

# Korkeiden potilaskohtaisten sosiaali- ja terveystalustusten ennustaminen

Larissa Magdaleena Laamanen

Helsingin Yliopisto

Valtiotieteellinen tiedekunta

Yhteiskuntapolitiikka

Maaliskuu 2019



HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty Valtiotieteellinen tiedekunta		Laitos – Institution – Department Sosiaalitieteiden laitos	
Tekijä □– Författare – Author Larissa Magdaleena Laamanen			
Työn nimi – Arbetets titel – Title Korkeiden potilaskohtaisten sosiaali- ja terveystalustusten ennustaminen			
Oppiaine – Läroämne – Subject Yhteiskuntapolitiikka			
Työn laji – Arbetets art – Level Pro gradu -tutkielma	Aika – Datum – Month and year Maaliskuu 2019	Sivumäärä – Sidoantal – Number of pages 60	
Tiivistelmä – Referat – Abstract			
<p>Aikaisempien tutkimusten mukaan pieni väestönosa aikaansaa suurimman osan terveydenhuollon kustannuksista. Potilaan kuulumista kalliiseen tai paljon palveluita käyttävien potilaiden ryhmään selittävät korkea ikä, diabetes, keuhkohtaumatauti, sydämen vajaatoiminta, masennus ja muut mielenterveyden ja käyttäytymisen häiriöt, tuki- ja liikuntaelinsairaudet, verisuonisairaudet sekä vanhuspalveluiden ja lastensuojelun asiakkuus. (Leskelä et al. 2013; Chechulin et al. 2014; Reho et al. 2018; Kapiainen et al. 2010; Reid et al. 2003; Rais et al. 2013; Ash et al. 2001; Calver et al. 2006.)</p> <p>Tässä poikkileikkaustutkielmassa selvitetään, mitkä yksittäiset tekijät selittävät terveydenhuollon potilaan päättymistä kaikista kalleimpaan 10 prosenttiin potilaista. Tämän lisäksi selvitetään, lisääkö sosiaalihuollon asiakkuus potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään.</p> <p>Tutkimusaineisto koostuu kolmesta rekisteriaineistosta: perusterveydenhuollon Effic-a-datasta, sosiaalihuollon ProConsona-datasta ja erikoissairaanhoidon laskudatasta. Kaikki aineistojen tiedot koskevat vuotta 2017, ja ne käsittävät 6000 asukaan kunnan koko julkisen sektorin sosiaali- ja terveydenhuollon käynnit, asiakkuustiedot, diagnoosit ja toimintolajit. Tutkielman otoskoko on 5621 ja puuttuvien tietojen määrä 1,4 prosenttia. Analyysimenetelmänä käytettiin logistista regressioanalyysia. Kalliin potilaan määritelmänä käytettiin kuulumista kalleimpaan 10 prosenttiin potilaista. Analyysiin valittiin 25 erillistä muuttujaa, jotka käsittävät iän ja sukupuolen lisäksi yksittäisiä erikoissairaanhoidon diagnooseja ja sairausryhmiä, perusterveydenhuollon tulosityitä ja toimintolajeja sekä tiedon sosiaalihuollon asiakkuudesta.</p> <p>Tulokset vahvistavat suurelta osin aikaisempien paljon palveluita käyttäviä ja kaikista kalleimpia potilaita selvittäneiden tutkimusten tuloksia. Kallein 10 % potilaista aikaansai tutkittavan kunnan terveydenhuollon menoista vuonna 2017 lähes 90 %. Yksittäisen potilaan kuulumista kalliiden potilaiden ryhmään ennusti vahvasti masennukseen ja muihin mielenterveyden ja käyttäytymisen häiriöihin, tuki- ja liikuntaelinsairauksiin ja eteisvärinäan liittyvät perusterveydenhuollon käynnit, aikuistyyppin diabetesdiagnoosi sekä sosiaalihuollon ja vanhuspalveluiden asiakkuus. Tuloksissa nousi esille myös eteisvärinäan liittyvien perusterveydenhuollon käyntien, kotihoidon ja kehitysvammahuollon tapahtumien sekä perusterveydenhuollon ja erikoissairaanhoidon vuodeosastojaksojen yhteys korkeisiin potilaskohtaisiin kustannuksiin.</p> <p>Sosiaali- ja terveydenhuollon kustannusten ennustaminen ja paljon palveluita käyttävien ja hoitoa tarvitsevien potilaiden tunnistaminen ovat ajankohtaisia aiheita, joilla on suuri yhteiskunnallinen ja kansantaloudellinen merkitys. Paljon palveluita käyttävien potilaiden ja heidän hoidontarpeensa tunnistaminen tarjoaa palveluntuottajille mahdollisuuden luoda näille potilasryhmille yksilöidympiä hoitopolkua, parantaa hoidon kustannusvaikuttavuutta, keskittyä ongelmien ehkäisyyn sairauksien parantamisen sijaan, ja mahdollisesti myös hidastaa kustannusten kasvua.</p>			
Avainsanat – Nyckelord – Keywords sosiaali- ja terveystalustukset, paljon palveluita käyttävät potilaat, terveydenhuollon kustannukset, kustannusten ennustaminen			

# Sisällys

1 Johdanto	1
2 Teoreettinen viitekehys	5
2.1 Pieni ryhmä aikaansaa suuren osan kustannuksista	5
2.2 Asiakas vie tilaa potilaalta	10
2.3 Kallis potilas	11
2.4 Odotettavissa olevat tulokset	16
3 Tutkimuksen tarkoitus, tavoite ja tutkimuskysymykset	18
4 Aineisto ja menetelmät	19
4.1 Aineiston kuvaus	19
4.2 Muuttujien valinta	22
4.3 Muuttujien esittely	24
4.4 Regressiomalli ja vetosuhteet	30
5 Analyysi ja tulokset	32
5.1 Kuvailevaa tarkastelua	32
5.2 Mallin muodostaminen ja tulokset	38
6 Pohdinta	43
6.1 Keskeiset tulokset ja niiden tulkinta	43
6.2 Luotettavuuden arviointi	46
6.3 Tutkimuseettiset näkökulmat	51
7 Johtopäätökset	53
Lähteet	56

# 1 Johdanto

Suomalainen universaali sosiaali- ja terveydenhuolto on jo vuosikymmenten ajan toiminut yhtenä suomalaisen hyvinvointivaltion peruspilareista. Alati kasvavat resurssipula, talouspaineet, väestön vanheneminen, sosioekonomiset terveyserot sekä kroonisten sairauksien yleistyminen ovat kuitenkin tällä vuosituhannella aiheuttaneet julkiseen sosiaali- ja terveysalaan suuria muutospaineita. Vuonna 2016 pelkästään terveydenhuollon menot olivat 20,5 miljardin euron suuriset, mikä on 9,5 prosenttia Suomen bruttokansantuotteesta (THL 2018). Väestöllinen huoltosuhde, eli työikäisen väestön suhde huollettavan väestön määrään oli vuonna 2017 60 ja sen ennustetaan heikkenevän sekä lyhyellä että pitkällä aikavälillä nuoren väestön vähenemisen ja ikääntyneiden määrän lisääntymisen johdosta (Tilastokeskus 2018a).

Sosiaali- ja terveystalouden ennustetaan kasvavan nykyisellään 2,4 prosenttia vuodessa (STM 2016). Sosiaali- ja terveyspalveluiden ongelmakohdat on tunnustettu myös korkeimpia poliittisia tahoja myöten, minkä johdosta nykyinen hallitus on tehnyt parhaansa järjestelmän uudistamiseksi ja kustannusten kasvun hillitsemiseksi. Sote- ja aluehallintouudistuksen tavoitteena on muun muassa tehdä palveluiden rahoituksesta läpinäkyvää, lisätä kansalaisten työ- ja toimintakykyä, käyttää resursseja tehokkaasti ja oikea-aikaisesti, sujuvoittaa hoito- ja palveluketjuja sekä siirtää painopiste sairauksien hoidosta niiden ennaltaehkäisyyn ja hyvinvoinnin edistämiseen (STM 2016). Puhuttaessa hoito- ja palveluketjujen sujuvoittamisesta ja terveystalouden ongelmien ennaltaehkäisystä, tulee ajankohtaiseksi kysymys hoidon kustannusvaikuttavuudesta ja kustannusten ennustamisesta. Tutkimukset osoittavat, että sosiaali- ja terveydenhuollon kustannukset jakautuvat yksilöiden kesken hyvin epätasaisesti, ja palveluita paljon käyttävät ja tarvitsevat potilaat aikaansaavat määräänsä suuremman osan sosiaali- ja terveystalouksista (esim. Leskelä et al. 2013; Chechulin et al. 2014; Rais et al. 2013; Kapiainen et al. 2010).

Aihe ei ole ajankohtainen pelkästään akateemisissa piireissä, vaan sen yhteiskunnallinen merkitys tunnustetaan valtion ylimpiä toimielimiä myöten. Väitettä vahvistaa esimerkiksi Valtion tarkastusviraston vuonna 2017 Sosiaali- ja terveysministeriölle antama lausunto, jonka mukaan paljon palveluita tarvitsevien ja käyttävien potilaiden määrää tai heidän käyttämiensä palveluiden kustannuksia ei ole selvitetty kansallisella tasolla riittävän laajasti, vaikka aiheen valtiontaloudellinen merkitys on kiistaton. Jos näitä asiakkaita, tai heidän palveluntarvettaan ja heille sopivia hoitoketjuja ei tunnusteta, voi se johtaa potilaiden terveydentilan heikkenemiseen, kalliimpien sosiaali- ja terveyspalveluiden käyttöön ja sitä kautta myös kustannusten kasvuun. (VTV 2017.)

Tämän pro gradu -tutkielman aiheena on potilaskohtaiset sosiaali- ja terveydenhuollon kustannukset ja niiden epätasainen jakautuminen. Tutkielman tavoitteena on selvittää, mitkä yksittäiset tekijät ennustavat korkeita potilaskohtaisia sosiaali- ja terveydenhuollon kustannuksia. Tutkielman tarkoituksena on toteuttaa poikkileikkaustutkimus yhden suomalaisen, alle 6000 asukkaan, kunnan terveys- ja sosiaalipalveluiden käyttäjistä ja selvittää logistisen regressiomallinnuksen avulla, mitkä yksittäiset tekijät ennustavat potilaan kuulumista kaikista kalleimpaan potilaiden kymmenykseen.

Tutkielma toteutetaan toimeksiantona Mehiläiselle, joka tuottaa tutkittavan kunnan julkisia sosiaali- ja terveyspalveluita. Tutkielmassa löydettävien potilaskohtaisiin sosiaali- ja terveydenhuollon kustannuksiin vaikuttavien tekijöiden pohjalta Mehiläisen on tarkoitus rakentaa ennustava työkalu, joka auttaa tunnistamaan ne asiakkaat, joilla on korkein todennäköisyys siirtyä palveluiden suurkäyttäjiksi tai kalliiksi potilaiksi, ja täten aiheuttaa kasvavia resurssipaineita sekä ylisuuria sosiaali- ja terveydenhuollon kustannuksia. Ennustavan työkalun tavoitteena on tunnistaa kohonneen riskin potilasryhmät hyvissä ajoin ennen ennustettavia päätetapahtumia, jolloin tämän riskin toteutuminen on vielä mahdollista välttää. Korkeita potilaskohtaisia sosiaali- ja terveydenhuollon kustannuksia ennustavien tekijöiden sekä paljon palveluita käyttävien ja tarvitsevien potilaiden tunnistaminen auttavaa Mehiläistä kehittämään tutkimuksen kohteena olevan kunnan sosiaali- ja terveyspalveluita yhtenä kokonaisuutena,

parantamaan hoitoketjujen kustannusvaikuttavuutta ja ennaltaehkäisemään terveysongelmia.

Tutkimuksen kvantitatiivisessa analyysissä käytettävät muuttujat on valittu aikaisempaa tutkimuskirjallisuutta hyödyntäen. Aiheesta löytyy kiitettävästi sekä kansainvälistä että suomalaista tutkimuskirjallisuutta. Useissa tutkimuksissa on onnistuttu identifioimaan käyntimäärien lisäksi useita yksittäisiä diagnooseja ja sairausryhmiä, jotka ennustavat yksittäisten potilaiden kuulumista kalleimpaan potilasryhmään (ks. esim. Chechulin et al. 2014; Ash et al. 2001; Reid et al. 2003; Kapiainen et al. 2010; Reho et al. 2018). Tutkielman ensisijainen tutkimuskysymys on: *Mitkä terveydenhuollon suoritteet lisäävät yksittäisen potilaan todennäköisyyttä kuulua kaikista kalleimpien potilaiden ryhmään?*

Tutkimuskirjallisuuteen tutustuminen osoittaa, että pieni väestönosa aikaansaa yhteiskunnalle suurimman osan terveydenhuollon kustannuksista sekä Suomessa että kansainvälisesti (ks. esim. Leskelä et al. 2013; Reho et al. 2018; Kapiainen et al. 2010; Chechulin et al. 2014; Reid et al. 2003; Rais et al. 2013). Suomessa kalliimpia asiakasryhmiä ovat esimerkiksi vanhuspalveluiden ja lastensuojelun asiakkaat sekä kalliita somaattisia hoitoja tarvitsevat asiakkaat. Myös pitkäaikaissairaudet sekä heikko sosioekonominen asema ennustavat palveluiden suurenkäyttöä ja korkeita terveydenhuollon kustannuksia. Kustannukset näyttävät muodostuvan korkeiksi etenkin niillä asiakkailla, jotka käyttävät terveystalveluiden lisäksi vähintään yhtä sosiaalipalvelua. (Leskelä et al. 2013.)

Tutkimuksessa yhdistetään dataa kolmesta eri tilastosta. Tilastot kuvaavat tutkittavan kunnan asukkaiden sosiaali- ja terveystalveluiden käynti- ja asiakkuustietoja vuodelta 2017. Perusterveydenhuollon suoritiedata kerättiin Effica-potilastietojärjestelmästä ja sosiaalipalveluiden data ProConsonasta. Erikoissairaanhoidon aineistona hyödynnetään erikoissairaanhoidon laskuista saatavaa kustannusdataa. Edellä mainitut kolme tilastoa yhdistetään keskenään, minkä avulla päästään tarkastelemaan suoritteita, diagnooseja ja

kustannuksia potilaskohtaisesti. Tutkimusaineistosta etsitään logistisen regressioanalyysin avulla niitä tekijöitä, jotka ennustavat asiakkaiden kuulumista korkeimpia kustannuksia aiheuttavaan potilasryhmään.

Tutkielma etenee siten, että toisessa luvussa hahmotetaan tutkielman teoreettista viitekehystä keskeisten käsitteiden määrittelyn sekä aikaisempaan tutkimuskirjallisuuteen tutustumisen kautta. Kolmannessa luvussa esitetään ja tarkennetaan tutkielman tarkoitus, tavoitteet sekä tutkimuskysymykset. Neljännessä luvussa kuvataan tutkielmassa käytettyä tilastoaineistoa ja sen valintaa, esitellään muuttujat sekä aineiston analyysimenetelmä. Tämän jälkeen (luku 5) siirrytään tutkielman analyysivaiheeseen, jonka aluksi raportoidaan yhdistetyn aineiston kuvailevia tunnuslukuja, rakennetaan logistinen regressiomalli, jolla pyritään ennustamaan tekijöitä, jotka vaikuttavat potilaiden todennäköisyyteen kuulua kaikista kalleimpien potilaiden ryhmään, ja lopuksi esitellään analyysin tulokset. Pohdintaluvussa (luku 6) kootaan yhteen analyysivaiheen tärkeimmät tulokset, verrataan niitä aikaisempien tutkimusten tuloksiin, arvioidaan tulosten luotettavuutta ja tutkimuseettisiä näkökulmia. Tutkielman lopuksi (luku 7) pohditaan tulosten yhteiskunnallista merkitystä ja ehdotetaan aiheita jatkotutkimukselle.

