryhmän osalta merkittävää tietoa.

Terveysriskien (taulukko 2) osalta tietoja haettiin vain kahden vuoden ajalta, jotta tulosten voitiin arvioida olevan riittävän tuoreita. Laboratoriotuloksista haettiin tiedot kolesterolitasoon liittyvästä LDL-arvosta sekä pitkäaikaisverensokerista HbA1C. Terveysriski-indikaattoreista valittiin ne, jotka oli asiakasryhmälle kirjattu vähintään 60 prosentin kattavuudella [25]. Nämä indikaattorit olivat BMI (painoindeksi), tupakointistatus sekä verenpaine (tarkasteluun systolinen paine).

**Taulukko 1.** Tutkimusaineiston sisällön kuvaus valittujen sairauksien osalta.

MUUTTUJAT	Tarkennin	Poiminat	Tulkinta	TIEDONLÄHDE JA AIKARJAUS
Diagnoositiedot (ICD-10 ja käyntisytyt ICPC -2 seuraavassa sarakkeessa esitettyjen koodien perusteella)	Luokitus	Koodi	Ryhmitys	SQL-kysely tietokannasta pitkäaikais-/kansansairauksien ICD10 ja ICPC2 -koodien mukaisesti ilman aikarajausta
	ICD-10	E10-14	DM = Diabetes	
	ICPC-2	T89-T90		
	ICD-10	I10-I15	RR = Verenpainetauti	
	ICPC-2	T85-T87		
	ICD-10	J43- J44	COPD = Keuhkohtaumatauti	
	ICPC-2	R95		
	ICD-10	I20-I25	MCC = Sepelvaltimotauti	
	ICPC-2	K74-K76		
	ICD-10	I42 ja I50	CHF = Sydämen vaajatoiminta	
	ICPC-2	K77		
	ICD-10	F30-F31	PSYK = Psykkinen sairaus	
	ICPC-2	P71-P73, P98		
	ICD-10	M05-M06	REUMA	
	ICPC-2	L88		
	ICD-10	F32-F38	DEPR = Depressio	
	ICPC-2	P76		
ICD-10	J45-J46	ASTMA		
ICPC-2	R96			
ICD-10	K05.30-K05.31	PARO = Parodontiitti eli hampaan kiinnityskudossairaus		
ICD-10	K02.1-K02.2, K02.8-K02.9	KARIES = Hampaan reikiytyminen		

**Taulukko 2.** Tutkimusaineiston sisällön kuvaus terveysriskien osalta.

MUUTTUJAT	Tarkennin	Poiminnot	Tulkinta	TIEDONLÄHDE JA AIKARJAUS
Laboratoriotuloksista pitkäaikaista verensokeritasoa kuvaava sokerihemoglobiini HbA1C ja "low density lipoprotein" LDL	Laboratorioindikaattori ja tiedon kattavuus asiakasryhmässä (n = 769)	Riskiraja	Riskimerkitys	SQL-kysely kahdelta edeltävältä vuodelta (vuodet 2013 ja 2014) - riskirajan ylittäneet tulokset
	HbA1C (tietojen kattavuus 90 %)	> 53 mmol/mol	Diabetekseen liittyvät elinriskit (Terveyskirjasto)	
	LDL (tietojen kattavuus 93 %)	> 3 mmol/l	Valtimokovettumatautiiriski (Terveyskirjasto)	
Keskeiset terveysriski-indikaattorit, jotka on asiakasryhmässä kirjattu yli 60 % kattavuudella kahden vuoden merkintöjen osalta	Indikaattori ja kirjaamiskattavuus asiakasryhmässä (n = 769)	Riskikirjaus	Riskimerkitys	SQL-kysely mittareittain kahdelta edeltävältä vuodelta (vuodet 2013 ja 2014) -riskirajan ylittäneet tai riskikirjaus (tupakointi +)
	TUPAK -mittari tupakointistatus, (kirjaamiskattavuus 63 %)	"+" = tupakoi	Tupakointi on monien sairauksien riskitekijä ja pahentaa jo todettuja sairauksia. (Terveyskirjasto)	
	RR Syst systolinen verenpaine (kirjaamiskattavuus 98 %)	> 140 mmHg	Kohonnut verenpaine vahingoittaa valtimoita ja aiheuttaa aivohalvauksia ja sydäninfarkteja. (Terveyskirjasto)	
	BMI painoindeksi (kirjaamiskattavuus 87%)	> 30 = merkittävä ylipaino	Mm. aikuistyyppin diabetesriski, sydäntautien ja aivohalvausten vaaraa. (Terveyskirjasto)	