## 2 Teoreettinen viitekehys

Kirjallisuuskatsaus yhdistää tämän tutkielman aihepiirin tutkimuskenttään ja antaa eväät soveltuvan analyysimenetelmän valitsemiseen ja käyttöön. Tässä luvussa tutustutaan aiheesta aikaisemmin tehtyihin tutkimuksiin, syvennetään ymmärrystä sosiaali- ja terveyspalveluiden kustannusten jakautumisesta sekä etsitään määritelmiä tutkimuksen kannalta oleellisille käsitteille. Luvussa kuvataan, miten aikaisemmat suomalaiset ja kansainväliset tutkimukset kuvaavat palveluiden suurkäyttöä ja sosiaali- ja terveydenhuollon kustannusten jakautumista, sekä mitä käsitteitä, menetelmiä ja muuttujia niissä on käytetty. Tutkimuskirjallisuuden etsimiseen käytettiin Google Scholar ja EBSCOhost -hakukoneita sekä Helka-kirjaston tietokannan artikkelihakua. Hakuja tehtiin sekä suomeksi että englanniksi ja käytettyjä hakusanoja olivat: *paljon terveyspalveluita käyttävät, paljon sosiaalipalveluita käyttävät, terveyspalveluiden suurkäyttö, terveyspalveluiden laajakäyttö, terveydenhuollon kustannukset, health care costs, social care costs, high utilization ja high cost utilizers.*

### 2.1 Pieni ryhmä aikaansaa suuren osan kustannuksista

Tutustuminen aikaisempaan tutkimuskirjallisuuteen (ks. esim. Chechulin et al. 2014; Rais et al. 2013; Reid et al. 2003; Ash et al. 2001; Kapiainen et al. 2010; Lee et al. 2017; Reho et al. 2018) osoittaa, että hyvin pieni väestönosa aikaansaa yhteiskunnille suurimman osan terveydenhuollon kustannuksista, ja että tämä näyttäisi olevan sekä pysyvä, että maailmanlaajuinen ilmiö. Chechulin et al. (2014) onnistuivat tutkimuksessaan *Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada)* identifioimaan useita tekijöitä, jotka ennustivat yksittäisten potilaiden siirtymistä kalleimpaan potilasryhmään. Kalliita kustannuksia ennustavia diagnooseja olivat muun muassa diabetes, mielenterveys- ja käytöshäiriöt, keuhkohtaumatauti ja sydämen vajaatoiminta. Myös iän huomattiin ennustavan korkeita kustannuksia: mitä korkeampi ikä, sitä suurempi riski päätyä kalliimpaan potilasryhmään. Tutkimuksen tulokset osoittivat myös, että aikaisempien vuosien



palveluiden käyttö on yksi merkittävimmistä korkeita potilaskohtaisia kustannuksia selittävästä tekijöistä. (Chechulin et al. 2014.) Toisen kanadalaisen tutkimuksen mukaan kallein 5 % potilaista aikaansaa jopa 61 % kaikista Ontarion provinssin terveydenhuollon kustannuksista (Rais et al. 2013).

Vastaavia tutkimuksia on tehty myös Suomessa. Esimerkiksi Oulussa tehdyn Leskelä et al. (2013) tutkimuksen mukaan kalliimpia asiakasryhmiä ovat vanhuspalveluiden ja lastensuojelun asiakkaat sekä kalliita somaattisia hoitoja tarvitsevat asiakkaat. Myös pitkäaikaissairaudet sekä heikko sosioekonominen asema ennustavat palveluiden suorkäyttöä ja tätä kautta suuria terveydenhuollon kustannuksia. Kustannukset näyttivät muodostuvan hyvin korkeiksi niillä asiakkailla, jotka käyttivät terveystalveluiden lisäksi vähintään yhtä sosiaalipalvelua. Kustannusvertailu osoitti, että 10 % Oulun asukkaista kerryttää jopa 81 % kaikista sotekustannuksista. (Leskelä et al. 2013.) Kapiainen ym. (2010) Suomen pääkaupunkiseudun asukkaita koskevan tutkimuksen mukaan alle 10 % potilaista aiheutti 80 % kaikista terveydenhuollon kustannuksista. Tutkimus osoitti myös, että mielisairausdiagnoosi oli kaikista merkittävin yksittäinen korkeita kustannuksia selittävä tekijä. Muita selittäviä tekijöitä olivat dementia, diabetes, keuhkohtaumatauti, astma, masennus, Parkinsonin tauti, sydämen vajaatoiminta, syöpä sekä verenpainetauti. (Kapiainen et al. 2010.)

Ash kollegoineen (2001) ovat tutkineet terveystalveluksia Yhdysvalloissa. Tutkimus koski yksityisillä terveystalveluksilla vakuutettuja potilaita. Tutkimuksessa osoitettiin, että 1 % potilaista saa aikaan jopa 31 % kaikista terveystalveluksista. Tutkimuksessa vertailtiin kahta erilaista kustannusten ennustamiseen soveltuvaa menetelmää, joista toinen perustui aikaisempien vuosien kustannusdataan ja toinen aikaisempina vuosina tehtyihin diagnooseihin. Tärkeiksi korkeita kustannuksia selittäviksi tekijöiksi nousivat muissakin tutkimuksissa esiin nousseet diagnoosit: diabetes, sydämen vajaatoiminta ja keuhkohtaumatauti. (Ash et al. 2001.) Paljon palveluita käyttäviä on tutkittu Yhdysvalloissa enemmänkin; Moturu et al. (2010) toteuttaman tutkimuksen tulokset osoittivat, että kallein 10 % kaikista potilaista saivat aikaan jopa kaksi kolmasosaa kaikista terveydenhuollon kustannuksista Arizonan osavaltiossa. Toisen tutkimuksen

mukaan kallein kymmenys potilaista aikaansai 60 % kaikista kustannuksista (Lee et al. 2017).

Myös masennuksen on huomattu olevan merkittävä korkeita terveydenhuollon kustannuksia selittävä tekijä (Ash et al. 2001). Samaan suuntaan viittaavat myös muut tutkimustulokset, joiden mukaan mielenterveyden ja käyttäytymisen häiriöt olivat yksi parhaiten korkeita kustannuksia selittävistä tekijöistä (Chechulin et al. 2014). Tutkimukset osoittavat myös, että fyysisten terveysongelmien lisäksi psykososiaaliset ongelmat ovat kalleimpien potilasryhmien edustajilla yleisiä: esimerkiksi Reid kollegoineen (2003) nostivat tutkimuksessaan esiin, kuinka jopa joka kolmannella kaikista kalleimman potilasryhmän edustajista oli diagnosoitu mielenterveysongelmia. Suomessa Kapiainen ym. (2010) selvittivät, että jopa 44 prosentilla kalliista potilaista oli todettu joko masennus tai mielisairaus. Viime vuonna valmistuneessa suomalaisessa tutkimuksessa tutkittiin puolestaan paljon palveluita käyttäviä työterveyshuollon potilaita, ja vertailussa tultiin siihen tulokseen, että palveluiden suuren käyttöönnä ennustivat mielenterveyteen ja käyttäytymisen häiriöihin liittyvät diagnoosit, etenkin masennusjaksot ja ahdistushäiriöt, sekä tuki- ja liikuntaelinsairauksiin liittyvät diagnoosit (Reho et al. 2018). Myös yleinen ahdistuneisuushäiriö sekä somaattinen oireilu ovat yhteydessä palveluiden suuren käyttöön (Kujanpää et al. 2017).

Pidemmän aikavälin tutkimukset osoittavat selkeästi, että kustannusten vinoutunut jakautuminen on myös pysyvä ongelma. Esimerkiksi Berk ja Monheit (2001) osoittavat Yhdysvaltojen terveystalouden perusteella, että jakauma on ollut vinoutunut jo 1930-luvulla, jolloin kallein 10 % potilaista käyttivät 59 % kaikista terveydenhuollon rahoista. Vuodesta 1970 lähtien tämä luku on ollut noin 70 %. (Berk & Monheit 2001.) Yhtä johdonmukaisia tuloksia ei kuitenkaan saada, kun tarkastellaan vinoutuneiden kustannusten pysyvyyttä yksilötasolla. Näyttäisi siltä, että kallein tai eniten palveluita käyttävä potilasryhmä ei koostu vuodesta toiseen samoista potilaista, vaan yksittäisten potilaiden liikkuminen kalleimpaan tai paljon palveluita käyttävään potilasryhmään ja siitä pois on tutkimusten mukaan yleistä (Ash et al. 2001; Andersson et al. 2004; Smits et al. 2009). Chechulin et al. (2014) tutkimuksen mukaan jopa 46,2 % vuoden 2010

kalleimpaan potilasryhmään kuuluneista eivät olleet kalleimmassa ryhmässä vuotta aiemmin. Suomalaisessa tutkimuksessa laskettiin puolestaan, että 53,7 % kalleimman potilasryhmän edustajista pysyivät tässä ryhmässä myös seuraavana vuonna (Lappalainen 2015). Toisaalta, vaikka aikaisempien tutkimusten valossa näyttäisi siltä, että kaikista kalleimman ryhmän ja muiden potilasryhmien välillä tapahtuukin vaihtelua joidenkin liikkuesssa siitä pois, jää kalleimmassa ryhmässä jäljelle kuitenkin karkeasti noin puolet, joilla näyttäisi olevan pysyvästi kohonnut tarve sosiaali- ja terveyspalveluiden käytölle. Tutkimusten mukaan niillä potilailla, joilla pysyvät paljon palveluita käyttävässä potilasryhmästä vuodesta toiseen (persistent frequent attenders), on muita paljon palveluita käyttäviä potilaita enemmän sosiaalisia ja psykologisia ongelmia, kroonisia sairauksia ja somaattisia oireita (Smits et al. 2009). Suomessa Koskela et al. (2010) havaitsivat muun muassa korkean painoindeksin (BMI > 30), ärtyvän suolen oireyhtymän, kuolemanpelon ja aikaisemman palveluiden suuren käytön ennustavan potilaan kuulumista paljon palveluita käyttävään ryhmään vuodesta toiseen.

Koska kalleimmat ja eniten palveluita käyttävät potilaat ovat usein juuri niitä potilaita, jotka ovat kaikista sairaimpia ja joiden palveluntarve on suurin, on luonnollista, että aiheesta tehdyt tutkimukset myös säännönmukaisesti osoittavat, että ikä on yksi merkittävimmistä kustannuksia ennustavista muuttujista. Kanadalaisen tutkimuksen mukaan melkein puolet kalleimman potilasryhmän edustajista olivat yli 60-vuotiaita (Reid et al. 2003). Rais et al. (2013) tutkimuksen mukaan jopa 60 % kalleimmista potilaista oli yli 65-vuotiaita. Suomessa Peruspalvelukuntayhtymä KALLIO:ta tutkineen Lappalaisen (2015) mukaan ikääntyneiden osuus paljon palveluita käyttävistä oli noin 50 %. Australiassa Calver kollegoineen (2006) sai tulokseksi 55,9 %. Samansuuntaisia tuloksia raportoivat myös Kapiainen et al. (2010), mutta heidän muodostamansa regressiomallin mukaan yli 65-vuotiaiden lisäksi alle 18-vuotiailla oli keskiarvoa suurempi riski kuulua erittäin kalliiden potilaiden ryhmään. Kalleimmassa potilasryhmässä olevat yli 65-vuotiaat eivät kuitenkaan näytä olevan potilastasolla yhtään sen kalliimpia kuin muut kalleimman ryhmän edustajat (ks. esim. Calver 2006; Reid et al. 2003). Calver et al. (2006) tutkimuksen johtopäätöksissä nostettiin esille huoli siitä, että kalleimpien potilaiden määrä tulee todennäköisesti vain kasvamaan väestön ikääntyessä.

Yksittäisten korkeita kustannuksia ennustavien diagnoosien tai diagnoosiryhmien lisäksi tutkimukset antavat selkeää näyttöä sille, että pelkästään terveysongelmien lukumäärä toimii myös itsessään hyvänä korkeita potilaskohtaisia kustannuksia ennustavana tekijänä. Komorbiditeetillä (comorbidity), tai yhteissairastavuudella, tarkoitetaan tilannetta, jossa potilaalla on yhtäaikaaisesti monta sairautta. Tätä voidaan mitata tutkimuksissa esimerkiksi diagnoosien, lääkemääräysten tai vastaanottokäyntien lukumäärällä. Reid et al. (2003) selvittivät tutkimuksessaan, että kalleimmassa potilasryhmässä jopa 80 % potilaista olivat monisairaita. Rais et al. (2013) tutkimuksen mukaan 83 % kalleimmista potilaista olivat saaneet hoitoa useampaan kuin yhteen terveysongelmaan. Fleishman ja Cohen (2010) selvittivät yhteissairastavuuteen perustuvien mallien käyttöä korkeiden terveydenhuollon kustannusten ennustamisessa ja saivat tutkimuksensa tulokseksi, että etenkin krooniset sairaudet ja niiden lukumäärä olivat merkittäviä korkeita kustannuksia selittäviä tekijöitä. Oulun väestöä koskevassa tutkimuksessa Leskelä et al. (2014) puolestaan selvittivät, että kalliit potilaat käyttävät sosiaali- ja terveystalvueluita myös muita potilaita laajemmin. Kalleimmassa potilasryhmässä käytettiin keskimäärin neljää eri palvelukokonaisuutta, kun keskiarvo kaikilla potilasryhmillä oli 1,1. Tulokset osoittavat, että paljon palveluita käyttävä 10 % potilaista vastasivat lähes kaikesta sosiaalipalveluiden käytöstä samana vuonna. (Leskelä et al. 2014.)

Monet tutkimukset ovat tarkastelleet iän lisäksi myös sukupuolen vaikutusta terveydenhuollon kustannuksiin, mutta sen osalta tulokset eivät ole yhtä johdonmukaisia. Chechulin et al. (2014) tutkimuksessa miespuolisilla potilailla oli suurempi riski kuulua kaikista kalleimpaan 5 prosenttiin. Samaan suuntaan viittaa myös Kapiaisen et al. (2010) Suomen pääkaupunkiseudun terveystalvueluiden potilaskohtaisia kustannuksia selvittäneen tutkimuksen tulokset. Reho kollegoineen (2018) puolestaan saivat tutkimuksessaan vastakkaisia tuloksia; naisilla oli miehiä suurempi todennäköisyys kuulua paljon työterveydenhuollon palveluita käyttävien potilaiden ryhmään. Samaan suuntaan osoittaa Lappalaisen (2015) vertailu, jonka mukaan Peruspalvelukuntayhtymä KALLIO:n kaikista kalleimmista potilaista naiset olivat hieman ylliedustettuina (54 %).

## 2.2 Asiakas vie tilaa potilaalta

Tutkittaessa paljon palveluita käyttäviä tai kaikista kalleimpia yksilöitä, tulee vastaan kysymys siitä, onko tarpeenmukaisempaa käyttää käsitettä potilas vai asiakas. Perinteisenä jakona voidaan pitää sitä, että sosiaalipalveluissa puhutaan asiakkaista ja terveydenhuollossa potilaista. Tämä on nähtävissä myös sosiaali- ja terveydenhuollon lainsäädännössä, jossa käsitteiden erot näkyvät jo lakien otsikkotasolla (ks. esim. Laki potilaan asemasta ja oikeuksista 1992; Laki sosiaalihuollon asiakkaan asemasta ja oikeuksista 2000).

Tällä hetkellä julkisessa keskustelussa on kuitenkin nähtävissä käsitteellinen siirtyminen pois potilaasta ja kohti asiakasta. Esimerkiksi yksi Juha Sipilän hallituksen kärkihankkeista on *asiakaslähtöisyyden* lisääminen terveystaloudessa. Kärkihankkeen tavoitteena on muokata sosiaali- ja terveystalouksien toimintaprosesseista asiakaslähtöisiä ja asiakkaan valinnanvapautta sekä omatoimisuutta tukevia. (Valtioneuvoston kanslia 2018.) Asiakas-sanankäytön aikaisempaa yleisempi käyttö terveydenhuollossa liittyy laajempaan, kansainväliseen ilmiöön, jossa potilaan ja lääkärin suhde on muuttunut aikaisempaa tasavertaisemmaksi, ja potilaille annetaan enemmän autonomiaa sekä valtaa päättää omasta hoidostaan (Deber et al. 2005). Asiakas-sanankäytön koetaan muuttavan potilaan asemaa terveydenhuollossa tekemällä heistä aikaisempaa aktiivisempia toimijoita (Karppinen 2018). Lisäksi muutoksen koetaan siirtävän vastuuta lääkäriltä potilaalle ja tekevän potilaasta ennemminkin kuluttajan, jolloin terveystalouksien päämäärä siirtyy potilaan terveydentilan parantamisesta palveluiden myymiseen (Lääkärilehti 2018; Potilaan lääkäri 2018).

Karppinen (2018) selvitti pro gradu -tutkielmassaan Potilas- ja Asiakas-sanojen käyttöä terveydenhuollon organisaatioiden verkkosivuilla ja tuli analyysissään siihen tulokseen, että käsitteitä käytetään usein rinnasteisesti ja epätarkasti, ja että Potilas-sanankäytön asemaa voidaan pitää terveydenhuollossa edelleen melko vakaana, vaikka Asiakas-sana tuleekin yhä useammin sen rinnalle. Tutkielman johtopäätöksissä mainitaan myös, että vaikka

terveydenhuollossa käytettäisiinkin sanaa potilas, on heitä mahdollista kohdella kuin asiakkaita. (Karppinen 2018.)

Tässä tutkielmassa pitäydytään yksinkertaistamisen vuoksi käsitteiden yleisemmissä määritelmässä, eli jatkossa tutkielman eri osissa viitataan järjestelmällisesti perusterveydenhuollon ja erikoisairaanhoidon osalta potilaisiin ja sosiaalihuollon osalta asiakkaisiin.

## 2.3 Kallis potilas

Tutkittaessa terveydenhuollon kustannuksia ja kalleimpia potilasryhmiä, on tärkeää määritellä mitä tarkoitetaan ”kalliilla potilaalla”. Riippuen siitä, onko vertailussa käytetty kustannusdataa, käyntimääriä, diagnooseja tai kenties näitä kaikkia, puhutaan aikaisemmissa tutkimuksissa usein joko paljon palveluita käyttävistä potilaista (high utilizers) tai kalliista potilaista (high cost users). Tässä tutkielmassa käytetään pääsääntöisesti käsitettä kallis potilas. Monissa tutkimuksissa kalleimman potilasryhmän määritelmänä on käytetty kalleinta 5 % potilaista (Chechulin et al. 2014; Reid et al. 2003; Rais et al. 2013; Zhao et al. 2005 ja Lappalainen 2015). Toinen yleisesti käytetty käytäntö on määritellä kalliiksi potilaiksi potilaiden kallein kymmenes (ks. esim. Fleishman & Cohen 2010; Leskelä et al. 2014; Moturu et al. 2010).

Kivelä et al. (2018) julkaisema artikkeli *Frequent Attenders in Primary Health Care: A Concept Analysis* antaa kuvan hajanaisesta ja vailla varsinaista teoriapohjaa olevasta tutkimuskentästä, jossa edes tutkimusten peruskäsitteiden määrittely ei ole kunnossa. Heidän tarkoituksenaan oli konseptoida ja operationalisoida paljon palveluita käyttävien määritelmä. Kirjallisuuskatsauksen perusteella paljon palveluita käyttävistä potilaista on tehty paljon tutkimuksia parina viime vuosikymmenenä, mutta määritelmä on suurimmassa osassa tutkimuksista erilainen, ja valittujen määritelmien perustelut ontuvat. Yleensä paljon palveluita käyttävien ryhmä määritellään sen perusteella, kuinka monta yksittäistä hoito- tai konsultaatiokertaa asiakkaalla tai potilaalla on esimerkiksi vuoden aikana (toisin sanoen siis käyntimäärän perusteella). Toinen yleinen

tapa on käyttää suhteellista määritelmää, esimerkiksi 5 tai 10 % eniten palveluita käyttäviä. (Kivelä et al. 2018.)

Joissakin tutkimuksissa on päädytty tarkastelemaan vain kalleinta 0,5 % potilaista (top users). Tämä voi käydä tarkoituksenomaiseksi esimerkiksi sellaisessa tapauksessa, jossa tarkoituksena on identifioida juuri ne kaikista kalleimmat, yksittäiset potilaat, ja kohdentaa hallinnolliset toimet heihin tapauskohtaisesti. (Ash et al. 2001.) Jotkut ovat verranneet tutkimuksissaan kalleimpia 1, 5 ja 10 prosenttia potilaista, ja testanneet näillä eri jaoilla rakentamansa kustannuksia ennustavan mallin luotettavuutta (ks. esim. Chechulin et al. 2014; Powers et al. 2005). Chechulin et al. (2014) käyttivät tutkimuksessaan kalliin potilaan kriteerinä kalleimpaan 5 % kuulumista, mutta testasivat myös 1 ja 10 % mukaan tehtyjä malleja, joiden ennustusvoima oli myös hyvä.

Fleishman ja Cohen (2010) lisäsivät omassa tutkimuksessaan vertailuun kustannusten lisäksi myös subjektiivisesti arvioidun terveydentilan (self-reported health status). Kuten monet muutkin tutkijat, tarkastelivat he kalleinta 10 % potilaista ja tulivat tutkimuksessaan siihen tulokseen, että kroonisten sairauksien lukumäärä ja itse arvioitu terveydentila ennustivat korkeita terveystaloustalouksia tulevaisuudessa. (Fleishman & Cohen 2010.) Monissa aikaisemmissa aiheesta tehdyissä tutkimuksissa on tutkittu myös potilaiden sosioekonomisen aseman (social and material deprivation) sekä asuinpaikan (rurality index) yhteyttä korkeisiin potilaskohtaisiin kustannuksiin (ks. esim. Chechulin et al. 2014). Tämän tutkielman aineistossa ei ole kattavaa tietoa potilaiden koulutuksesta, ammatista, tulotasosta tai heille maksetuista tulonsiirroista, joten ne on jätetty kokonaan vertailun ulkopuolelle. Lisäksi asuinpaikan merkitys on tämän tutkielman tutkimusasetelman kannalta epäolennainen, sillä lähes kaikki tämän kyseisen kunnan terveystalouksien potilaat ja sosiaalihuollon asiakkaat asuvat lähtökohtaisesti samassa kunnassa.

Korkeita potilaskohtaisia kustannuksia voidaan ennustaa pääsääntöisesti kahdella tavalla. Ensimmäinen tapa on tarkastella pelkästään aikaisempaa kustannusdataa (prior-

cost), ja selvittää kuinka hyvin esimerkiksi aikaisempien vuosien korkeat kustannukset tai kalleimpaan potilasryhmään kuulumisen selittää nykyisiä korkeita kustannuksia. Toinen tapa on tarkastella diagnooseja (diagnosis-based risk models), mikä antaa pelkkiä kustannuksia yksityiskohtaisemman kuvan käsillä olevasta ongelmasta ja mahdollistaa ongelmaan puuttumisen. Aikaisemmat tutkimukset osoittavat, että diagnooseihin perustuvat ennustavat mallit ovat vähintäänkin yhtä hyviä kuin kustannuksiin perustuvat mallit, ja että näiden menetelmien yhdistäminen tuo vieläkin luotettavamman tuloksen. (Ash et al. 2001.) Tässä tutkimuksessa käytetään kustannus- ja diagnoosipohjaista logistista regressiomallia.

Lee et al. (2017) esittävät artikkelissaan väitteen siitä, ettei kalliiden potilaiden ryhmä ole homogeeninen, ja ettei terveydenhuollon kustannuksia pystytä riittävästi selittämään pelkästään korkeiden käyntimäärien avulla. He lähtivät tarkastelemaan potilaiden kalleinta 10 prosenttia muodostamalla heistä erillisen ”hot-spotters”-ryhmän, johon kuuluivat kaikki ne kalliiksi määritellyn ryhmän potilaat, joilla oli vähintään neljä sairausjaksoa tai päivystyskäyntiä vuoden aikana. Erilaisiin palvelunkäyttöprofiileihin perustuva klusterianalyysi paljasti, että hot-spotters -ryhmä muodosti vain 9 prosenttia potilaiden kalleimmasta kymmenyksestä, ja että suurin osa kalleimpaan 10 prosenttiin kuuluvista potilaista eivät olleet palveluiden suurkäyttäjiä, vaan heistä muodostui ylisuuria kustannuksia lukuisista muista syistä. Tutkimus antaa viitteitä siitä, että pelkästään terveydenhuollon käyntimääriin puuttumalla ei kustannuksia pystytä riittävän tehokkaasti supistamaan, vaan tähän tarvitaan tarkempaa analyysiä kaikista kalleimpien potilaiden tunnusmerkeistä. Tutkijat esittivät myös huolensa siitä, että suuri osa kalliiden potilaiden aiheuttamista kustannuksista ovat erittäin vaikeasti ehkäistävissä (Lee et al. 2017).

Taulukkoon 1 on kerätty vertailua aikaisempien tutkimusten käyttämistä kalliiden potilaiden määritelmistä, analyysimenetelmistä, selittävistä muuttujista sekä otoskoosta.



Taulukko 1

*Teoreettisen viitekehyksen yhteenveto*

Tutkimus	Kalliin potilaan määritelmä	Otoskoko	Analyyssimenetelmä	Käytetyt muuttujat
Chechulin et al. 2014	5 %	10 300 856	Logistinen regressio	ICD10, käyntimäärät, kustannukset
Moturu et al. 2010	10 %	139 039	Data mining	MDC (perustuu ICD-luokitukseen), käyntimäärät, lääkemääräykset
Ash et al. 2001	0,50 %	n. 2,7 milj.	Logistinen regressio	DCG (perustuu ICD-luokitukseen), käyntimäärät, kustannukset
Fleishman & Cohen 2010	10 %	53 918	Logistinen regressio	Kroonisten terveysongelmien lukumäärä (DCG)
Reid et al. 2003	5 %	2 997 808	Kuvailevat tunnusluvut	ADG ja EDC, käyntimäärät, kustannukset
Rais et al. 2013	5 %	n. 13,6 milj.	Kuvailevat tunnusluvut	EDC, käyntimäärät, kustannukset
Zhao et al. 2005	5 %	> 1 milj.	Logistinen regressio	DCG/HCC, kustannukset, lääkemääräykset
Powers et al. 2005	1, 5 ja 10 %	344 832	Logistinen regressio	PHD (muokattu), lääkemääräykset, kustannukset
Leskelä et al. 2014	10 %	152 494	Logistinen regressio	Käyntimäärät, diagnoosit, kustannukset
Lappalainen et al. 2015	5 %	33 352	Kuvailevat tunnusluvut	Käyntimäärät, diagnoosit, lääkemääräykset, kustannukset
Kapiainen et al. 2010	Yli 50 000 tai yli 75 000 e/v	984 591	Logistinen regressio	ICD10, kelakorvaukset, lääkemääräykset, kustannukset

*DCG = Diagnostic Cost Groups*

*EDC = Extended Diagnosis Clusters*

*ADG = Aggregated Diagnosis Groups*

*MDC = Major Diagnostic Categories*

*HCC = Hierarchial Condition Category*

*PHD = Pharmacy Health Dimensions*

Taulukosta 1 on nähtävissä, että suurimmassa osassa tutkimuksista on käytetty analyysimenetelmänä logistista regressioanalyysiä. Logistinen regressioanalyysi on valittu menetelmäksi silloin, kun tarkoituksena on ollut luoda korkeita kustannuksia ennustava tilastollinen malli. Joissakin tutkimuksissa on tyydytty pelkästään kuvailemaan paljon palveluita käyttäviä, ilman erillisen ennustavan mallin muodostamista (esim. Reid et al. 2003; Rais et al. 2013). Suuressa osassa tutkimuksista kalliin potilaan määritelmä on 5 % eniten kustannuksia aiheuttavista potilaista. Valitun prosentiosuuden perusteena näyttäisi olevan useimmiten vain hyväksi todettu käytäntö, vaikkakin joissakin tapauksissa muusta tutkimuskentästä poikkeavaa määritelmää on perusteltu hallinnollisin syin (ks. esim. Ash et al. 2001).

Tarkastelemalla Taulukkoa 1, voidaan nähdä, että aikaisempien tutkimusten otoskoot vaihtelevat kymmenistä tuhansista kymmeneen miljooniin, ja otoskoko on suoraan verrannollinen tutkittavan alueen kokoon ja väkimäärään. Otoskoot näyttäisivät näissä tutkimuksissa olevan pääsääntöisesti riittävän suuria, jotta tutkimustulosten yleistettävyyttä ei vaarannu. Tutkimuksissa käytetyissä muuttujissa on myös verrattain paljon vaihtelua. Suurimmassa osassa tutkimuksista on kuitenkin käytetty sekä käyntimäärä, kustannus, että diagnoositietoja sekä lääkemääräyksiä niissä tapauksissa kuin se on ollut mahdollista. Diagnoosien valinnassa ja ryhmittelyssä on paljon eroja, mutta useimmat ryhmittelyt perustuvat kansainväliseen ICD10-tautiluokitukseen.

Aikaisemmat tutkimukset osoittavat, että diagnoosien lisääminen vertailuun on hyödyllistä, muttei välttämätöntä; pelkästään korkeiden kustannusten ennustaminen edellisten vuosien kustannusdatalla näyttäisi tuottavan myös toimivia malleja (Ash et al. 2001; Chechulin et al. 2014). Pelkkä kustannusdata tai käyntimäärät eivät kuitenkaan anna tietoa ongelmista yksilötasolla, eivätkä siten välttämättä tarjoa riittäviä työkaluja tuotettavien sotepalveluiden kehittämiseksi.

Teoreettisen viitekehyksen perusteella voidaan todeta, että terveydenhuollon kustannuksien ennustamiseen liittyvää kansainvälistä tutkimuskirjallisuutta on saatavilla

kattavasti. Sosiaalihuollon palveluiden osalta vastaavia tutkimuksia ei ole vielä samassa mittakaavassa tehty, sosiaali- ja terveydenhuollon kustannuksia yhdistävistä tutkimuksista puhumattakaan. Näiden tutkimusten puute, selkeiden käsitteiden määrittelyiden sekä teoriapohjan puuttumisen lisäksi, kertoo selvästi siitä, että tutkimusaihe on verrattain uusi, ja että sosiaali- ja terveydenhuollon kustannuksia ja niiden epätasaista jakautumista kartoittaville tutkimuksille on tarvetta. Tähän vaikuttanee olennaisesti myös se, että sosiaali- ja terveystaloudilla on perinteisesti ollut eri järjestäjät, eikä niitä ole täten pidetty yhtenä kokonaisuutena. Tämä seikka tulee ainakin järjestämistä osalta muuttumaan suunnitteilla olevan soteuudistuksen myötä. Kuten Lappalainen (2015) huomauttaa omassa tutkimuksessaan, ei sosiaalipalveluilla ole olemassa mitään universaalia määritelmää, vaan se vaihtelee maasta toiseen. Tästä johtuen joissakin terveystalouksien käytössä ja kustannuksia käsittelevissä tutkimuksissa voi olla mukana joitakin tietoja myös sosiaalipalveluiden käytöstä.

## 2.4 Odotettavissa olevat tulokset

Tutkimuskirjallisuuteen tutustumisen perusteella voidaan tehdä jonkinlaisia alustavia oletuksia tai hypoteeseja siitä, minkä tekijöiden uskotaan nousevan tässä tutkielmassa hahmoteltavassa logistisessa regressiomallissa esiin korkeita potilaskohtaisia sosiaali- ja terveydenhuollon kustannuksia ennustavina tekijöinä. Yksi näistä tekijöistä on oletettavasti ikä, jonka monissa tutkimuksissa on huomattu vaikuttavan kustannuksiin niin, että yli 65-vuotiaiden todennäköisyys kuulua kalliiden potilaiden ryhmään on muita ikäluokkia suurempi (ks. esim. Rais et al. 2013; Reid et al. 2003; Calver et al. 2006). Iän lisäksi myös tiettyjen yksittäisten tulosten ja diagnoosien voidaan aikaisempien tutkimusten tulosten perusteella olettaa selittävän korkeita kustannuksia myös tämän tutkielman aineistossa. Näitä ovat esimerkiksi mielenterveyteen ja käyttäytymisen häiriöihin liittyvät diagnoosit (esimerkiksi masennus), tuki- ja liikuntaelinsairaudet, diabetes, sydämen vajaatoiminta, keuhkoastma ja verisuonisairaudet (ks. esim. Chechulin et al. 2014; Kapiainen et al. 2010; Reid et al. 2003; Ash et al. 2001; Reho et al. 2018). Näiden lisäksi myös vanhuspalveluiden

hoitajaksojen sekä sosiaalihuollon asiakkuuden oletetaan lisäävän potilaan riskiä kuulua kalliiden potilaiden ryhmään. Tätä hypoteesia tukee Leskelän et al. (2013) paljon sosiaali- ja terveystalveluita käyttäviä potilaita koskevan tutkimuksen tulos, jonka mukaan vanhuspalveluiden ja lastensuojelun asiakkaat ovat sosiaali- ja terveystalveluiden kalleimpia asiakasryhmiä.

On huomionarvoista todeta, että tämän tutkielman otoskoon pienuus ( $N < 6000$ ) voi vaikuttaa tulosten merkitsevyyteen ja mallin ennustusvoimaan. Yksittäisiin tulosityihin, hoitajaksoihin tai diagnooseihin liittyvien tapahtumien määrä jää joissakin tapauksissa hyvin pieneksi ( $n < 100$ ), mikä todennäköisesti aiheuttaa verrattain suuria luottamusvälejä ja joissakin tapauksissa myös voi myös aiheuttaa yllättävän näköisiä korrelaatioita ja ylisuuria vetosuhteita. Näitä seikkoja sekä niiden vaikutusta regressiomallin tuloksiin pohditaan kattavammin tutkielman Pohdinta-luvussa.

### 3 Tutkimuksen tarkoitus, tavoite ja tutkimuskysymykset

Tässä rekisteriaineistoon perustuvassa poikkileikkaustutkimuksessa selvitetään, mitkä yksittäiset tekijät selittävät terveydenhuollon potilaan kuulumista kaikista kalleimpaan 10 prosenttiin potilaista. Tämän lisäksi pyritään selvittämään, lisääkö sosiaalihuollon asiakkuus potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään.