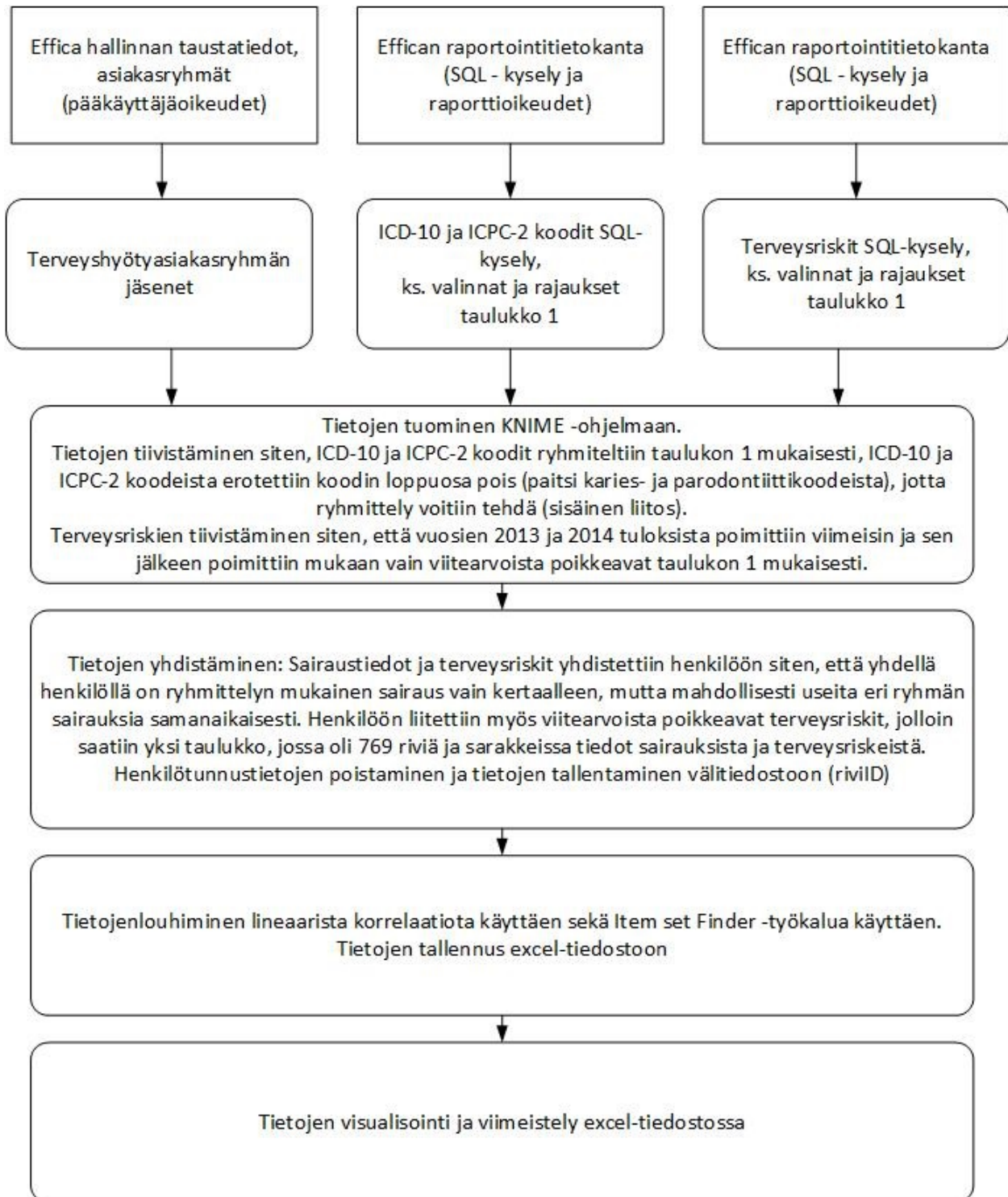
Tiedonlouhintaa varten on olemassa useita eri työkaluja ja menetelmiä. Tässä työssä tiedot haettiin SQL-kyselyjen avulla ja tuotiin käsittelyä ja analysointia varten KNIME-ohjelmaan (Konstanz Information Miner) [29]. KNIME-ohjelmassa tiedot yhdistettiin, suodatettiin, tiivistettiin, analysoitiin ja lopuksi tallennettiin anonyyminä tietona Excel-taulukkoon tietojen lopullista visualisointia ja raportointia varten.

Erilaisia diagnoosi- ja riskitietojen yhdistelmiä ja niiden yleisyyttä tutkittiin KNIME-ohjelman Item Set Finder-solmulla. Solmu sisältää useita algoritmeja, joiden avulla aineistosta voidaan etsiä erikokoisia muuttujien yhdistelmiä ja laskea niiden esiintyvyydet sekä määrinä että prosentteina suhteessa aineistossa olevaan henkilömäärään [30]. Menetelmän avulla tarkasteltiin, millaisia yhdistelmiä muodostuu, kun yhteisten muuttujien määrää vaihdellaan (esimerkit kuviossa 3 ja 4). Käytetyt muuttujat on kuvattu taulukoissa 1 ja 2, niiden osalta oltiin kiinnostuneita, oliko muuttuja henkilöllä: kyllä tai ei. Muuttujia saattoivat olla esimerkiksi DM=kyllä, RR=e, korkea verenpaine=kyllä tai korkea\_HbA1c=e. Yhteisten muuttujien lisäksi asiakkaalla voi olla myös muita diagnooseja ja riskejä, joten siksi kukin yhdistelmä tulee suhteuttaa koko asiakasryhmään (n=769), eikä

verrata toisiinsa. Muuttujayhdistelmille voidaan laskea niiden esiintyvyys aineistossa, josta käytetään nimitystä tuki (support). Tuki kertoo kuinka monelta henkilöltä kyseinen muuttujayhdistelmä löytyy. Tässä analyysissä minimiueksi, alla `RelativeItemSetSupport%`, määriteltiin 5 %. Tämä raja on syytä huomioida tuloksia tarkasteltaessa. Tuloksia tarkasteltaessa on huomioitava myös, että sama asiakas voi olla samanaikaisesti eri muuttuja-/muuttujayhdistelmäriveillä, jos hänellä on yhtäaikaista myös muita kuin rivillä mainittuja sairauksia ja riskejä. Muuttujayhdistelmät eivät siis ole toisiaan poissulkevia, vaan potentiaalisesti päällekkäisiä.

Aineistosta analysoitiin myös, miten eri sairaus- ja riskitiedot liittyivät toisiinsa. Menetelmänä käytettiin KNIME-työkalun Linear Correlation-solmua [31]. Solmu laskee korrelaation ristiintaulukoinnin perusteella Pearsonin  $\chi^2$ -testiä käyttämällä, jonka jälkeen korrelaation aste normalisoidaan välille [0-1]. Arvo 1 kuvaa vahvaa korrelaatiota.

Kuviossa 2. on esitetty tiedonlouhinnan prosessi SQL-kyselyistä lopputuloksiin tiivistetyssä muodossa.



Kuvio 2. Tiedonlouhinnan prosessi tiivistetysti.

## Tulokset

### *Erilaisten diagnoosi- ja terveysriskien ja niiden yhdistelmien esiintyvyydet*

Kuviossa 3 esitetään yli 5 %:lla asiakasryhmästä esiintyvät yksittäiset diagnoosit ja terveysriskit, jolloin mukaan tulevat kaikki taulukoissa 1 ja 2 esitetyt muuttujat. Niistä viisi yleisintä ovat verenpainetauti, diabetes, ylipaino, kohonnut verenpaine ja kohonnut kolesteroli, joiden esiintyvyys on yli 25 % asiakasryhmässä. Verenpainetauti on kirjattu lähes 80 %:lle asiakkaita.

Moniriskisyyden tunnistamiseksi analysoitiin, mikä on suurin muuttujayhdistelmämäärä, joka löytyy aineistosta määritellyllä 5 %:n minimituella. Tulos osoitti, että

maksimissaan näillä ehdoilla tästä aineistosta löytyi neljän muuttujan pituinen muuttujayhdistelmä kuudessa eri ilmenemismuodossa. Yleisin niistä on yhdistelmä, jossa mukana oli kohonnut verenpaine, ylipaino, diabetes ja verenpainetauti. Tämä yhdistelmä esiintyi lähes 10 %:lla terveyshyötyasiakkaista (n=769). Myös tässä tuloksessa on huomioitava, että samaa asiakasta voi koskea useampi eri yhdistelmä, jos hänellä oli enemmän kuin neljä sairautta tai riskiä.