Aikaisempaan tutkimuskirjallisuuteen nojaten, käytetään kustannuksia selittävinä tekijöinä potilaiden perusterveydenhuollon tulosityitä sekä erikoissairaanhoidon diagnoositietoja, vuodeosastojaksojen lukumäärää sekä tietoja potilaiden vanhainkoti-, kotihoito-, lastensuojelu- ja vammaispalveluasiakkuuksista.

Aikaisemmat tutkimukset antavat kuvan terveystalouden epätasaisen jakautumisen universaalista ja stabiilista luonteesta, mutta tutkimuksia sosiaalihuollon kustannusten yhteydestä terveydenhuollon kustannuksiin näyttävät olevan harvassa (mm. Leskelä et al. 2013). Tämä tutkielma pyrkiikin laajentamaan tutkimuskentän tietopohjaa selvittämällä tarkemmin myös sosiaalihuollon suhdetta korkeisiin perusterveydenhuollon ja erikoissairaanhoidon kustannuksiin.

Tutkielmassa pyritään vastaamaan seuraaviin kysymyksiin:

- 1. Mitkä terveydenhuollon suoritteet (diagnoosit, tulosityt, hoitotapahtumat, vanhainkoti-, kehitysvammaisuus- tai kotihoitotapahtumat) lisäävät yksittäisen potilaan todennäköisyyttä kuulua kaikista kalleimpien potilaiden ryhmään?*
- 2. Lisääkö sosiaalipalveluiden asiakkuus (lastensuojelu-, vammais- tai muu aikuis- ja perhetyöasiakkuus) potilaan todennäköisyyttä kuulua kaikista kalleimpien potilaiden ryhmään?*

## 4 Aineisto ja menetelmät

Tässä luvussa kuvaillaan tutkimusaineistoa, joka muodostuu kolmesta eri tilastosta, perustellaan aineiston keräysvaiheessa tehtyjä valintoja, rajauksia sekä lopulliset muuttujavalinnat. Tämän jälkeen analyysissä käytetyt taustamuuttujat, luokitellut ja jatkuvat muuttujat sekä kustannusmuuttujat esitellään havainnollistavien ristiintaulukointien ja frekvenssitaulukoiden avulla. Lopuksi tarkennetaan käytettävää analyysimenetelmää, eli logistista regressioanalyysia ja sen ominaispiirteitä sekä selvennetään vetosuhteiden tulkintaa.

### 4.1 Aineiston kuvaus

Tutkielman aineisto koostuu kolmesta erillisestä tilastoaineistosta:

perusterveydenhuollon Effica-datasta, sosiaalihuollon ProConsona-datasta ja erikoissairaanhoidon laskudatasta. Kaikkien tilastoaineistojen keräyksiin hankittiin vaadittavat tutkimusluvut hyvää tutkimuseettistä käytäntöä noudattaen.

Lupaviranomaisina toimivat tutkimuksen kohteena olevan kunnan johtava virkalääkäri ja johtava sosiaalityöntekijä. Tutkimuseettisiä näkökulmia pohditaan tarkemmin luvussa 6.

Kyseessä on tilastoaineistolla toteutettu poikkileikkaustutkimus. Kaikkien aineistojen tiedot koskevat vuotta 2017, ja ne käsittävät tutkittavan kunnan koko julkisen sektorin sosiaali- ja terveydenhuollon käynnit, hoitajakset, asiakkuustiedot ja päädiagnoosit (N = 5702). Erikoissairaanhoidon aineiston osalta saatiin käyttöön myös kustannustiedot henkilöittäin, koska kyseessä oli sairaanhoitopiirin laskudata. Perusterveydenhuollon aineiston kustannustiedot ovat arvioita, jotka on yhdistetty aineistoon jälkikäteen, käyttäen hyväksi Terveyden ja hyvinvoinnin laitoksen raporttia *Terveyden- ja sosiaalihuollon yksikkökustannukset Suomessa vuonna 2011* (Kapiainen ym. 2014). Tutkimusaineisto ei sisällä lääkemääräyksiä, tutkimuksia tai toimenpiteitä, joten ne on rajattu tämän tutkimuksen ulkopuolelle. Tarkasteltavaan aineistoon on valikoitunut

kaikki ne henkilöt, jotka ovat käyttäneet alueen sosiaali-, perusterveydenhuollon tai erikoissairaanhoidon palveluita vuoden 2017 aikana, eikä sitä ole rajattu demografisten tekijöiden, kuten iän tai sukupuolen mukaan. Aineistosta rajattiin jo tutkimussuunnitelmavaiheessa pois potilaskertomukset niiden sisältämän arkaluontoisen, yksittäisen potilaan tunnistamisen mahdollistavan tiedon vuoksi.

Perusterveydenhuollon aineisto on kerätty tutkimuksen kohteena olevan kunnan potilastietojärjestelmästä Efficasta, ja se koostuu yksittäisistä järjestelmään kirjatusta suoritteista. Aineistonkeruu ulkoistettiin Tieto Oyj:lle, joka on myös Efficajärjestelmän ylläpitäjä. Sosiaalihuollon aineistona toimii CGI:n ProConsona Sosiaalityö -ohjelman asiakasrekisteri, johon on kerätty asiakkuustiedot aikuis- ja perhetyöstä, vammaispalveluista ja lastensuojelusta. Toisin kuin kaksi muuta aineistoa, erikoissairaanhoidon aineisto on laskudataa, ja se koostuu sairaanhoitopiiriin tehdyistä läheteistä aiheutuneista käynneistä, tutkimuksista ja toimenpiteistä. Vuodelta 2017 kerätty sosiaalihuollon aineisto koostuu 340 yksittäisestä asiakkaasta, joilla oli yksi tai useampi sosiaalihuollon asiakkuus -merkintä tuon vuoden aikana. Aineisto sisälsi myös tiedot lastensuojelun päätöksistä, ilmoituksista, ratkaisuksista ja tarvearvioinneista, mutta niiden päätettiin olevan tutkimuskysymyksen kannalta irrelevantteja, joten ne rajattiin lopullisesta analyysiaineistosta pois ja mukaan otettiin pelkästään asiakkuustiedot. Erikoissairaanhoidon aineistossa on 1925 erillistä henkilötunnusta ja se sisältää tiedot potilaiden sukupuolesta ja syntymävuodesta, avosairaanhoito- ja vuodeosastojaksot, päädiagnoosit (ICD10), toimenpiteet (DRG) sekä erikoisalut. Erikoissairaanhoidon aineistosta lopulliseen analyysiin sisällytettiin vain päädiagnoosit, potilaskohtaiset kustannukset ja tiedot vuodeosasto- ja avohoitojaksoista.

Perusterveydenhuollon potilastietojärjestelmästä alun perin kerätyt suoritelistaukset valittiin hyödyntäen aikaisempaa tutkimuskirjallisuutta. Diagnoositietojen lisäksi Efficasta haettiin tiedot käyntimääristä ja tulosyistä, aikaleimat, henkilötunnukset (salattuna), tulosalueet ja toimintolajit (n = 5626). Näistä potilastietojärjestelmästä haetuista suoritetiedoista päädyttiin lopullisessa analyysissä käyttämään vain muutamaa tutkimusasetelman kannalta oleellista suoritetta, joita ovat salattu henkilötunnus,

syntymävuosi, sukupuoli ja tulosyy (ICPC-koodi). Perusterveydenhuollon aineiston osalta jouduttiin tekemään päätös diagnoositietoja (ICD10) sisältävien muuttujien poisjättämisestä, sillä aineisto oli siltä osin erittäin puutteellinen, eikä diagnoositietoja ollut systemaattisesti kirjattu järjestelmään. ICD10-koodien sijaan analyysiin valittiin ICPC-koodeja sisältävät muuttujat, koska ne oli merkitty lähes jokaiseen yksittäiseen suoritteeseen. ICPC, eli perusterveydenhuollon kansainvälinen luokitus, kuvaa vastaanotokäynnin syytä, toisin sanoen tulosyytä. ICPC-luokittelun koodit muistuttavat monilta osin ICD-koodeja, ja niillä onkin yhteiset juuret. ICPC-koodisto on luotu täydentämään kansainvälistä tautiluokitusta, ja molempia luokitteluja käytetään nykyään usein yhdessä. (Kuntaliitto 2017.) ICPC-koodeja on käytetty sairastavuuden mittaamiseen myös muissa tutkimuksissa (ks. esim. Smits et al. 2009).

Taulukko 2 kuvaa lopullisen analyysiaineiston sukupuolijakauman. Naisten osuus (50 %) on hieman miesten osuutta (48,7 %) suurempi. Puuttuvia tietoja Sukupuoli-muuttujassa on yhteensä 76 kappaletta. Nämä ovat yksittäisiä terveydenhuollon potilaita tai sosiaalihuollon asiakkaita, joilla potilas- tai asiakastietojärjestelmän sukupuolikenttä on tuntemattomasta syystä jätetty tyhjäksi. Analyysiaineiston sukupuolijakauma ei kuvaa suoraan tutkittavan kunnan sukupuolijakaumaa, vaan kertoo kuinka suuri osa palveluiden käyttäjistä vuonna 2017 on ollut kunkin sukupuolen edustajia. Analyysiaineiston sukupuoli jakauma eroaa hieman tutkittavan kunnan sukupuolijakaumasta, jossa miesten osuus oli vuonna 2017 53,24 % ja naisten osuus 46,75 % (Tilastokeskus 2018b).

Taulukko 2

*Analyysiaineiston sukupuolijakauma*

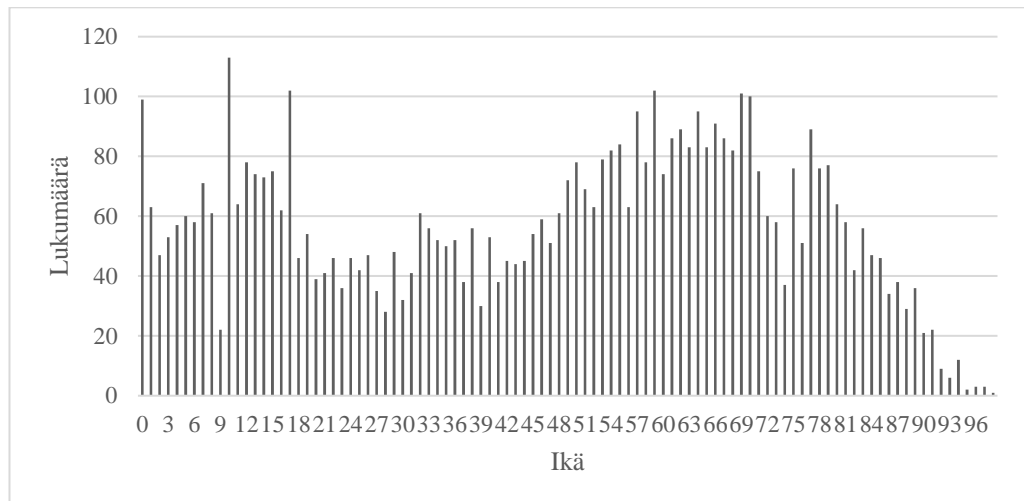
	N-luku	%
Mies	2775	48,7
Nainen	2851	50
Puuttuvat	76	1,3
Yhteensä	5702	100



Alla oleva pylväsdiagrammi (Kuvio 1) esittää analyysiaineiston ikäjakauman. Aineisto sisältää henkilöitä aina 0-vuotiaista 100-vuotiaisiin ja keski-ikä on noin 46 vuotta.

Kuvio 1

*Analyysiaineiston ikäjakauma (0-100 vuotta)*



## 4.2 Muuttujien valinta

Suurimmassa osassa tutkielman teoreettisessa viitekehyksessä esitellyistä kansainvälisistä ja suomalaisista tutkimuksista tutkittavaa ilmiötä selittävinä muuttujina käytettiin saatavuudesta riippuen tietoja potilaiden diagnooseista, käyntimääristä, lääkemääräyksistä ja potilaskohtaisista kustannuksista. Diagnoositietojen osalta tutkimuksissa käytetyt luokittelujärjestelmät poikkeavat kuitenkin toisistaan niiden yksityiskohtaisuuden osalta. (ks. Chechulin et al. 2014; Ash et al. 2001; Reid et al. 2003; Rais et al. 2013; Reho et al. 2018; Leskelä et al. 2013; Kapiainen et al. 2010.)

Chechulin ym. (2014) ovat muodostaneet ennustavan mallinsa muuttujalistauksen Maailman terveysjärjestön ICD-10- ja ICD-9-tautiluokitusten (International Statistical Classification of Diseases and Related Health Problems) pohjalta ja valinneet sen avulla

malliinsa 20 yksittäistä selittävää muuttujaa. He ovat nostaneet vertailuun ICD-pääluokkien lisäksi myös muutaman yksittäisen diagnoosin: keuhkohtaumataudin, diabeteksen sekä sydämen vajaatoiminnan, sillä nämä ovat länsimaissa yleisiä, ja täten myös oletettavasti kalliita kansantauteja (Chechulin et al. 2014). ICD-pääluokkien lisäksi monet tutkimukset ovat hyödyntäneet Diagnosis Clusters (EDC) -tautiluokitusta (ks. esim. Reid et al. 2003; Rais et al. 2013). EDC-luokittelu on erittäin kattava; siinä on 24 diagnoosiryhmää ja 114 yksittäistä diagnoosia (Province of British Columbia 2017). EDC-luokittelun on todettu aikaisemmissa tutkimuksissa olevan muihin samankaltaisiin listauksiin verrattuna erittäin tehokas työkalu terveystalouden kustannusten ennustamisessa (Brilleman et al. 2014).

Tämän tutkielman lopulliseen analyysiin valittiin 25 erillistä muuttujaa, jotka käsittävät taustamuuttujien lisäksi yksittäisiä diagnooseja, sairausryhmiä, tulosityitä sekä toimintolajeja (esim. perusterveydenhuollon kotihoidon hoitojakso). Analyysiin sisällytettiin niitä yksittäisiä diagnooseja, joita oli tutkittavassa aineistossa eniten, ja joissa oli myös eniten vaihtelua. Vertailun ulkopuolelle jätettiin kaikki diagnoosikoodit, joissa tapahtumien määrä oli alle 100. Jotkin logistisen regressioanalyysin selittävästä muuttujista on valittu suoraan tutkielman teoreettisessa viitekehysessä esitetyn tutkimuskirjallisuuden ja aikaisempien tutkimustulosten perusteella. Näitä muuttujia ovat esimerkiksi aikaisemmissa tutkimuksissa esitetyt kalliiseen potilasryhmään kuulumista ennustavat yksittäiset diagnoosit: masennus, diabetes, keuhkohtaumatauti ja sydämen vajaatoiminta (ks. esim. Chechulin et al. 2014; Kapiainen et al. 2010; Reid et al. 2003; Ash et al. 2001; Reho et al. 2018). Tämän lisäksi tutkielman regressiomallin selittäviksi muuttujiksi on otettu mukaan Suomessa yleisimpiin kansantauteihin liittyvät tulosityt ja diagnoosit, kuten verisuonitaudin riski, sepelvaltimotauti, astma, allergia, tuki- ja liikuntaelinsairaudet sekä muut mielenterveyden ja käyttäytymisen häiriöt (ks. THL 2017).

Sosiaalihuollon aineistosta jätettiin lopullisessa mallissa pois erilliset päätökset, ratkaisut, ilmoitukset, sopimukset yhteydenotot ja suunnitelmat niihin liittyvien tapahtumien pienen määrän takia ( $n < 100$ ). Erikoissairaanhoidon aineistosta rajattiin

pois toimenpidekoodit, sillä ne sisälsivät monelta osin päädiagnoosien kanssa päällekkäisiä tietoja. Analyysissa on päädytty käyttämään kahta taustamuuttujaa, jotka ovat ikä ja sukupuoli. Tutkimusaineistossa on potilaita kaikista eri ikäryhmistä, aina 0-vuotiaista 100-vuotiaisiin. Kaikki yli 100-vuotiaat on rajattu aineistosta pois, sillä ne olivat järjestelmiin automaattisesti luotuja testihenkilöitä. Ikä-muuttujaa ei luokitella erikseen, vaan se on jatkuva muuttuja. Aineistojen yhdistämisvaiheessa päädyttiin käyttämään suurimman osa-aineiston, eli perusterveydenhuollon aineiston Ikä-muuttujaa. Tästä syystä niillä henkilöillä, jotka ovat puuttuneet alkuperäisestä Effica-aineistosta, ei ikätietoa välttämättä lopullisessa yhdistetyssä aineistossa ole. Taustamuuttujista Sukupuoli-muuttuja on koodattu kaksiluokkaiseksi muuttujaksi, jossa mies saa arvon 1 ja nainen arvon 0.