Kolmen eri muuttujan yhdistelmiä (kuvio 4) löytyy jo 30 erilaista 5 % minimituella. Yleisin yhdistelmä on ylipaino, diabetes ja verenpainetauti, joka esiintyy lähes neljäsosalla asiakasryhmästä (n=769).

Yhdistelmän koko	Yhdistelmien määrä asiakasryhmässä	Yhdistelmän esiintyvyyden prosenttiosuus asiakasryhmässä (määrä*100/769 (n))	Yhdistelmän muuttuja 1
1	609	79,19	Verenpainetauti
1	429	55,79	Diabetes
1	379	49,28	Ylipaino
1	275	35,76	Kohonnut verenpaine
1	212	27,57	LDL_riski
1	172	22,37	Karies
1	172	22,37	Astma
1	155	20,16	Sepelvaltimotauti
1	124	16,12	Depressio
1	112	14,56	Kohonnut verensokeri
1	109	14,17	Tupakointi
1	68	8,84	Sydämen vajaatoiminta
1	65	8,45	Reuma
1	58	7,54	Keuhkohtaumatauti

**Kuvio 3.** Yhden sairauden tai terveysriskimuuttujan esiintyvyys asiakasryhmässä (n=769).



Yhdistelmän koko	Yhdistelmien määrä asiakasryhmässä	Yhdistelmän esiintyvyyden prosentiosuus asiakasryhmässä (määrä*100/769 (n))	Yhdistelmän muuttuja 1	Yhdistelmän muuttuja 2	Yhdistelmän muuttuja 3
3	199	25,88	Ylipaino	Diabetes	Verenpainetauti
3	134	17,43	Kohonnut verenpaine	Diabetes	Verenpainetauti
3	106	13,78	Kohonnut verenpaine	Ylipaino	Verenpainetauti
3	90	11,70	Kohonnut verensokeri	Verenpainetauti	Diabetes
3	90	11,70	Kohonnut verenpaine	Ylipaino	Diabetes
3	79	10,27	Karies	Diabetes	Verenpainetauti
3	75	9,75	LDL_riski	Ylipaino	Verenpainetauti
3	75	9,75	LDL_riski	Diabetes	Verenpainetauti
3	72	9,36	Kohonnut verensokeri	Ylipaino	Diabetes
3	70	9,10	LDL_riski	Kohonnut verenpaine	Verenpainetauti
3	69	8,97	Astma	Ylipaino	Verenpainetauti
3	66	8,58	Sepelvaltimotauti	Diabetes	Verenpainetauti
3	65	8,45	Astma	Diabetes	Verenpainetauti
3	64	8,32	Karies	Ylipaino	Verenpainetauti
3	62	8,06	Karies	Ylipaino	Diabetes
3	56	7,28	Kohonnut verensokeri	Ylipaino	Verenpainetauti
3	52	6,76	Sepelvaltimotauti	Kohonnut verenpaine	Verenpainetauti
3	52	6,76	Sepelvaltimotauti	Ylipaino	Verenpainetauti
3	51	6,63	Depressio	Ylipaino	Verenpainetauti
3	51	6,63	Depressio	Diabetes	Verenpainetauti
3	49	6,37	Kohonnut verensokeri	Kohonnut verenpaine	Diabetes
3	49	6,37	LDL_riski	Ylipaino	Diabetes
3	48	6,24	Astma	Kohonnut verenpaine	Verenpainetauti
3	47	6,11	Astma	Ylipaino	Diabetes
3	45	5,85	Tupakointi	Diabetes	Verenpainetauti
3	45	5,85	Karies	Kohonnut verenpaine	Verenpainetauti
3	41	5,33	Kohonnut verensokeri	Kohonnut verenpaine	Verenpainetauti
3	41	5,33	Sepelvaltimotauti	Ylipaino	Diabetes
3	41	5,33	Astma	LDL_riski	Verenpainetauti
3	39	5,07	Depressio	LDL_riski	Verenpainetauti

**Kuvio 4.** Kolmen sairaus- tai terveystilayhdistelmän esiintyvyys asiakasryhmässä (n=769).

Muuttuja	Muuttujan riippuvuus suhteessa depressioniin
Verenpainetauti	0,04
Keuhkohtaumatauti	0,01
Diabetes	0,05
Astma	0,09
Depressio	1,00
Sydämen_vajaatoiminta	0,05
Sepelvaltimotauti	0,11
Psykoottinen_sairaus	0,06
Karies	0,02
Parodontiitti	0,04
Reuma	0,01
Ylipaino	0,07
Tupakointi	0,13
LDL-riski	0,13
Kohonnut_verenpaine	0,02
Kohonnut_verensokeri	0,07

**Kuvio 5.** Keskinäiset riippuvuussuhteet depressioniin potilastietojärjestelmään kirjattujen tietojen perusteella terveyshyötyasiakasryhmässä (n=769).

Muuttuja	Muuttujan riippuvuus suhteessa tupakointiin
Verenpainetauti	0,08
Keuhkohtaumatauti	0,28
Diabetes	0,01
Astma	0,07
Depressio	0,13
Sydämen_vajaatoiminta	0,00
Sepelvaltimotauti	0,07
Psykoottinen_sairaus	0,02
Karies	0,01
Parodontiitti	0,10
Reuma	0,02
Ylipaino	0,02
Tupakointi	1,00
LDL-riski	0,04
Kohonnut_verenpaine	0,02
Kohonnut_verensokeri	0,03

**Kuvio 6.** Keskinäiset riippuvuussuhteet tupakointiin potilastietojärjestelmään kirjattujen tietojen perusteella terveyshyötyasiakasryhmässä (n=769).

### ***Pitkäaikaissairauksien ja terveysriskien keskinäiset riippuvuudet***

Linearisesta korrelaatiotaulukosta (Liite 1) on esiin nostettu kaksi esimerkkiä, toinen sairaus- ja toinen riskilähtöisenä. Kuviossa 5 on tarkasteltu depression korrelaatiota suhteessa muihin tarkastelussa oleviin sairauksiin ja terveysriskeihin. Korrelaatiot ovat heikkoja, mutta tulokset ovat samansuuntaisia kuin muuttujien yhdistelmiä kuvaavassa kuviossa 4. Eniten riippuvuutta ilmenee tupakoinnin, LDL-riskin, sepelvaltimotaudin ja astman kanssa.

Kuviossa 6 on puolestaan tarkasteltu tupakoinnin riippuvuutta suhteessa muihin sairauksiin ja terveysriskeihin. Suurimmat riippuvuudet löytyivät keuhkohtaumatautiin, depression sekä parodontiittiin.

Muita riippuvuuksia tarkasteltaessa (Liite 1) voidaan todeta, että erityisesti diabeteksella ja depressiolla on korrelaatioita useampaan muuhun muuttujaan.

### **Pohdinta**

#### ***Eettisyys ja luotettavuus***

Potilastietojärjestelmien sisältämä tiedon käsittely määritellään tarkasti lainsäädännössä [32]. Yksilön tietosuoja tulee turvata kaikissa tutkimuksen vaiheissa.

Tietojen tietoturvalisesta käsittelystä tehtiin tarkka suunnitelma ja organisaatiolta haettiin tutkimuslupaa sekä laadittiin rekisteriseloste aineiston käytöstä. Henkilötunnusellista tietoa käytettiin vain tietojen yhdistämiseen, jonka jälkeen ne korvattiin yksilöllisellä rivitunnusella. Sen jälkeen analyysi ja raportointi toteutettiin vain anonyymiä tietoa käsitellen. Yksilön tietosuojasta huolehdittiin koko prosessin ajan.

Kvantitatiivisessa tutkimuksessa on tärkeää arvioida sen validiteettia eli pätevyyttä. Validiteetti tarkoittaa karkeasti sitä, kuinka hyvin tutkimustulokset pystyvät kattamaan tutkimuskysymykset [33,34]. Tässä artikkelissa tutkimuskysymykset koskivat valitun asiakasryhmän potilastietoihin kirjattuja sairaus- ja terveysriskitietoja. Tässä mielessä taulukoiden 1 ja 2 mukainen aineisto vastasi tutkimuskysymyksiin. Validiteettia lisäsi se, että analyysiin poimitut muuttujat olivat tarkasti määriteltäviä ja yksiselitteisiä tilastotietoja. Tulosten tarkastelussa on tärkeää huomioida aineiston ulkopuolella rajattujen tietojen merkitys. Jos kirjaamiskattavuus olisi ollut riittävä kaikkien riskitietojen osalta, niin asiakkaiden samanaikaisten terveysongelmien määrä olisi näyttänyt todellisempana ja oletettavasti kumuloituneempana. Myös aineiston ulkopuolelle jääneiden terveysriskien, esimerkiksi alkoholiriippuvuutta mittaavan AUDITin osalta, on huomioitava, että niiden osalta yhteydet muihin riskeihin ja sairauksien jäävät tässä tutkimuksessa avoimeksi. Jos potilastietojärjestelmään

syötetty tieto on puutteellista, väärää tai väärässä paikassa, vaikeuttaa se muun muassa potilaan hoitamista ja tiedon toisiokäyttöä [4,23.] Tämän tutkimuksen osalta se vääristää tutkimustuloksia.

Reliabiliteetti kuvaa luotettavuutta ja sen avulla voidaan arvioida tutkimuksen toistettavuutta [33,34]. Tutkimusaineisto kerättiin potilasrekisteritiedoista, joiden kirjaamistavan ja järjestelmän tutkija tunsi hyvin. Tämä lisää osaltaan tulosten luotettavuutta. Terveydenhuollon ammattilaiset kirjaavat tiedot potilastietojärjestelmään rakenteisesti yhtenäisesti sovitulla mittarimerkinällä. Ammatilainen saattaa kirjata tiedon väärin järjestelmään, jolloin se voi vaikuttaa lopputuloksiin. Tiedon louhinnan menetelmillä poistettiin aineistosta selvästi vääriksi merkinnöiksi tunnistetut tiedot (esimerkiksi painoindeksilukema 250). Nämä mahdolliset virheet tietojen tallennuksessa on syytä ottaa huomioon lopputuloksia analysoitaessa. Tutkimuksen vaiheet on mahdollista toistaa, sillä tietojen keruuseen liittyvät SQL-kyselyt ja tietojen jatkokäsittely on toteutettu pääosin KNIME-ohjelmistoa hyödyntäen. Tietojen käsittely KNIME:llä tapahtuu muodostamalla tietovirtakaavio, joka määrittää mistä käsiteltävä tieto tulee ja miten sen vaiheittainen käsittely etenee. Tietojen käsittely voidaan toistaa tietovirtakaavion mukaisesti.

### ***Tutkimustulosten pohdintaa***

Tutkimuksen tarkoitus oli kuvata tietojärjestelmään rakenteisesti kirjattujen potilastietojen avulla millaisia sairaus- ja terveysriskien muodostamia yhdistelmiä ja korrelaatioita ilmenee asiakasryhmässä, jonka jäsenillä oli useita sairauksia ja terveysriskejä yhtäaikaaisesti. Tutkimustulosten ohella tärkeänä tavoitteena oli myös kuvata, miten potilastietoa louhimalla voidaan analysoida asiakkaiden sairastamiseen ja terveysriskeihin liittyvää tietoa.

Artikkeli kuvaa, miten perusterveydenhuollon organisaatio, joka on valinnut uudistumisen viitekehikseen CCM-mallin, hyödyntää tiedonlouhinnan menetelmiä ja potilastietojärjestelmän tietoja asiakasryhmän kansansairauksien ja terveysriskien tunnistamiseksi [25,26]. Tutkimuksen kohteena olevan organisaation johto päät-

ti, että tavoitteena on kansansairauksien hoidon tulosten paraneminen ja valitsi ne strategiset indikaattorit, jotka asiakkaista kirjataan potilastietojärjestelmään rakenteisesti sekä uudisti toimintamallejaan.