### 4.3 Muuttujien esittely

Taulukko 3

*Regressiomallin jatkuvien muuttujien nimet ja koodit*

<b>Muuttuja</b>	<b>Koodit</b>
Ikä	0-100
Sydämen vajaatoimintaan liittyvä perusterveydenhuollon käynti	K77 (ICPC1)
Eteisvärinään liittyvä perusterveydenhuollon käynti	K78 ja K80 (ICPC1)
Verisuonitaudin riskitekijöihin liittyvä perusterveydenhuollon käynti	K22 (ICPC1)
Masennukseen liittyvä perusterveydenhuollon käynti	P76 (ICPC1)
Muihin mielenterveyden häiriöihin liittyvä perusterveydenhuollon käynti	P01-P99 (ICPC1, pois lukien P76)
Tuki- ja liikuntaelinsairauksiin liittyvä perusterveydenhuollon käynti	L01-L99 (ICPC1)
Perusterveydenhuollon vanhainkotipalveluiden tapahtuma	53 Vanhainkoti
Perusterveydenhuollon kehitysvammahoidon tapahtuma	70 Kehitysvammahoito
Perusterveydenhuollon vuodeosastojakso	09 Vuodeosasto
Perusterveydenhuollon kotihoidon tapahtuma	22 Kotihoito

Taulukossa 3 on listattu logistisen regressiomallin jatkuvat muuttujat ja niihin yhdistetyt perusterveydenhuollon tulosity-koodit (ICPC), hoitajaksot ja hoitotapahtumat. Jatkuvia muuttujia ovat kaikki ne muuttujat, joissa verrattain paljon vaihtelua, ja ne voivat saada minkä tahansa arvon esimerkiksi lukujen 0 ja 100 välillä (esim. Ikä-muuttuja). Tästä syystä niihin lukeutuu pelkästään rajattu määrä perusterveydenhuollon aineiston muuttujia, eikä yhtään sosiaalihuollon tai erikoissairaanhoidon muuttujaa.

#### Taulukko 4

##### *Regressiomallin luokiteltujen muuttujien nimet, koodit ja luokittelut*

<b>Muuttuja</b>	<b>Koodit</b>	<b>Luokittelu</b>
Sukupuoli	M, N	1, 0 (1 = mies, 0 = nainen)
Astmaan tai allergiaan liittyvä perusterveydenhuollon käynti	R96, R97 A92 (ICPC1)	1, 0*
Diabetekseen liittyvä perusterveydenhuollon käynti	T90 (ICPC1)	1, 0
Keuhkohtaumatautiin liittyvä perusterveydenhuollon käynti	R95 (ICPC1)	1, 0
Erikoissairaanhoidon diabetesdiagnoosi	E10.2–E13.9 (ICD10)	1, 0**
Erikoissairaanhoidon masennusdiagnoosi	F31.31, F32.00–F33.9, F41.2, F43.21 (ICD10)	1, 0
Erikoissairaanhoidon sydämen vajaatoiminta -diagnoosi	I50.1–I50.9 (ICD10)	1, 0
Erikoissairaanhoidon sepelvaltimotautidiagnoosi	I21.01–I25.9 (ICD10)	1, 0
Erikoissairaanhoidon keuhkohtaumatautidiagnoosi	J44.0–J44.8 (ICD10)	1, 0
Erikoissairaanhoidon eteisvärinädiagnoosi	I48 (ICD10)	1, 0
Erikoissairaanhoidon mielenterveyden ja käyttäytymisen häiriöt -pääluokkaan kuuluva diagnoosi	F00-F99 (ICD10)	1, 0
Erikoissairaanhoidon tuki- ja liikuntaelinsairaudet -pääluokkaan kuuluva diagnoosi	M00–M99 (ICD10)	1, 0
Erikoissairaanhoidon vuodeosastojakso	VOS	1, 0***
Sosiaalihuollon asiakkuus	Lastensuojelu-, vammaispalvelu-, muu aikuis- ja perhetyö-, vammaisten kuljetuspalvelu-, iäkkäiden palvelu- ja kuntouttavan työtoiminnan asiakkuus	1, 0****

\* 1 = vähintään yksi käynti, 0 = ei käyntiä

\*\* 1 = vähintään yksi diagnoosimerkintä, 0 = ei diagnoosimerkintää

\*\*\* 1 = vähintään yksi vuodeosastojakso, 0 = ei vuodeosastojaksoja

\*\*\*\* 1 = vähintään yksi sosiaalihuollon asiakkuusmerkintä, 0 = ei asiakkuusmerkintää

Analyysiin valikoituneet luokitellut muuttujat on esitelty Taulukossa 4. Kaikki luokitellut muuttujat ovat dikotomisissa, eli kaksiluokkaisia muuttujia, ja ne voivat saada joko arvon 0 tai 1. Luokitelluiksi muuttujiksi muokattiin kaikki ne muuttujat, joiden esiintyvyydessä ei ollut merkittävästi vaihtelua. Muuttujasta riippuen vaihtelun pieni määrä johtuu joko aineiston suhteellisen pienestä koosta (ja tapahtumien pienestä n-luvusta) tai siitä, ettei siinä aidosti ole suurta vaihtelua. Diagnoosimuuttujien koodaamista kaksiluokkaisiksi muuttujiksi puoltaa myös se, ettei esimerkiksi pitkäaikaissairauksissa saman sairauden diagnoosimerkintöjen lukumäärä yhdellä ihmisellä välttämättä tarkoita, että sairaus olisi todettu samalla henkilöllä monta kertaa.

Erikoissairaanhoidon aineiston muuttujien kustannustiedot saatiin suoraan laskuaineistosta, joten erillisiä kustannusmuuttujia ei tarvinnut niiden osalta analyysiaineistoon koodata. Erikoissairaanhoidon aineiston kustannustiedot on merkitty euron tarkkuudella suoritekohtaisesti ja tutkimusaineistossa jokaiselle henkilölle on muodostunut oma kustannustieto, jossa kaikkien potilaiden suoritteiden hinnat on laskettu yhteen. Niillä tutkimusaineiston henkilöillä, joilla ei ole vuoden 2017 aikana ollut erikoissairaanhoidon avohoidon tai vuodeosaston hoitajaksoja tai muita suoritteita, on erikoissairaanhoidon kustannukseksi merkitty 0 euroa.

Perusterveydenhuollon hintatietojen määrittelyn apuna on käytetty Terveystieteiden ja hyvinvoinnin laitoksen yksikkökustannusraporttia (Kapiainen et al. 2014). Hintatiedot on koodattu aineistoon niiden muuttujien (tulosityt, sairausryhmät ja hoitajakset) osalta, jotka olivat suoraan löydettävissä yksikkökustannusraportilta. Analyysia varten muodostetut kustannusmuuttujat on listattu alla olevassa taulukossa (Taulukko 5).

Taulukko 5

*Perusterveydenhuollon kustannusmuuttujalistaus ja arvioidut kustannustiedot per  
diagnoosi- tai käyntimerkintä.*

Kustannusmuuttuja	Yksikkökustannusraportin määritelmä	Arvioitu kustannus per diagnoosi tai käynti (€)
PTH_diabetes_hinta	Diabetes	52
PTH_raskaus_hinta	Raskaus	62
PTH_masennus_hinta	Depressio	83
PTH_eteisvärinä_hinta	Sydämen rytmihäiriöt	66
PTH_tule_hinta	Muu tuki- ja liikuntaelinsairaus	66
PTH_verenpainetauti_hinta	Verenpainetauti	39
PTH_muut_mielenterveyden_häiriö_hinta	Myy psyyken sairaus tai ongelma	76
PTH_sydämen_vajaatoiminta_hinta	Sydämen vajaatoiminta tai muu sydänsairaus	54
PTH_huumeidenkäyttö_hinta	Päihteiden väärinkäyttö	78
PTH_astma_tai_allergia_hinta	Allergiaoireet	57
PTH_kasvaimet_ja_syöväet_hinta	Pahanlaatuinen kasvain	54
PTH_kehitysvammahuolto_hinta	Kehitysvamma tai epämuodostumat	55
PTH_keuhkohtaumatauti_hinta	Krooninen keuhkosairaus	62
PTH_kotihoito_hinta	Kotihoito (ka)	40,5
PTH_vuodeosasto_hinta	Perusterveydenhuollon vuodeosasto, keskimääräinen e/hoitopäivä	213
PTH_vanhainkoti_hinta	Vanhainkoti, hoitovuorokausi, osto	154
PTH_fysioterapia_hinta	Fysioterapiakäynnit (ka)	114
PTH_kilpirauhanen_hinta	Muu endokrinologinen sairaus	34
PTH_dementia_ja_muistihäiriö_hinta	Dementia tai vanhuudenheikkous	91
PTH_mielenterveysneuvola_hinta	Avohoidon mielenterveysyksikkö (ka)	73
PTH_muut_neuvolapalvelut_hinta	Terveystieteiden vastuu	47

*ka = aritmeettinen keskiarvo (esim. eripituisten vastaanottoaikojen tai eri ammattiryhmien vastaanottojen kustannusten keskiarvo)*

Kaikkiin aineiston muuttujiin ei pystytty yhdistämään kustannustietoja, sillä niille ei löytynyt suoraa vastinetta lähteenä käytettävästä yksikkökustannusraportista. Näiden muuttujien hintatiedoksi päädyttiin merkitsemään 0 euroa. Joidenkin muuttujien osalta päädyttiin käyttämään kahden tai useamman kustannustiedon aritmeettista keskiarvoa, esimerkiksi eripituisten vastaanottoaikojen tai eri ammattiryhmien vastaanotokustannusten keskiarvoa. Onkin huomionarvoista mainita, että tässä tutkielmassa ei tutkita sosiaali- ja terveydenhuollon kokonaiskustannuksia, eivätkä tutkimuksen kohteena ole potilaskohtaiset kustannukset *per se*, vaan niiden suhteet. Voidaan myös olettaa, että kustannukset noudattavat kutakuinkin samaa jakaumaa kuin jos käytettävissä olisi tiedot kokonaiskustannuksista, ja että korkeita potilaskohtaisia sotekustannuksia ennustavat tekijät saadaan selville myös olemassa olevan tutkimusasetelman avulla. Vertailun vuoksi tämän tutkielman luvussa 5 esitetyt kustannuslaskelmat on laskettu sekä puuttuvien tietojen kanssa, että korvaamalla puuttuvat kustannustiedot vuoden 2017 perusterveydenhuollon asukaskohtaisten nettokustannusten keskiarvolla (Tilastokeskus 2018c).

Kun kaikille mahdollisille perusterveydenhuollon muuttujille oli luotu tutkimusaineistoon erillinen kustannusmuuttuja, laskettiin nämä uudet muuttujat yhteen yhdessä erikoissairaanhoidon kustannustietojen kanssa. Tästä muodostui uusi Yhdistettykustannus-muuttuja, jonka avulla myöhemmin muodostettiin regressiomallin selitettävä Kallispotilas-muuttuja. Alla Taulukko 6, josta nähdään Yhdistettykustannus-muuttujan kuvailevat tunnusluvut.

Taulukko 6

*Yhdistettykustannus-muuttuja*

<b>Muuttuja</b>	<b>N-luku</b>	<b>Keskiarvo (€)</b>	<b>Mediaani (€)</b>	<b>Min (€)</b>	<b>Max (€)</b>
Kustannukset/hlö	5702	1632,2333	66	0*	130630,5

\* puuttuvat arvot muutettu nolaksi (0€)

Analyysin selitettävänä tekijänä toimii Kallispotilas-muuttuja, johon kuuluu kallein 10 prosenttia potilaista. Kaikki yli 1740 euron vuotuisen hinnan saaneet saivat arvon 1, ja muut potilaat arvon 0. Lopulliseen kalliiden potilaiden tutkittavaan ryhmään (muuttujan arvo 1) kuului yhteensä 568 henkilöä. Tätä muuttujaa käytettiin regressiomallissa selitettävänä tekijänä, kun tarkoituksena oli selittää, mitkä yksittäiset suoritteet (diagnoosit, käyntimäärät, vuodeosastojaksot, sairausryhmät ym.) vaikuttavat potilaan todennäköisyyteen kuulua kalliiden potilaiden ryhmään.

Taulukot 7 ja 8 näyttävät Kallispotilas-muuttujan kuvailevia tunnuslukuja sekä muuttujan sukupuolijakauman. Jakaumasta voidaan huomata, että naisten suhteellinen osuus Kallispotilas-muuttujassa on hieman miehiä suurempi. Sukupuolijakauman Puuttuvat-sarakkeessa on ne henkilöt (yhteensä 76 kpl), joilta sukukupuolimerkintä puuttui kokonaan yhdistetystä aineistosta. Taulukko 9 kuvaa Kallispotilas-ryhmään kuuluvien ja Muut-ryhmään kuuluvien potilaiden keski-ikä. Kalliiden potilaiden keski-ikä on yli 69 vuotta ja muiden potilaiden keski-ikä noin 43 vuotta.

Taulukko 7

*Kallispotilas-muuttuja*

Muuttuja	Muuttujan arvo	N-luku	%
Kallispotilas	<i>Ei (0)</i>	5134	90
	<i>Kyllä (1)</i>	568	10

Taulukko 8

*Ristiintaulukointi: Kallispotilas ja Sukupuoli*

Muuttuja	Muuttujan arvo	Puuttuva		Mies		Nainen	
		N-luku	%	N-luku	%	N-luku	%
Kallispotilas	<i>Ei (0)</i>	76	1,5	2520	49,1	2538	49,40
	<i>Kyllä (1)</i>	0	0	255	44,9	313	55,1



## Taulukko 9

*Ristiintaulukointi: Kallispotilas ja Ikä*

Muuttuja	Muuttujan arvo	Ikä		
		Keskiarvo	Min	Max
Kallispotilas	<i>Ei (0)</i>	43,23	0	97
	<i>Kyllä (1)</i>	69,87	13	99

### 4.4 Regressiomalli ja vetosuhteet

Tämä tutkielma on tilastoaineiston avulla tehty poikkileikkaustutkimus.

Analyysimenetelmänä käytetään logistista regressioanalyysia, jota voidaan käyttää sekä todellisuuden mallintamiseen että havaintojen ennustamiseen. Logistisessa regressiossa selitettävä muuttuja on aina kaksiluokkainen, eli saa joko arvon 1 tai 0, ja se sopii parhaiten sellaiseen tutkimusasetelmaan, jossa koitetaan selvittää, tapahtuuko joku asia vai ei. Logistinen regressioanalyysi on lääketieteessä yleisesti käytetty analyysimenetelmä. (Metsämuuronen 2006, s. 670; Bland & Altman 2000.) Esimerkiksi suurin osa tutkielman teoreettisessa viitekehyksessä (ks. Taulukko 1) esitellyistä tutkimuksista on toteutettu logistisen regressiomallin avulla (ks. Chechulin et al. 2014; Fleishman & Cohen 2010; Ash et al. 2001; Powers et al. 2005; Kapiainen et al. 2010; Leskelä et al. 2014; Zhao et al. 2005). Tässä tutkielmassa regressiomallin selitettävänä muuttujana, eli sinä muuttujana, jota halutaan ennustaa, toimii Kallispotilas-muuttuja, jossa kaikista kalleimpaan kymmenykseen kuuluvat potilaat saavan arvon 1 ja muut 0. Regressioanalyysin tulokset kertovat sen, vaikuttavatko valitut selittävät muuttujat ennustettavan tapahtuman todennäköisyyteen, ja jos vaikuttavat niin kuinka paljon. Luottamusvälinä käytetään 95 %:n luottamustasoa.

Regressiomallin tulokset ilmaistaan vetosuhteina. Vetosuhde on todennäköisyyksien (tai osuuksien) etäisyyden mitta ja sillä voidaan kuvata, kuinka etäällä kaksi todennäköisyyttä ovat toisistaan (Rita 2004). Vetosuhde kertoo, mihin suuntaan tapahtuman todennäköisyys muuttuu selittävän muuttujan muuttuessa yhden yksikön

verran. Veto on aina positiivinen luku, eikä se kuvaa suoraan todennäköisyyksien tai osuuksien suhdetta, vaan vetojen suhteen. Vetosuhteen ollessa alle 1, on selittävän ja selitettävän muuttujan korrelaatio negatiivinen. Kun vetosuhte on puolestaan yli 1, vaikuttaa selittävä muuttuja tapahtuman todennäköisyyteen positiivisesti. Selittävät muuttujat voivat olla joko jatkuvia tai luokiteltuja. Logistisessa regressiossa selitettävän ja selittävän muuttujan suhteen oletetaan noudattavan logaritmista asteikkoa, eli kasvavan niin sanotusti korkoa korolle. (Bland & Altman 2000; Szumilas 2010; Rita 2004.) Tämän tutkielman analyysissä selittävinä muuttujina toimivat diagnoosit, tulokset, sosiaalihuollon asiakkuus ja hoitajaksot, ja ne sisältävät sekä luokiteltuja että jatkuvia muuttujia.

Muuttujien valinnassa käytettiin askeltavan menetelmän ja pakotetun mallin yhdistelmää. Askeltavassa menetelmässä muuttujia lisätään tai poistetaan mallista mallin selitysasteen muutoksen perusteella. Pakottavassa mallissa selittävät muuttujat valitaan teorian tai aikaisemman tutkimuksen perusteella ja koitetaan selvittää, miten tietyt muuttujat selittävät tutkimuksen kohteena olevaa ilmiötä. Yhtäältä tutkimusasetelma on konfirmatorinen, eli siinä selvitetään jo entuudestaan tunnettuja mielenkiinnon kohteena olevaa ilmiötä selittäviä tekijöitä ja niiden vetosuhteita, ja toisaalta pyritään löytämään uusia yhteyksiä. (Metsämuuronen 2006, s. 670–676.)

Logistisessa regressiomallissa havaintojen määrän tulee olla kohtuullinen malliin mukaan otettavien muuttujien lukumäärään nähden, sillä muuten muuttujilla ei välttämättä ole risteäviä havaintoja, tai mallin selitysaste nousee teknisistä syistä liian korkeaksi (Metsämuuronen 2006, s. 672). Tämän tutkielman lopullisen regressioanalyysin muuttujalistauksen ulkopuolelle on tästä syystä jätetty kaikki ne yksittäiset diagnoosit, tulokset, tapahtumat, hoitajaksot ja asiakkuustiedot, joiden tapahtumien lukumäärä on ollut alle 100.

## 5 Analyysi ja tulokset

Tässä luvussa kuvaillaan regressiomallin muodostaminen ja analyysin tulokset. Ensimmäisessä alaluvussa esitellään mallin selittävien muuttujien sekä selitettävän Kallispotilas-muuttujan tilastollisia tunnuslukuja havainnollistavien frekvenssitaulukoiden avulla. Toisessa alaluvussa kuvataan regressiomallin muodostamisen eri vaiheita ja esitetään lopullinen regressiomalli, sen tulokset ja mallin ennustusvoima.

### 5.1 Kuvailevaa tarkastelua

Taulukossa 10 kuvataan logistisen regressiomallin otos (N=5621) ja puuttuvien tietojen määrä (1,4 %). Taulukko 11 kuvaa malliin sisällytettyjen jatkuvien muuttujien kuvailevia tunnuslukuja. Mallissa on mukana potilaita aina vastasyntyneistä 99-vuotiaisiin ja koko analyysiaineiston keski-ikä on noin 46 vuotta. Jatkuvista muuttujista selvästi eniten vaihtelua on kolmessa perusterveydenhuollon muuttujassa: vanhainkodin tapahtumissa, kehitysvammahuollon tapahtumissa ja kotihoidon tapahtumissa. Vaihtelun suuri määrä tarkoittaa tässä tapauksessa sitä, että joillakin analyysiaineiston henkilöillä ei näitä tapahtumia ole ollenkaan, kun taas joillakin tapahtumien lukumäärä voi liikkua kymmenissä, sadoissa tai jopa tuhansissa (esim. perusterveydenhuollon vanhainkotitapahtumat). Näin suuri määrä on seurausta tilanteista, joissa hoidontarve on päivittäinen ja jatkuva. Huomionarvoista on se, että kaikki tapahtumamerkinnot eivät välttämättä ole fyysisiä hoitajan tai lääkärin koti- tai vastaanottokäyntejä vaan ne sisältävät myös puhelinkonsultaatiot ja asiakirjamerkinnät, joita tehdään joissakin tapauksissa myös ilman potilaan tapaamista.

Taulukko 10

*Regressiomallin otoskoko ja puuttuvat tiedot*

	<b>N-luku</b>	<b>%</b>
Sisällytetty analyysiin	5621	98,6
Puuttuvat	81	1,4
<b>Yhteensä</b>	<b>5702</b>	<b>100</b>

Taulukko 11

*Frekvenssitaulukko: analyysiaineiston jatkuvat muuttujat ja niiden kuvailevat tunnusluvut*

<b>Muuttuja</b>	<b>Keskiarvo</b>	<b>Keskihajonta</b>	<b>Min</b>	<b>Max</b>
Ikä	46,09	26,13	0	99
Sydämen vajaatoimintaan liittyvä perusterveydenhuollon käynti	0,006	0,11	0	4
Eteisvärinään liittyvä perusterveydenhuollon käynti	0,2055	1,99	0	37
Verisuonitaudin riskitekijöihin liittyvä perusterveydenhuollon käynti	0,3816	2,18	0	38
Masennukseen liittyvä perusterveydenhuollon käynti	0,0677	0,74	0	17
Muihin mielenterveyden häiriöihin liittyvä perusterveydenhuollon käynti	0,1132	1,01	0	30
Tuki- ja liikuntaelinsairauksiin liittyvä perusterveydenhuollon käynti	0,8877	2,73	0	53
Perusterveydenhuollon vanhainkotipalveluiden tapahtuma	0,403	5,47	0	124
Perusterveydenhuollon kehitysvammahoidon tapahtuma	0,363	7,71	0	231
Perusterveydenhuollon vuodeosastojakso	0,4648	2,31	0	52
Perusterveydenhuollon kotihoidon tapahtuma	23,0626	160,92	0	3077

Taulukossa 12 kuvataan mallin luokitellut muuttujat ja niiden frekvenssit ja prosenttiosuudet muuttujien eri arvoilla (0 ja 1). Taulukkoa lukemalla voidaan havaita, että analyysissä mukana olevista potilaista 76:lle ei ole merkitty tietoa sukupuolesta, mikä näkyy puuttuvana tietona. Muutamaa poikkeusta lukuun ottamatta taulukkoon merkittyjen diagnoosien, tulosten ja hoitotapahtumien esiintyvyydessä ei ole suuria sukupuolten välisiä eroja. Perusterveydenhuollon tulostyistä vain keuhkohtaumatautiin liittyvissä käynneissä sukupuoliero on huomattava; jopa 65 prosenttia käynneistä on miespuolisten tutkimuskohteiden käyntejä. Samansuuntainen jakauma on nähtävissä myös erikoissairaanhoidon keuhkohtaumatautidiagnooseissa, joissa miesten osuus on 62,5 %. Yleisempiä erikoissairaanhoidon diagnooseja ovat miehillä myös mielenterveyden ja käyttäytymisen häiriöihin liittyvät diagnoosit (61,5 %), diabetesdiagnoosit (62,8 %) ja sepelvaltimotautidiagnoosit (67,2 %). Naisilla yleisempiä diagnoosimerkintöjä ovat puolestaan sydämen vajaatoiminnan (60 %) ja tuki- ja liikuntaelinsairauksien (60 %) diagnoosit. Monissa luokitelluissa muuttujissa tapahtumien määrä kokonaisuudessaan on kuitenkin verrattain pieni, joten niistä voidaan johtaa yleistettävissä olevia johtopäätöksiä.

Taulukko 12

*Frekvenssitaulukko: analyysiaineiston luokitellut muuttujat ja niiden kuvailevat tunnusluvut*

*(1 = vähintään yksi käynti/diagnoosimerkintä, 0 = ei käyntiä/diagnoosimerkintää)*

Muuttuja	Muuttujan arvo	Mies N	Mies %	Nainen N	Nainen %	Ei arvoa N	Ei arvoa %
Astmaan tai allergiaan liittyvä perusterveydenhuollon käynti	1	64	45,7%	76	54,3%	0	0,0%
	0	2711	48,7%	2775	49,9%	76	1,4%
Diabetekseen liittyvä perusterveydenhuollon käynti	1	171	51,7%	160	48,3%	0	0,0%
	0	2604	48,5%	2691	50,1%	76	1,4%
Keuhkohtaumatautiin liittyvä perusterveydenhuollon käynti	1	13	65,0%	7	35,0%	0	0,0%
	0	2762	48,6%	2844	50,1%	76	1,3%
Verenpainetautiin liittyvä perusterveydenhuollon käynti	1	60	44,4%	75	55,6%	0	0,0%
	0	2715	48,8%	2776	49,9%	76	1,4%
Erikoissairaanhoidon diabetesdiagnoosi	1	27	62,8%	16	37,2%	0	0,0%
	0	2748	48,6%	2835	50,1%	76	1,3%
Erikoissairaanhoidon masennusdiagnoosi	1	15	50,0%	13	43,3%	2	6,7%
	0	2760	48,7%	2838	50,0%	74	1,3%
Erikoissairaanhoidon sydämen vajaatoiminta -diagnoosi	1	12	40,0%	18	60,0%	0	0,0%
	0	2763	48,7%	2833	49,9%	76	1,3%
Erikoissairaanhoidon sepelvaltimotautidiagnoosi	1	39	67,2%	19	32,8%	0	0,0%
	0	2736	48,5%	2832	50,2%	76	1,3%
Erikoissairaanhoidon keuhkohtaumatautidiagnoosi	1	5	62,5%	3	37,5%	0	0,0%
	0	2770	48,6%	2848	50,0%	76	1,3%
Erikoissairaanhoidon eteisvärinädiagnoosi	1	23	51,1%	20	44,4%	2	4,4%
	0	2752	48,6%	2831	50,0%	74	1,3%
Erikoissairaanhoidon mielenterveyden ja käyttäytymisen häiriöt -pääluokkaan kuuluva diagnoosi	1	56	61,5%	30	33,0%	5	5,5%
	0	2719	48,5%	2821	50,3%	71	1,3%
Erikoissairaanhoidon tuki- ja liikuntaelinsairaudet -pääluokkaan kuuluva diagnoosi	1	89	38,7%	138	60,0%	3	1,3%
	0	2686	49,1%	2713	49,6%	73	1,3%
Erikoissairaanhoidon vuodeosastojakso	1	253	48,2%	265	50,5%	7	1,3%
	0	2522	48,7%	2586	50,0%	69	1,3%
Sosiaalihuollon asiakkuus	1	174	59,2%	120	40,8%	0	0,0%
	0	2601	48,1%	2731	50,5%	76	1,4%

Taulukko 13 kuvaa Kallispotilas-ryhmään kuuluvien terveydenhuollon potilaiden tai sosiaalipalveluiden asiakkaiden aikaansaaminen käyntien prosentuaalisen osuuden kaikista käynneistä regressiomallin selittävien muuttujien osalta. Kallispotilas-ryhmä koostuu niistä 10 prosentista tutkittavan kunnan sosiaali- ja terveystaloukselta vuonna 2017 käyttäneistä potilaista, jotka aiheuttavat eniten terveydenhuollon kustannuksia per henkilö. Kuten taulukosta voidaan huomata, aikaansaa tämä ryhmä monien muuttujien osalta määräänsä suuremman osuuden kaikista käynneistä. Taulukkoon listatuista muuttujista ainoastaan astmaan tai allergiaan sekä verenpainetautiin liittyvien perusterveydenhuollon käynneissä kaikista kallein potilaiden kymmenyksen osuus kaikista käynneistä on noin 10 %. Kaikkien muiden muuttujien osalta tämän potilasryhmän osuus kaikista käynneistä on määräänsä nähden vähintään noin kaksinkertainen.

Perusterveydenhuollossa kallein kymmenys aikaansaa kehitysvammahoidon tapahtumista 54,8 %, kotihoidon tapahtumista 67,3 %, vuodeostastajaksista 52,4 % sekä sydämen vajaatoimintaan, masennukseen ja muihin mielenterveyden ja käyttäytymisen häiriöihin liittyvistä käynneistä hieman alle 50 %.

Perusterveydenhuollon vanhainkodin hoitajaksoista jopa 86,3 prosenttia on potilaiden kalleimman kymmenyksen aikaansaamia. Erikoissairaanhoidossa sydämen vajaatoiminnan diagnooseista kalleimpien potilaiden osuus on 66,7 %, keuhkohtaumatauti-diagnooseista 62,5 % ja masennusdiagnooseista 40 %. Muiden erikoissairaanhoidon muuttujien osalta luvut ovat hieman matalampia, esimerkiksi vuodeostastajaksista ja sepelvaltimotautidiagnooseista kalleimpien potilaiden osuus on noin 34 %. Sosiaalihuollon asiakkuuksista kallein potilaiden kymmenys vastasi noin 28 prosentista.

Taulukko 13

*Selitettävän Kallispotilas-muuttujan prosenttiosuudet selittävistä muuttujista*

Muuttuja	Kallispotilas	
	n	% kaikista
Verisuonitaudin riskitekijöihin liittyvä perusterveydenhuollon käynti	99	21,90 %
Diabetekseen liittyvä perusterveydenhuollon käynti	82	24,80 %
Keuhkohtaumatautiin liittyvä perusterveydenhuollon käynti	4	20,00 %
Sydämen vajaatoimintaan liittyvä perusterveydenhuollon käynti	10	45,50 %
Masennukseen liittyvä perusterveydenhuollon käynti	37	48,70 %
Muihin mielenterveyden häiriöihin liittyvä perusterveydenhuollon käynti	69	47,30 %
Astmaan tai allergiaan liittyvä perusterveydenhuollon käynti	15	10,70 %
Verenpainetautiin liittyvä perusterveydenhuollon käynti	13	9,60 %
Tuki- ja liikuntaelinsairauksiin liittyvä perusterveydenhuollon käynti	259	18,50 %
Eteisvärinä liittyvä perusterveydenhuollon käynti	55	36,70 %
Perusterveydenhuollon kehitysvammahoidon jakso	17	54,80 %
Perusterveydenhuollon kotihoidon jakso	323	67,30 %
Perusterveydenhuollon vanhainkodin jakso	82	86,30 %
Perusterveydenhuollon vuodeosastojakso	320	52,40 %
Erikoissairaanhoidon vuodeosastojakso	180	34,30 %
Erikoissairaanhoidon diabetesdiagnoosi	14	32,60 %
Erikoissairaanhoidon masennusdiagnoosi	12	40,00 %
Erikoissairaanhoidon sydämen vajaatoiminta -diagnoosi	20	66,70 %
Erikoissairaanhoidon mielenterveyden ja käyttäytymiset - pääluokkaan kuuluva diagnoosi	21	23,10 %
Erikoissairaanhoidon keuhkohtaumatautidiagnoosi	5	62,50 %
Erikoissairaanhoidon sepelvaltimotautidiagnoosi	20	34,50 %
Erikoissairaanhoidon eteisvärinädiagnoosi	13	28,90 %
Erikoissairaanhoidon tuki- ja liikuntaelinsairaudet - pääluokkaan kuuluva diagnoosi	58	25,20 %
Sosiaalihuollon asiakkuus	81	27,60 %



Taulukkoon 14 on laskettu kustannustiedot Kallispotilas-muuttujan perusteella. Analyysiaineiston mukaiset yhteenlasketut kustannukset olivat tutkittavassa kunnassa vuonna 2017 yli 9 miljoonaa euroa. Näistä kaikista kalleimpaan potilaiden kymmenykseen kuuluvat aikaansaivat yli 8 miljoonaa euroa, eli lähes 90 %. Nämä luvut eivät kuitenkaan kerro koko totuutta, sillä kaiken kaikkiaan 2043:lta aineistossa mukana olleelta yksittäiseltä henkilöltä puuttui kustannustieto kokonaan. Kallispotilas-muuttujan sekä kustannuslaskelmien muodostamista varten nämä puuttuvat tiedot korvattiin hinnalla 0 €. Mikäli nuo puuttuvat tiedot korvattaisiin Tilastokeskuksen vuoden 2017 Kuntataloustilastossa raportoidulla keskimääräisellä perusterveydenhuollon nettokustannuksella (593 €/asukas), olisi Kallispotilas-ryhmän edustajien kustannusten suhteellinen osuus 79,08 %, eli se tippuisi noin kymmenen prosenttiyksikköä analyysiaineiston tietoihin verrattuna (Tilastokeskus 2018c).

Taulukko 14

*Kustannustaulukko: Kallispotilas-ryhmään kuuluvien potilaiden kustannukset vuonna 2017 euroina ja prosenttiosuuksina (puuttuvat tiedot korvattu summalla 0 €).*

Muuttuja	Muuttujan arvo	€	%
Kallispotilas	<i>Ei (0)</i>	998 908	10,66
	<i>Kyllä (1)</i>	8 318 086	89,37
	<b>Yhteensä</b>	<b>9 306 994</b>	<b>100</b>

*1 = Kallispotilas-ryhmään kuuluvat*

*0 = muut*

## 5.2 Mallin muodostaminen ja tulokset

Taulukossa 15 esitelty lopullinen logistinen regressiomalli muodostuu Ikä- ja Sukupuoli- muuttujien lisäksi 23 erillisestä tulosyy-, diagnoosi- ja hoitojaksomuuttujasta. Muuttujista iän lisäksi sydämen vajaatoimintaan, eteisvärinä, verisuonitaudin riskitekijöihin, masennukseen, muihin mielenterveyden ja käyttäytymisen häiriöihin, tuki- ja liikuntaelinsairauksiin sekä vanhainkoti-, kotihoito-, vuodeosasto- ja kehitysvammapalveluiden hoitojaksoihin liittyvät

perusterveydenhuollon käynnit on ilmaistu mallissa jatkuvina muuttujina. Kaikki loput muuttujat ovat kaksiluokkaisia muuttujia, ja tähän liittyneet valinnat on perusteltu tarkemmin luvussa 4.