Tutkimustulosten mukaan kerätty terveystieto korkean riskin asiakkaista osoitti valituilla ja rajatuilla diagnooseilla ja terveysriskeillä niiden yhdistelmien laajan kirjon. Huomioiden aineistosta pois rajatut sairaudet ja terveysriskit, syntyy monisairaista hyvin heterogeeninen kuva. Monisairastavuuteen liittyen tunnistettiin erilaisia riskien yhdistelmiä, kuten depression osalta riippuvuudet moneen muuhun terveysriskiin tai sairauteen. Korrelaatiot ovat tosin heikkoja, alle 0.14 kuviossa 5, kun vahva olisi lähempänä 1.0.aa. Tulokset auttavat kuitenkin havainnoimaan sitä, että asiakkaan hoidossa ei kannata keskittyä vain yhteen ongelmaan. Tavoitteena on, että keskenään yhdessä esiintyvät terveysriskit tulevat huomioituksi ja ratkaistuiksi. Tutkimustulokset haastavat perinteisen hoitomallin, joka perustuu yksittäisten terveysongelmien hoitosuositukseen, sekä sairauskohtaisesti erikoistuneisiin ammattilaisiin, kuten diabetes-, depressio-, astma-, sydän- ja päihdehoitajiin. Monisairastavuuden hallinta ja hoito ohjautuu väistämättä kohti moniammatillista mallia, jossa omahoidon tuki ja sen myötä toteutettava elämäntapamuutos muodostuu hoidon perustaksi [25,26].

Tuloksia voidaan hyödyntää tämän asiakasryhmän palvelutarpeiden tunnistamisessa ja palvelujen kehittämisessä. Tulokset osoittivat, että asiakasryhmässä jopa 25 %:lla oli samanaikaisesti diabetes, verenvainetauti ja ylipaino. Tämän tiedon perusteella voidaan suunnitella asiakkaiden omahoitoa tukevia palveluja sellaisiksi, että ne tukevat asiakkaan hoidon tuloksia kaikilla näillä terveysongelma-alueilla. Eri sairauksien ja terveysriskien yhtäaikainen esiintyvyyden tunteminen auttaa ammatillaisia huomioimaan ne asiakkaan palvelutarvetta arvioidessaan, hoitosuunnitelmaa tehtäessä, hoidon toteutuksessa ja arvioinnissa. Tässä tutkimuksessa käytettiin lineaarista korrelaatioanalyysiä näiden yhtäaikaisten esiintyvyyksien tunnistamisessa. Kohderyhmän erityisyydestä johtuen tuloksia tarkasteltiin lähinnä kokonaisuutena ja ilmenemien keskinäisenä vertailuna. Asiakkaan kokonaisuhoitoon suunnittelussa ja vaikkapa ryhmätoiminnan sisällön suunnittelussa voidaan ottaa

huomioon esimerkiksi tupakoinnin ja depression keskinäinen riippuvuus, joka on todettu myös Yoon ym. tutkimuksessa [19]. Käytännössä tämä voi tarkoittaa sitä, että tupakoinnin lopettamiseen tukevassa ryhmässä käsitellään myös masennukseen liittyviä teemoja.

Tutkimuksen tuloksia ei voi yleistää muihin kuin tutkimuksen kohteena olevaan asiakasryhmään, sillä kohde-ryhmä edustaa erityistä terveystieteiden valikoitua ryhmää. Tulokset ovat kuitenkin tutkimuksen tehneen organisaation ja asiakasryhmän kannalta kiinnostavia ja osoittavat, että tällaiset tarkastelut ovat mahdollisia suomalaisen perusterveydenhuollon kontekstissa. Hämeenlinnassa on kehitetty tiedolla johtamisen mallia, jossa potilastietojärjestelmästä saatava tieto, kuten nämä tutkimustulokset otetaan palvelutoiminnan suunnittelun pohjaksi. Lisäksi myös hoitotiimeille ja yksittäisille ammattihenkilöille on alettu suunnitella omia raportteja, joiden avulla voidaan saada ajantasais- ta tietoa myös kliinisen työn tueksi ja palautteeksi omasta työstä.

Tutkimustulokset osoittavat myös mahdollisuudet jatkokutkimuksiin. Asiakasryhmän aineistoa voidaan tutkia pitkittäistutkimuksena, jolloin olisi mahdollista tarkastella, millaisia muutoksia asiakkaiden terveystieteiden ja sairauksien määrissä sekä ilmenemismuodoissa tapahtuu ajallisesti esimerkiksi ennen ja jälkeen jonkin intervention. Jatkokutkimuksissa aineistoon kannattaisi liittää myös palvelutapahtumiin liittyvää dataa.

### ***Tiedonlouhinnan ja -hallinnan mahdollisuudet***

Perusterveydenhuollon kehittämistarpeet nousevat erityisesti havaituista ongelmista hoidon tuloksissa [1,2]. Valmisteilla oleva sosiaali- ja terveydenhuollon (sote) –uudistus tulee aikanaan osoittamaan millaisia tiedontarpeita se tulee synnyttämään [3]. On tärkeää, että keskeisiin haasteisiin löydetään ne mittarit, jotka tukevat asiakkaan saamaa terveyshyötyä sekä palvelu- prosessien uudistamistarpeita. Edelleen on varmistettava, että haluttu tieto kertyy järjestelmiin systemaattisesti ja mahdollisimman tehokkaasti.

Vaikka terveydenhuoltolaissa (1326/2010) [5] viitataan kunnan velvollisuuteen määrittellä ne hyvinvointi- ja terveysosoittimensa, jotka luovat perustan kunnan terveyden edistämisen tavoitteille ja toimenpiteille, on monessa kunnassa ja terveydenhuollon organisaatiossa kamppailtu ajantasaisen ja tarpeellisen tiedon saamiseksi. Ongelmien taustalla saattaa olla monia syitä: tiedon tarpeita ei ole tunnustettu, tietoja ei ole tallennettu järjestelmiin, tietoja ei saada järjestelmästä ulos, raportit eivät palvele dynaamista tiedon tarvetta tai tietoa ei osata analysoida. Potilastietojärjestelmä ei sinällään tarjoa riittävää toiminnanohjausta erityisesti asiakkaan terveystieteiden tunnistamiseen ja kirjaamiseen. Tämä vähentää tärkeiden tietojen kirjaamiskattavuutta. Hyppönen [23] ym. ovat tuoneet esiin, että valtakunnallisesti tulisi käynnistää tiedon laadun arviointi ja että mm. rakenteiden käytön kattavuudesta, herkkyydestä ja tarkkuudesta tarvittaisiin seurantatietoa. Myös potilaskertomusjärjestelmien älykkyyden lisääminen ja kirjaajan työtä ohjaavien rakenteisiin perustuvien muistutusten todettiin lisäävän suosituksen aikaista hoidon ajoittamista. Erityisesti perusterveydenhuollon osalta tuotiin esiin rakenteisen tiedon merkitys prosessi- ja kustannusvaikutusten seuraamiseksi. Tässäkin tutkimuksessa olisi voitu hyödyntää useampia organisaation tärkeiksi määrittämiä terveystieteiden, kuten masennuksen oireita mittaavaa Beckin depressioindeksiä (BDI) tai alkoholiriippuvuutta mittaavaa AUDIT-mittaria, mutta niitä ei ollut kirjattu riittävän kattavasti eli tässä tutkimuksessa 60 % kattavuudella asiakasryhmän jäsenistä.

Monet terveydenhuollon organisaatiot kokevat haasteellisena potilastietojärjestelmään kirjatun tiedon hyödyntämismahdollisuudet. Järjestelmätoimittajat tarjoavat organisaatioille valmiita ja räätälöityjä raportteja, mutta ne eivät useinkaan pysty vastaamaan nopeisiin tiedon tarpeen muutoksiin. Tiedon lisääntyminen synnyttää yleensä uusia kysymyksiä ja tiedontarpeita, joten tiedonlouhinnan prosessin tulisi olla iteratiivinen. Tiedonlouhinta tarjoaa valtavat mahdollisuudet tutkia potilastietojärjestelmiin tallennettuja tietoja. Sen avulla voidaan yhdistää useita eri tietoaineistoja, joiden avulla saadaan käsitys sekä potilaan terveystilasta että hänen

elintavoistaan. Tämä auttaa esimerkiksi ennaltaehkäisevässä terveydenhuollossa [18,19].

Eduksi on, jos tiedon louhimisen osaaminen on lähellä tiedon käyttäjiä eli organisaatioilla itsellään. Tällöin voidaan yhdessä tarkastella potilastietojärjestelmän raportointitietokannasta louhittua tietoa ja tehdä tarkennuksia tiedon tarpeisiin tulosten mukaan. Tässä artikkelissa tutkituista kroonisista kansansairauksista tallennettavista tietosisällöistä löytyy kansallinen ohjeistus [21], mutta monisairastuvuus käsite on vieras, toisin kuin kansainvälisesti [12,13]. Hämeenlinnassa päädyttiin varmistamaan oma tiedonlouhinnan osaaminen palkkaamalla organisaatioon tiedonhallinnan osajia [9]. Osaamisen lisääntyä kyettiin rajallisten valmisraporttien sijasta tekemään raportointitietokantaan suoraan SQL-kyselyjä ja analysoimaan tietoja monipuolisesti tiedonlouhinnan ohjelmilla, kuten tässä tutkimuksessa käytetyllä KNIME-ohjelmalla. Tiedon käyttäjien eli johdon ja klinikoiden kanssa käytiin jatkuvaa vuoropuhelua, jolloin vastauksia saatiin jatkuvasti kehittyviin ja tarkentuviin tiedontarpeisiin. Tämä artikkeli kuvaa tutkimusta, jossa perusterveydenhuollon kannalta kiinnostaviin kysymyksiin voitiin tuottaa tiedonlouhinnan keinoin vastauksia.

## Johtopäätökset

Systemaattisella päätöksenteolla, kirjaamisen ja tiedonhallinnan kehittämällä voidaan potilastietojärjestelmään luoda tietovaranto, jota analysoimalla voidaan tunnistaa korkeimman riskin asiakkaita. Heille voidaan räätälöidä juuri heidän tarpeitaan vastaavia palveluita, joissa kokonaisuuden hallinta ja omahoidon tuki ovat tärkeitä kulmakiviä. Systemaattinen kirjaaminen parantaa päätöksenteon tuen ja muiden digitaalisten palveluiden mahdollisuutta arvonlisään, kun tietoperusta on laajempi. Artikkelissa kuvatussa systeemissä kehittämistyössä voidaan nähdä paljon asiakastiedon hyötykäytön mahdollisuuksia – potentiaalisesti asiakkaan parhaaksi ne muuttuvat vain aktiivisella tietovaraston rakentamisella ja hyödyntämisellä.

## Lähteet

- [1] Terveyden ja hyvinvoinnin laitos. Hoitoonpääsy perusterveydenhuollossa. Verkkosivu ja tilastokatsaus. THL; 2015. Saatavissa: <https://www.thl.fi/fi/tilastot/tiedonkeruut/hoitoonpaasy-perusterveydenhuollossa>. (Viitattu 6.9.2015).
- [2] OECD ja EU. Health at a Glance: Europe 2014. Verkkojulkaisu. Paris: OECD Publishing; 2014. Saatavissa: DOI: [http://dx.doi.org/10.1787/health\\_glance\\_eur-2014-en](http://dx.doi.org/10.1787/health_glance_eur-2014-en) (Viitattu 6.9.2015).
- [3] Sosiaali- ja terveys. Verkkosivu. Suomen Kuntaliitto 2015. Saatavissa: <http://www.kunnat.net/fi/asiantuntijapalvelut/soster/Sivut/default.aspx>. (Viitattu 6.9.2015).
- [4] Tieto hyvinvoinnin ja uudistuvien palvelujen tukena - Sote-tieto hyötykäyttöön -strategia 2020. STM 2014 Saatavissa: <http://www.julkari.fi/handle/10024/125500> (Viitattu 24.11.2015).
- [5] Finlex. Terveydenhuoltolaki (30.12.2010/1326). Verkkosivu. Saatavissa: <https://www.finlex.fi/fi/laki/ajantasa/2010/20101326>. (Viitattu 31.12.2015).
- [6] Bodenheimer T, Wagner E, Grumbach K. Improving Primary Care for Patients With Chronic Illness. (Reprinted) JAMA, October 9, 2002;288(14):1775-1779.
- [7] Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient Self-management of Chronic Disease in Primary Care. (Reprinted) JAMA, November 20, 2002;288(19):2469-2475.
- [8] POTKUlla terveyshyötyä. Verkkosivu. Väli-Suomen Kaste-hanke; 2012–2014. Saatavissa <http://www.potkuhanke.fi/potku-hanke?lang=fi-FI>. (Viitattu 22.2.2016).
- [9] Potku 2-hanke. Verkkosivu. Saatavissa: <http://www.potkuhanke.fi/>. Väli-Suomen Kaste-hanke 2012–2014. (Viitattu 15.8.2015).
- [10] The Chronic Care model. Verkkosivu. Improving chronic illness care 2006-2016. Saatavissa: [http://www.improvingchroniccare.org/index.php?p=The\\_Chronic\\_Care\\_Model&s=2](http://www.improvingchroniccare.org/index.php?p=The_Chronic_Care_Model&s=2). (Viitattu 21.5.2013).