Regressiomallissa sukupuoli ja ikä sekä tulositytät, diagnooseja ja hoitajaksoja kuvaavat muuttujat ovat selittäviä muuttujia, kun taas selitettävänä muuttujana toimii Kallispotilas-muuttuja, jossa 10 % kaikista kalleimmista potilaista saavat arvon 1 ja muut arvon 0. Mallin tulokset on ilmaistu vetosuhteina (*OR*), jotka kertovat lisääkö vai laskeeko kukin muuttuja potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään. Malli näyttää jokaisen muuttujan vetosuhteen, kun kaikki muut muuttujat on vakioitu. Raportoidun vetosuhteen perässä oleva tähti (\*) ilmaisee suhteen tilastollisen merkitsevyyden ( $p < 0,05$ ). Vetosuhteiden lisäksi taulukkoon on kirjattu jokaisen muuttujan kohdalle myös muuttujan vaihtelu 95 prosentin luottamusvälillä. Lopullinen regressiomalli on hyvin ennustusvoimainen (97,7 %).

Selittäivistä muuttujista kahdellatoista on mallin mukaan tilastollisesti merkitsevä yhteys Kallispotilas-muuttujan kanssa. Vetosuhdetta tulkitsemalla voidaan esimerkiksi sanoa, että diabetekseen liittyvä perusterveydenhuollon käynti lisää potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään ( $OR = 8,429$ ). Potilaalla, jolla on vuoden 2017 aikana ollut yksi tai useampi diabetekseen liittyvä perusterveydenhuollon käynti, veto on yli kahdeksankertainen verrattuna potilaaseen, jolla ei ole aineistossa merkintää diabetekseen liittyvästä perusterveydenhuollon käynnistä. Myös erikoissairaanhoidon diabetesdiagnoosi näyttäisi vetosuhteen perusteella kasvattavan potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään, mutta ero ei ole tilastollisesti merkitsevä ( $p > 0,05$ ).

Taulukko 15

*Logistinen regressioanalyysi, jossa selitettävänä muuttujana potilaan todennäköisyys kuulua Kallispotilas -ryhmään, kun kukin selittävä muuttuja muuttuu yhden yksikön verran (N=5621)*

<b>Muuttuja</b>	<b>OR</b>	<b>95 % luottamusväli</b>
Sukupuoli (L)	0,708	0,473–1,059
Ikä (J)	1,002	0,991–1,013
Diabetekseen liittyvä perusterveydenhuollon käynti (L)	8,429*	4,967–14,305
Sydämen vajaatoimintaan liittyvä perusterveydenhuollon käynti (J)	0,381	0,034–4,216
Eteisvärinäan liittyvä perusterveydenhuollon käynti (J)	1,276*	1,215–1,340
Verisuonitaudin riskitekijöihin liittyvä perusterveydenhuollon käynti (J)	0,919*	0,861–0,980
Keuhkohtaumatautiin liittyvä perusterveydenhuollon käynti (L)	3,208	0,443–23,221
Astmaan tai allergiaan liittyvä perusterveydenhuollon käynti (L)	2,487	0,928–6,668
Masennukseen liittyvä perusterveydenhuollon käynti (J)	1,853*	1,565–2,194
Muihin mielenterveyden häiriöihin liittyvä perusterveydenhuollon käynti (J)	2,475*	2,120–2,890
Tuki- ja liikuntaelinsairauksiin liittyvä perusterveydenhuollon käynti (J)	1,475*	1,398–1,556
Perusterveydenhuollon vanhainkotipalveluiden tapahtuma (J)	1,709*	1,443–2,024
Perusterveydenhuollon kotihoidon tapahtuma (J)	1,180*	1,145–1,215
Perusterveydenhuollon kehitysvammahoidon tapahtuma (J)	1,109*	1,030–1,194
Perusterveydenhuollon vuodeosastojakso (J)	2,289*	2,036–2,574
Erikoissairaanhoidon diabetesdiagnoosi (L)	1,086	0,179–6,592
Erikoissairaanhoidon sepelvaltimotautidiagnoosi (L)	0,744	0,200–2,765
Erikoissairaanhoidon sydämen vajaatoiminta -diagnoosi (L)	0,182	0,022–1,487
Erikoissairaanhoidon eteisvärinädiagnoosi (L)	0,715	0,169–3,030
Erikoissairaanhoidon keuhkohtaumatautidiagnoosi (L)	0,530	0,000–741,636
Erikoissairaanhoidon masennusdiagnoosi (L)	2,514	0,305–20,737
Erikoissairaanhoidon mielenterveys ja käyttäytymisen häiriöt - pääluokkaan kuuluva diagnoosi (L)	1,660	0,368–7,500
Erikoissairaanhoidon tuki- ja liikuntaelinsairaudet -pääluokkaan kuuluva diagnoosi (L)	0,707	0,361–1,386
Erikoissairaanhoidon vuodeosastojakso (L)	2,073*	1,217–3,529
Sosiaalihuollon asiakkuus (L)	3,593*	1,871–6,899
Vakio	0,004	

*J = Jatkuva muuttuja*

*OR = vetosuhde (engl. odds ratio)*

*L = Luokiteltu muuttuja*

*\* = tilastollisesti merkitsevä (p-arvo on < 0,05)*

Toinen kustannuksia ennustava tekijä näyttäisi mallin mukaan olevan muihin mielenterveyden tai käyttäytymisen häiriöihin liittyvä perusterveydenhuollon käynti. Kyseessä on jatkuva muuttuja, joten vetosuhte ( $OR = 2,475$ ) kertoo vedon muutoksen silloin kun selittävän muuttujan arvo nousee yhden yksikön verran. Logistisessa regressiossa selitettävän ja selittävien muuttujien suhteen oletetaan noudattavan logaritmista jakaumaa. Esimerkiksi Kallispotilas-muuttujan ja muihin mielenterveyden ja käyttäytymisen häiriöihin liittyvä perusterveydenhuollon käynti -muuttujan välinen veto kasvaa 2,5 kertaiseksi jos potilaalla on yksi mielenterveyden tai käyttäytymisen häiriöihin liittyvä perusterveydenhuollon käynti, ja jokainen yksittäinen käynti tämän jälkeen kasvattaa vetosuhteen toiseen potenssiin. Esimerkiksi jos potilaalla on kaksi mielenterveyden ja käyttäytymisen häiriöihin liittyvää perusterveydenhuollon käyntiä, kasvaa vetosuhteen potenssiin kaksi ( $2,475 \times 2,475 = 6,125$ ), ja jos kolme käyntiä, niin vetosuhte potenssiin kolme ja niin edelleen. Muita jatkuvia muuttujia, jotka Taulukossa 14 esitellyn mallin mukaan lisäävät potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään ovat masennukseen liittyvä perusterveydenhuollon käynti ( $OR = 1,853$ ), eteisvärinäan liittyvä perusterveydenhuollon käynti ( $OR = 1,276$ ), tuki- ja liikuntaelinsairauksiin liittyvä perusterveydenhuollon käynti ( $OR = 1,475$ ) sekä perusterveydenhuollon vuodeosastojaksot ( $OR = 2,289$ ), vanhainkodin hoitajakset ( $OR = 1,709$ ), kotihoidon tapahtumat ( $OR = 1,180$ ) ja kehitysvammahuollon tapahtumat ( $OR = 1,109$ ).

Erikoissairaanhoidon muuttujista ainoastaan vuodeosastojaksot näyttäisivät lisäävän todennäköisyyttä kuulua kalliiseen potilasryhmään ( $OR = 2,073$ ). Muiden erikoissairaanhoidon muuttujien osalla erot eivät ole tilastollisesti merkitseviä ( $p > 0,05$ ). Esimerkiksi erikoissairaanhoidon masennusdiagnoosin, muihin mielenterveyden ja käyttäytymisen häiriöihin liittyvän diagnoosin sekä diabetesdiagnoosin vetosuhteet ovat myös yli 1, mutta niiden luottamusvälit ovat verrattain suuria. Luottamusvälit voivat olla ylisuuria sen takia, että analyysiin mukaan otettu erikoissairaanhoidon aineisto on melko suppea ja otoskoko sen myötä pieni. Onkin oletettavaa, että jotkin suhteet, jotka tässä mallissa eivät näytä merkitseviltä, olisivat isommalla otoskolla tarkasteltuna olleet tilastollisesti merkitseviä. Toisaalta isoissa otoksissa hyvin pienetkin erot saattavat näyttäytyä tilastollisesti merkitsevinä, joten tutkijan tehtäväksi jää aina

niiden yhteiskunnallisen merkittävyyden tulkitseminen. Malliin otettiin mukaan myös yksi sosiaalihuollon muuttuja. Sosiaalihuollon asiakkuus -muuttuja kertoo, onko potilaalla yksi tai useampi sosiaalihuollon asiakkuusmerkintä (lastensuojelu-, aikuis- ja perhetyö- tai vammaispalveluasiakkuus) vuonna 2017. Vetosuhteesta voidaan tulkita, että sosiaalihuollon asiakkuusmerkintä lisää potilaan todennäköisyyttä kuulua Kallispotilas-ryhmään ( $OR = 3,593$ ).

Jotkut regressiomallin muuttujista näyttivät lisäävän potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään, kun niitä tarkasteltiin yksittäin, mutta merkitys hävisi, kun ne lisättiin lopulliseen regressiomalliin. Esimerkiksi Ikä-muuttujan ja perusterveydenhuollon verisuonitautien riskitekijä -muuttujan suhteet selitettävään tekijään olivat omissa malleissaan tilastollisesti merkitseviä, mutta tilanne muuttui, kun malliin lisättiin muut muuttujat. Tähän voi vaikuttaa tiettyjen muuttujien vaikutus toisiinsa, ja voidaan tulkita niin, että esimerkiksi perusterveydenhuollon diabetekseen liittyvä käynti vaikuttaa Ikä-muuttujan ja Kallispotilas-muuttujan väliseen suhteeseen.

Taulukko 16 kuvaa lopullisen mallin ennustusvoiman eli sen, kuinka hyvin malli ennustaa Kallispotilas-muuttujan arvoja tilanteessa, jossa se ei tiedä oikeita vastauksia. Taulukon mukaan malli ennustaa Muut-ryhmään (Kallispotilas-muuttujan arvo 0) kuuluvat potilaat 99,3 prosentin todennäköisyydellä oikein, ja kaikista kalleimpaan ryhmään kuuluvat potilaat 83 prosentin todennäköisyydellä. Kokonaisuudessaan malli luokitteli oikein 97,7 prosenttia havainnoista.

Taulukko 16

*Regressiomallin ennustusvoima*

Havaittu	Ennustettu		
		Kallis potilas	Oikein
		Ei (0)	Kyllä (1)
Kallis potilas	Ei (0)	5021	34
	Kyllä (1)	96	470
Yhteensä (%)			97,7

## 6 Pohdinta

Tutkielman pohdintaosion ensimmäisessä alaluvussa muistutetaan lukijaa tutkielman tavoitteista ja tutkimuskysymyksistä, esitellään tutkimuskysymysten kannalta keskeisemmät analyysin tulokset ja verrataan niitä aikaisempien tutkimusten tuloksiin. Pohdintaluvun toisessa alaluvussa arvioidaan tutkimuksen luotettavuutta, eli validiteettia ja reliabiliteettia, aineiston asettamien rajojen, kuten otoskoon ja puuttuvien tietojen sekä muuttujien valinnan ja luokittelun kannalta. Kolmannessa alaluvussa käsitellään tutkimuseettisiä näkökulmia.

### 6.1 Keskeiset tulokset ja niiden tulkinta

Tutkielman tavoitteena oli selvittää, mitkä tekijät ennustavat korkeita potilaskohtaisia sosiaali- ja terveystalouksia. Edellisessä pääluvussa raportoidun logistisen regressiomallinnuksen avulla vastattiin kahteen tutkimuskysymykseen

- 1. Mitkä terveydenhuollon suoritteet lisäävät yksittäisen potilaan todennäköisyyttä kuulua kaikista kalleimpien potilaiden ryhmään?*
- 2. Lisääkö sosiaalipalveluiden asiakkuus potilaan todennäköisyyttä kuulua kaikista kalleimpien potilaiden ryhmään?*

Kuvailevien tunnuslukujen tarkastelun perusteella voidaan todeta, että potilaiden kallein kymmenys aikaansai määräänsä suuremman osuuden monista erikoissairaanhoidon diagnooseista sekä perusterveydenhuollon käynneistä ja tapahtumista. Esimerkiksi vuonna 2017 tutkittavan kunnan kaikista perusterveydenhuollon vanhainkodin hoitajaksoista jopa 86 prosenttia koski pelkästään potilaiden kalleinta 10 prosenttia. Perusterveydenhuollon kotihoidon tapahtumien, erikoissairaanhoidon sydämen vajaatoiminnan diagnoosien sekä keuhkohtaumatauti-diagnoosien osalta kalleimman

kymmenyksen osuus oli yli 60 prosenttia. Perusterveydenhuollon kehitysvammahoidon tapahtumissa sekä vuodeosastojaksoissa vastaava luku oli yli 50 prosenttia.

Kustannusten tarkastelu osoittaa, että kallein kymmenys aikaansai lähes 90 prosenttia kaikista aineiston kustannuksista. Tulos on samansuuntainen Suomessa aiheesta aikaisemmin tehtyjen tutkimusten kanssa (ks. esim. Leskelä et al. 2013; Kapiainen et al. 2010).

Aikaisemman tutkimuksen perusteella tälle tutkielmalle määriteltynä hypoteesina oli, että korkeita potilaskohtaisia sotekustannuksia ennustaisivat etenkin ikä, masennus, muut mielenterveyden häiriöt, sydämen vajaatoiminta, keuhkohtaumatauti, diabetes, tuki- ja liikuntaelinsairaudet, verisuonisairaudet sekä sosiaalihuollon asiakkuus (Chechulin et al. 2014; Kapiainen et al. 2010; Reid et al. 2003; Ash et al. 2001; Reho et al. 2018; Rais et al. 2013; Calver et al. 2006). Regressioanalyysi todisti hypoteesin oikeaksi ainakin osittain. Analyysi paljastaa, että yksittäisen potilaan todennäköisyyttä kuulua kaikista kalleimpaan potilaiden ryhmään lisäävät sekä tietyt, yksittäisiin diagnooseihin että sairausryhmiin liittyvät perusterveydenhuollon käynnit, perusterveydenhuollon kotihoidon, vanhainkodin, kehitysvammahuollon ja vuodeosaston hoitajaksot, erikoissairaanhoidon vuodeosastojaksot sekä sosiaalihuollon asiakkuus. Kuten seuraavaksi huomataan, ovat tämän tutkielman tulokset samansuuntaisia aikaisempien tutkimusten tulosten kanssa.

Mallin mukaan yksi potilaskohtaisia kustannuksia ennustava tekijä on diabetekseen liittyvä perusterveydenhuollon käynti. Diabeteksen yhteys korkeisiin potilaskohtaisiin terveydenhuollon kustannuksiin on osoitettu myös muissa tutkimuksissa (ks. Chechulin et al. 2014; Ash et al. 2001; Kapiainen et al. 2010). Muut tutkimukset (mm. Chechulin et al. 2014; Kapiainen et al. 2010; Reid et al. 2003; Ash et al. 2001; Reho et al. 2018) ovat osoittaneet diabeteksen lisäksi myös sydämen vajaatoiminnan, keuhkohtaumataudin ja verisuonisairauksien lisäävän potilaan todennäköisyyttä päätyä kaikista kalleimpien potilaiden ryhmään, mutta näiden diagnoosien yhteys korkeisiin kustannuksiin ei ollut tässä tutkielmassa tilastollisesti merkitsevä.