- [11] POTKU-hanke. Terveyshyötymalli Väli-Suomen terveyskeskuksiin 2012–2014. Verkkosivu. POTKU-hanke. Luettavissa: <http://www.potkuhanke.fi/fi>. (Viitattu 9.1.2015).
- [12] de Groot V, Beckerman H, Lankhorst, GJ, Bouter LM. How to measure comorbidity: a critical review of available methods. *Journal of Clinical Epidemiology* 2003;56:221–229.
- [13] Roque F, Jensen P, Schmock H, Dalgaard M, Andreatta M, Hansen T, et al. Using Electronic Patient Records to Discover Disease Correlations and Stratify Patient Cohorts. *PLoS Computational Biology*, [www.ploscompbiol.org](http://www.ploscompbiol.org), 2011 Aug; 7(8):e1002141.
- [14] Dykes PC, Collins S A. Building Linkages between Nursing Care and Improved Patient Outcomes: The Role of Health Information Technology. *OJIN: The Online Journal of Issues in Nursing* 2013; 18 (3): Manuscript 4. September 30. Saatavilla: <http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Vol-18-2013/No3-Sept-2013/Nursing-Care-and-Improved-Outcomes.html>. (Viitattu 25.11.2016)
- [15] Turunen A. Data-analyysin monet mahdollisuudet. *Tieteessä tapahtuu* 2/2012; s. 54–56.
- [16] Bath P. Data Mining in Health and Medical Information. Teoksessa: *Annual Review of Information Science and Technology*, Volume 38, Issue 1, Article first published online: 22 SEP; 2005. s.331–361.
- [17] Bifet A. *Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams*. IOS Press, Amsterdam, Netherlands 2010. Saatavissa: <http://site.ebrary.com/lib/uef/Doc?id=10375704&ppg=15> (Käytettävissä Itä-Suomen yliopiston sekä KYSin verkoissa).
- [18] Alizadehsani R, Habibi J, Hosseini MJ, Mashayekhi H, Boghrati R, Ghandeharioun A, Bahadorian B, Sani ZA. A data mining approach for diagnosis of coronary artery disease. doi: 10.1016/j.cmpb.2013.03.004 . *Epub* 2013 Mar 25. *Comput Methods Programs Biomed.* 2013 Jul;111(1):52-61.
- [19] Yoon S, Taha B, Bakken S. Using a Data Mining Approach to Discover Behavior Correlates of Chronic Disease: A Case Study of Depression. Saatavilla: <http://ebooks.iospress.nl/publication/36575>. *Stud Health Technol Inform.* 2014; 201: 71–78.
- [20] Scopus –tietokanta.Opiskelijan käyttöoikeudella Itä-Suomen yliopistosta. Scopus® on Elsevier B.V:n rekisteröimä; 2015.
- [21] Virkkunen H, Mäkelä-Bengs P, Vuokko R (toim.). *Terveydenhuollon rakenteisen kirjaamisen opas. Keskeisten kertomusrakenteiden kirjaaminen sähköiseen potilaskertomukseen – Osa I. Versio 2015*. ISBN 978-952-302-479-3 (verkkojulkaisu). Terveyden ja hyvinvoinnin laitos (THL). Ohjaukset 2/2015. 120 sivua. Helsinki 2015. Saatavilla: <http://urn.fi/URN:ISBN:978-952-302-479-3> (Viitattu 29.11.2015)
- [22] Tilastotietoja suomalaisten terveydestä ja hyvinvoinnista. Verkkopalvelu. THL [Sotkanet.fi](http://www.sotkanet.fi); 2015. Saatavissa: <https://www.sotkanet.fi/sotkanet/fi/index>. (Viitattu 1.11.2015).
- [23] Hyppönen H, Vuokko R, Doupi P, Mäkelä-Bengs P (toim.). *Sähköisen potilaskertomuksen rakenteistaminen. Menetelmät, arviointikäytännöt ja vaikutukset 2014*. Terveyden- ja hyvinvoinnin laitos raportti 31/2014. [Verkkodokumentti] Saatavissa: [https://julkari.fi/bitstream/handle/10024/125442/URN\\_ISBN\\_978-952-302-381-9.pdf?sequence=1](https://julkari.fi/bitstream/handle/10024/125442/URN_ISBN_978-952-302-381-9.pdf?sequence=1) (Viitattu 24.11.2015).
- [24]. Potku 2 -hankkeen loppuarviointi. Ramboll Management Consulting Oy 2014. Saatavissa: <http://www.potkuhanke.fi/fi/dokumentit-ja-materiaalit-ii/viewdownload/84-lop-puraportti-vali-suomen-potku2/640-ulkoinen-loppuarviointi>. (Viitattu 10.10.2015).
- [25] Johanson J-E, Husman P, Uusikylä P. Viisto, väistö ja vuorovaikutus - Julkisen ja yksityisen terveydenhuollon kanssakäymisen muotoilu. NodeHealth-hankkeen synteisiraportti. Tampereen yliopisto, työterveyslaitos, Tekes. Suomen Yliopistopaino Oy – Juvenes Print, Tampere; 2015. Saatavissa: [http://tampub.uta.fi/bitstream/handle/10024/98376/viisto\\_vaisto\\_ja\\_vuorovaikutus\\_2015.pdf?sequence=1](http://tampub.uta.fi/bitstream/handle/10024/98376/viisto_vaisto_ja_vuorovaikutus_2015.pdf?sequence=1). (Viitattu 9.1.2016).
- [26] Väisänen L. Väli-Suomen Potku, väliraportti. Hämeenlinna/Häme POTKU osahanke 1.11.2012 –

- 31.10.2014. Saatavissa: <http://www.potkuhanke.fi/fi/dokumentit-ja-materiaalit-ii/finish/82-loppuraportit-hame-potku/630-hameenlinnan-loppuraportti>. (Viitattu 15.11.2015).
- [27] Kanavamallin väliarviointi. Kooste arviointiaineis-  
tosta Hämeenlinnan Terveyspalvelut - arviointi Kana-  
vamallin tavoitteet. Verkkojulkaisu. Hämeenlinnan  
terveyspalvelut. Saatavissa <http://www.potkuhanke.fi/fi/dokumentit-ja-materiaalit-ii/finish/82-loppuraportit-hame-potku/634-hameenlinnan-kanavamallin-arviointi>.  
(Viitattu 22.2.2016).
- [28] Rönkkö, I. Potilastiedot terveysaseman asiakas-  
ryhmäprofiilien kuvaajana. Itä-Suomen yliopisto, 2015  
Yhteiskuntatieteiden ja kauppatieteiden tiedekunta.  
Sosiaali- ja terveysjohtamisen laitos, Pro gradu-  
tutkielma (Viitattu 29.3.2016)
- [29] KNIME Analytics Platform. KNIME 2014. Saatavissa:  
[www.knime.org/files/Market-ing/Datasheets/KNIME\\_](http://www.knime.org/files/Market-ing/Datasheets/KNIME_Analytics_Platform_PDS.pdf)  
[Analytics\\_Platform\\_PDS.pdf](http://www.knime.org/files/Market-ing/Datasheets/KNIME_Analytics_Platform_PDS.pdf). (Viitattu 16.11.2015).
- [30] Item Set Finder (Borgelt). KNIME 2016. Saatavissa:  
[https://www.knime.org/files/nodedetails/\\_mining\\_sub](https://www.knime.org/files/nodedetails/_mining_subgroup_Item_Set_Finder_Borgelt_.html)  
[group\\_Item\\_Set\\_Finder\\_Borgelt\\_.html](https://www.knime.org/files/nodedetails/_mining_subgroup_Item_Set_Finder_Borgelt_.html). (Viitattu  
31.12.2015).
- [31] Linear Correlation. KNIME 2016. Saatavissa:  
[https://www.knime.org/files/nodedetails/\\_statistics\\_Li](https://www.knime.org/files/nodedetails/_statistics_Linear_Correlation.html)  
[near\\_Correlation.html](https://www.knime.org/files/nodedetails/_statistics_Linear_Correlation.html). (Viitattu 31.12.2015).
- [32] Potilasasiakirjojen laatiminen ja käsittely. Opas  
terveydenhuollolle. Sosiaali- ja terveysministeriön jul-  
kaisuja 2012:4. STM 2014. Saatavissa:  
[http://www.julkari.fi/bitstream/handle/10024/112073/](http://www.julkari.fi/bitstream/handle/10024/112073/URN%3ANBN%3Afi-fe201504225719.pdf?sequence=1)  
[URN%3ANBN%3Afi-fe201504225719.pdf?sequence=1](http://www.julkari.fi/bitstream/handle/10024/112073/URN%3ANBN%3Afi-fe201504225719.pdf?sequence=1).  
(Viitattu 13.10.2015).
- [33] Heikkilä T. Tilastollinen tutkimus. Edita Helsinki. 7.  
painos; 2008.
- [34] Hirsjärvi S, Remes P, Sajavaara P. Tutki ja kirjoita.  
20. painos. Kustannusosakeyhtiö Tammi. Helsinki; 2015.

Liite 1.

Muuttujat	Verenpainetauti	Keuhkohtaumatauti	Diabetes	Astma	Depressio	Sydämen_vajaatoiminta	Sepelvaltimotauti	Psykoottinen_sairaus	Karies	Parodontiitti	Reuma	Ylipaino	Tupakointi	LDL-riski	Kohonnut_verenpaine	Kohonnut_verensokeri
Verenpainetauti	1,00	0,01	0,08	0,02	0,04	0,04	0,00	0,01	0,07	0,04	0,06	0,03	0,08	0,00	0,09	0,01
Keuhkohtaumatauti	0,01	1,00	0,03	0,21	0,01	0,03	0,05	0,04	0,02	0,03	0,02	0,05	0,28	0,02	0,02	0,03
Diabetes	0,08	0,03	1,00	0,11	0,05	0,02	0,05	0,02	0,05	0,05	0,18	0,18	0,01	0,17	0,00	0,37
Astma	0,02	0,21	0,11	1,00	0,09	0,07	0,03	0,01	0,03	0,01	0,07	0,04	0,07	0,03	0,02	0,05
Depressio	0,04	0,01	0,05	0,09	1,00	0,05	0,11	0,06	0,02	0,04	0,01	0,07	0,13	0,13	0,02	0,07
Sydämen_vajaatoiminta	0,04	0,03	0,02	0,07	0,05	1,00	0,21	0,02	0,07	0,06	0,05	0,03	0,00	0,03	0,01	0,07
Sepelvaltimotauti	0,00	0,05	0,05	0,03	0,11	0,21	1,00	0,08	0,00	0,00	0,08	0,09	0,07	0,09	0,01	0,02
Psykoottinen_sairaus	0,01	0,04	0,02	0,01	0,06	0,02	0,08	1,00	0,06	0,03	0,02	0,09	0,02	0,00	0,06	0,00
Karies	0,07	0,02	0,05	0,03	0,02	0,07	0,00	0,06	1,00	0,17	0,01	0,03	0,01	0,04	0,03	0,01
Parodontiitti	0,04	0,03	0,05	0,01	0,04	0,06	0,00	0,03	0,17	1,00	0,02	0,05	0,10	0,04	0,03	0,00
Reuma	0,06	0,02	0,18	0,07	0,01	0,05	0,08	0,02	0,01	0,02	1,00	0,09	0,02	0,01	0,04	0,05
Ylipaino	0,03	0,05	0,18	0,04	0,07	0,03	0,09	0,09	0,03	0,05	0,09	1,00	0,02	0,03	0,02	0,12
Tupakointi	0,08	0,28	0,01	0,07	0,13	0,00	0,07	0,02	0,01	0,10	0,02	0,02	1,00	0,04	0,02	0,03
LDL-riski	0,00	0,02	0,17	0,03	0,13	0,03	0,09	0,00	0,04	0,04	0,01	0,03	0,04	1,00	0,03	0,11
Kohonnut_verenpaine	0,09	0,02	0,00	0,02	0,02	0,01	0,01	0,06	0,03	0,03	0,04	0,02	0,02	0,03	1,00	0,07
Kohonnut_verensokeri	0,01	0,03	0,37	0,05	0,07	0,07	0,02	0,00	0,01	0,00	0,05	0,12	0,03	0,11	0,07	1,00

Kuvio 1. Lineaarinen korrelaatiotaulukko asiakasryhmän sairaus- ja terveystietiedoista.