Myöskään iän yhteydestä korkeisiin potilaskohtaisiin sote-kustannuksiin ei tämän tutkielman tulosten perusteella voi tehdä selkeitä johtopäätöksiä, koska näiden välinen yhteys ei lopullisessa mallissa ollut tilastollisesti merkitsevä ( $p > 0,05$ ). Iän on kuitenkin havaittu muissa tutkimuksissa olevan merkittävä potilaskohtaisia kustannuksia ennustava tekijä siten, että yli 65-vuotiailla on selvästi suurempi todennäköisyys kuulua kalliiden potilaiden ryhmän (ks. esim. Rais et al. 2013; Reid et al. 2003; Calver et al. 2006). Toisaalta, Kallispotilas-ryhmän keski-ikä (yli 69 vuotta) oli huomattavasti korkeampi verrattuna koko analyysiaineiston keski-ikään (46 vuotta). Lisäksi, kun Ikä- ja Kallispotilas-muuttujien yhteyttä tarkasteltiin omassa mallissaan, oli iän vaikutus korkeisiin potilaskohtaisiin kustannuksiin tilastollisesti merkitsevä niin, että yhden yksikön lisäys Ikä-muuttujassa lisäsi potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään. Tämä vaikutus kuitenkin katosi, kun malliin lisättiin muut muuttujat. Tämä voi johtua siitä, että jotkin muut regressiomallissa olevat muuttujat saattavat vaikuttaa Ikä-muuttujan ja Kallispotilas-muuttujan väliseen vetosuhteeseen. Mallissa on esimerkiksi mukana sellaisia muuttujia (esim. perusterveydenhuollon vanhainkodin ja kotihoidon tapahtumat), jotka ovat yleisiä yli 65-vuotiailla potilailla, mikä voi mahdollisesti piilottaa iän vaikutuksen.

Regressioanalyysin perusteella voidaan tulkita, että myös masennus lisää potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään. Masennukseen liittyvä perusterveydenhuollon käynti nosti vedon lähes kaksinkertaiseksi. Tämä vahvistaa Kapiaisen et al. (2010) sekä Rehon et al. (2018) omissa tutkimuksissaan saamia tuloksia masennuksen osallisuudesta korkeiden potilaskohtaisen kustannusten ja palveluiden suurkäytön rakentumisessa. Muihin mielenterveyden ja käyttäytymisen häiriöihin liittyvä käynti nostaa mallin mukaan vedon 2,5 kertaiseksi. Myös tämä on samansuuntainen aikaisempien tutkimusten kanssa (ks. Chechulin et al. 2014; Reid et al. 2003; Kapiainen et al. 2010; Reho et al. 2018). Potilaiden kallein kymmenys aikaansaai lähes 50 % kaikista masennukseen ja muihin mielenterveyden ja käyttäytymisen häiriöihin liittyvistä perusterveydenhuollon käynneistä.



Myös tuki- ja liikuntaelinsairaudet nousevat analyysissä esiin yhtenä korkeita kustannuksia ennustavana tekijänä. Myös Reho et al. (2018) havaitsivat tämän yhteyden tutkimuksessaan, jossa he tarkastelivat suomalaisessa työterveyshuollossa paljon palveluita käyttäviä potilaita. Tämän tutkielman tulosten mukaan myös eteisvärinäan liittyvät käynnit, perusterveydenhuollon ja erikoissairaanhoidon vuodeosastojaksot sekä vanhainkodin, kotihoidon ja kehitysvammahuollon tapahtumat lisäävät potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään. Kaikista kalleimpien potilaiden ryhmä vastasi lähes 55 prosentista kaikista kehitysvammahuollon tapahtumista, 67 prosentista kotihoidon tapahtumista ja noin 53 prosentista perusterveydenhuollon vuodeosastojaksoista.

Tutkielman toisena tutkimuskysymyksenä oli, lisääkö sosiaalihuollon asiakkuus potilaan todennäköisyyttä kuulua kalliiden potilaiden ryhmään. Analyysi osoitti, että mikäli terveydenhuollon potilaalla on vähintään yksi sosiaalihuollon asiakkuus -merkintä, on hänen todennäköisyytensä kuulua kalliiden potilaiden ryhmään suurempi niihin potilaisiin verrattuna, joilla tällaista merkintää ei ole. Tämän lisäksi myös vähintään yksi perusterveydenhuollon vanhainkodin hoitojakso tai tapahtuma lisää potilaan todennäköisyyttä kuulua kalliiseen ryhmään. Kalliiseen potilasryhmään kuuluvat saivat aikaan yli 80 % kaikista tutkittavan kunnan perusterveydenhuollon vanhainkotitapahtumista. Tämä vahvistaa aikaisempien tutkimusten tuloksia siitä, että sosiaali- ja terveystalouden kalleimpia asiakasryhmiä ovat nimenomaan lastensuojelun ja vanhustalouden asiakkaat, ja että kustannukset nousevat hyvin korkeiksi etenkin niillä potilailla, jotka käyttävät terveystalouden lisäksi vähintään yhtä sosiaalipalvelua. (Leskelä et al. 2013.)

## 6.2 Luotettavuuden arviointi

Tutkielman aineistona käytettiin potilastietojärjestelmistä koostettua rekisteriaineistoa, jossa yksittäisiä henkilöitä oli noin 5700 kappaletta. Potilastietojärjestelmät sisältävät luotettavaa tietoa potilaiden terveydentilasta ja diagnooseista. Potilastietojärjestelmien

tutkimuskäyttöön liittyy usein kuitenkin myös ongelmia, joita myös tämän tutkielman toteutusvaiheessa kohdattiin.

Ensinnäkin tutkittavan kunnan koko (asukkaita alle 6000) on jopa Suomen mittakaavassa niin pieni, ettei pelkästään sen potilastietojärjestelmistä kerätyn aineiston avulla pystytä tekemään tutkimusta, joka olisi yleistettävissä laajemmin koko Suomeen. Tutkittavassa kunnassa ikääntyneiden, eli yli 65-vuotiaiden, osuus koko väestöstä on 29 %, mikä on jonkin verran koko maan keskiarvoa (21 %) korkeampi. Myös kunnan huoltosuhte (80,8) on selvästi heikompi kuin muualla suomessa keskimäärin (60). (Kuntaliitto 2018.) Ikääntyneiden osuus väestöpohjasta voi vaikuttaa tutkimustulosten luotettavuuteen, sillä se näkyy todennäköisesti myös kunnan sosiaali- ja terveystalouden asiakas- ja potilaskunnassa ja saattaa näin vääristää tulosten ikäjakaumaa. Tilastokeskuksen (2018d) mukaan suomalaisten keski-ikä vuonna 2017 oli 42,7 vuotta. Tässä tutkielmassa potilaiden keski-ikä on noin 46 vuotta, mikä on hieman maan keskiarvo suurempi, mutta eroa ei pidetty tutkimusasetelman kannalta merkittävänä.

Toiseksi, eri kuntien, sairaanhoitopiirien ja yksityisten palveluntuottajien potilastietojärjestelmien kirjauskäytännöt eroavat toisistaan merkittävästi. Esimerkiksi diagnoositietojen merkitseminen jokaisen vastaanottokäynnin tai muun asiakaskontaktin yhteydessä ei ole kaikissa näistä pakollista, tai se jää muusta syystä hoitohenkilökunnalta tekemättä. Tämä oli nähtävissä myös tämän tutkielman aineistossa; erikoissairaanhoidon aineistossa päädiagnoosit oli järjestelmällisesti merkitty jokaisen suoritteen kohdalle, kun taas perusterveydenhuollon aineistossa diagnoosikoodin merkitseminen näytti olevan pikemminkin poikkeus kuin sääntö. Perusterveydenhuollon Effica-datasta löytyi kuitenkin tulosyy-koodit (ICPC), joita pystyttiin käyttämään sairastavuuden mittarina diagnoositietojen sijaan.

Tutkimuksen luotettavuuteen vaikuttaa merkittävästi myös otoskoko, mikä täytyy ottaa tässä tutkielmassa huomioon etenkin erikoissairaanhoidon laskudatan sekä

sosiaalihuollon ProConsona-datan osalta. Molempien aineistojen osalta jouduttiin tekemään päätös siitä, että lopulliseen regressiomalliin mukaan valitut muuttujat luokiteltiin kaksiluokkaisiksi, niiden verrattain pienen vaihtelun vuoksi. Otokoko vaikuttaa myös ennustavan mallin luotettavuuteen. Mikäli otokoko jää hyvin pieneksi (esim. sosiaalihuollon aineistossa  $n = 340$ ), voidaan regressiomallissa havaita poikkeuksellisen suuria luottamusvälejä tai yllättävän näköisiä korrelaatioita (vetosuhteita). Analyysiluvussa raportoitua regressiomallia tarkasteltaessa huomataan, että monien selittävien muuttujien osalla luottamusvälit ovat melko suuria, mikä vaikuttaa niiden ja selitettävän muuttujan yhteyteen vetosuhdetta pienentäen. Voidaankin olettaa, että jotkin mallissa raportoidut muuttujien väliset yhteydet olisivat isommassa tutkimusaineistossa tilastollisesti merkitseviä (esimerkiksi keuhkohtaumatauti ja sydämen vajaatoiminta). Tätä oletusta tukee muiden tutkimusten tulokset, jotka osoittavat selkeästi näiden diagnoosien merkityksen kalliiseen potilasryhmään kuulumisen ennustajina (ks. esim. Chechulin et al. 2014; Kapiainen et al. 2001; Ash et al. 2001).

Luotettavuuden arvioinnissa tutkielman tekijän on otettava huomioon myös muuttujien muokkaus- ja valintavaiheessa tehdyt päätökset ja rajaukset. Esimerkiksi muuttujien koodaus voi vaikuttaa tutkimustuloksiin. Tässäkin tutkielmassa koodausvaiheessa on jouduttu tekemään tiettyjä rajauksia ja päätöksiä, jotka on kuitenkin pyritty raportoimaan mahdollisimman kattavasti luvussa 4. Muuttujien valinta -alaluku sisältää taulukoita, joihin on raportoitu selkeästi yksittäisiin muuttujiin yhdistetyt diagnoosi- ja tulosykköidit, muuttujien mitta-asteikot ja niiden selitykset.

Kallispotilas-muuttujan muodostamisen apuna käytetty Yhdistettyhinta-muuttuja sisälsi aineiston sisällön asettamien rajoitteiden takia paljon puuttuvia tietoja. Puuttuvat kustannustiedot korvattiin kuvailevaa tarkastelua ja Kallispotilas-muuttujan muodostamista varten luvulla 0 €, joten on tärkeää mainita, etteivät tuloksissa raportoidut kustannustiedot kuvaa terveydenhuollon kokonaiskustannuksia. Aineiston muokkausvaiheessa tehtiin oletus, että aineiston kustannustiedot (puuttuvista tiedoista huolimatta) olivat tässä tutkimusasetelmassa riittäviä selvittämään terveystalouden

epätasaista jakautumista sekä kalliiksi määriteltyjen ja muiden potilaiden välisiä eroja. Vertailun vuoksi Kallispotilas-ryhmän aikaansaamien kustannusten osuus kaikista kustannuksista on esitetty tutkielman luvussa 5 korvaamalla 0 € -summa vuoden 2017 perusterveydenhuollon potilaskohtaisilla nettokustannuksilla (Tilastokeskus 2018c).

Myös selittävien muuttujien valinta vaikuttaa analyysin tuloksiin ja on täten olennainen osa tutkimuksen luotettavuuden arviointia. Tämän tutkielman regressiomalliin pyrittiin ottamaan mukaan mahdollisimman kattava muuttujalistaus, joka sisältää muuttujia kaikista kolmesta osa-aineistosta. Pyrkimyksenä oli sisällyttää analyysiin kaikkia tutkittavan kunnan sosiaali- ja terveystalouden palvelumuotoja tulosten kattavuuden ja luotettavuuden lisäämiseksi. Muuttujien valintaan on vaikuttanut olennaisesti kirjallisuuskatsauksessa esitetyt aikaisemmat tutkimukset ja niiden tulokset. Tämän vuoksi regressiomallissa on mukana sairausryhmien lisäksi tiettyjä yksittäisiä diagnoosi- ja tulosyymereitä, kuten masennus, keuhkoastma, sydämen vajaatoiminta ja diabetes. Näihin sairauksiin liitetyt perusterveydenhuollon ja erikoissairaanhoidon käynnit on todettu aikeisemmissä tutkimuksissa olevan merkittäviä potilaskohtaisia kustannuksia ennustavia tekijöitä, mutta kuten tämän tutkielman luvun 5 raportoinnista on huomattavissa, ei kaikilla näillä muuttujilla ollut tässä tutkimusasetelmassa tilastollisesti merkitsevää yhteyttä kustannuksiin. Toisaalta niidenkin muuttujien osalta, jotka eivät mallissa tuottaneet tilastollisesti merkitsevää yhteyttä Kallispotilas-muuttujan kanssa, logistisen regressiomallin vetosuhteet olivat kuitenkin samansuuntaisia kuin tutkimuskirjallisuudessa. Voidaan olettaa, että käsillä olevan aineiston rajallisuus ja otoskoon pienuus vaikuttivat tuloksiin niin, että jotkut kansallisella (ja kansainvälisellä) tasolla tilastollisesti merkitsevät selitettävän ja selittävien muuttujien väliset yhteydet jäivät tässä tutkimuksessa piiloon.

Tutkimuksen luotettavuuden arvioinnissa ei tule unohtaa muuttujien luokittelun merkitystä tutkimustuloksiin. Tämän tutkielman logistinen regressiomalli sisälsi sekä luokiteltuja että jatkuvia muuttujia, ja luokitteluihin liittyvät valinnat on pyritty perustelemaan kattavasti tutkielman luvussa 4. Kaikki regressiomallin luokitellut muuttujat ovat kaksiluokkaisia muuttujia, joissa arvo 0 merkitsee tietyn diagnoosin,

tulosityyn tai hoitojakson puuttumista, ja arvo 1 sitä, että potilaalla on vähintään yksi tietyn diagnoosin, tulosityyn tai hoitojakson merkintä. Yhtenä kriittisenä huomiona tähän liittyen voidaan pitää sitä, että muuttujien luokittelu kaksiluokkaisiksi poistaa mahdollisuuden vertailla niiden lukumääriä ja toisaalta myös lukumäärien vaikutusta vetosuhteeseen. Jatkuvien ja useampiluokkaisten muuttujien avulla on mahdollista tehdä tarkempaa vertailua esimerkiksi käyntimäärien ja komorbiditeetin, eli yhteissairastavuuden vaikutuksesta kustannuksiin (ks. esim. Reid et al. 2003; Rais et al. 2013; Fleishman & Cohen 2010). Tutkijat joutuvat kuitenkin aina toimimaan aineiston asettamissa rajoissa, ja tässäkin tutkielmassa pieni otoskoko sekä joidenkin muuttujien puuttuvien tietojen suuri määrä tai yksittäisten tapahtumien harvinaisuus on vaatinut tiettyjen menetelmällisten kompromissien tekemistä.

Suurin osa työssäkäyvästä väestöstä ei käytä julkisia sosiaali- ja terveystalvveluita vaan hyödyntää työnantajan tarjoamia yksityisiä työterveystalvveluita. Käytännössä julkisia sotetalvveluita käyttää pääsääntöisesti siis alle 18 ja yli 65-vuotiaat. Tämä seikka on otettava huomioon tutkimustulosten luotettavuuden arvioinnissa, sillä julkisten sotetalvveluiden potilastietojärjestelmistä koostuvan aineiston ikäjakauma voi olla vinoutunut myös edellä mainitusta asiasta johtuen.

Monissa tutkielman luvussa 2 esitellyistä tutkimuksista paljon talvveluita käyttäviä tai kaikista kalleimpia potilaita on tutkittu hyödyntäen useammalta eri vuodelta koostettua aineistoa (ks. esim. Chechulin et al. 2014). Kahden tai useamman peräkkäisen vuoden vertailu antaa mahdollisuuden testata regressiomallia esimerkiksi niin, että malli muodostetaan vuoden X tietojen perusteella ja tämän jälkeen mallin toimivuutta testataan vuoden Y aineistolla. Tämä tutkielma on kuitenkin poikkileikkaustutkimus, jossa aineisto sisältää vain vuoden 2017 käynti- ja diagnoositietoja, joten edellä mainittua menetelmää ei ollut mahdollista hyödyntää. Tulokset ovat kuitenkin pääpiirteittäin samansuuntaisia aikaisempien tutkimusten kanssa, joten jo tämän itsessään voidaan todeta kertovan jotakin tutkimusasetelman ja menetelmien luotettavuudesta.

### 6.3 Tutkimuseettiset näkökulmat

Tämä Pro gradu -tutkielma on toteutettu hyvää tieteellistä käytäntöä noudattaen. Alustava tutkimussuunnitelma hyväksyttiin ensin tutkittavan kunnan kunnanjohtajalla, jonka jälkeen kaikille tutkielmassa käytettäville aineistoille hankittiin tutkimusluvut asianmukaisilta viranomaisilta. Perusterveydenhuollon aineiston tutkimusluvan allekirjoitti kunnan johtava virkalääkäri, ja sosiaalihuollon aineiston osalta johtava sosiaalityöntekijä. Molemmissa tapauksissa lupakäsittely eteni niin, että päättävälle tahoille lähetettiin alustava tutkimussuunnitelma, jota muokattiin yhdessä johtavien virkamiesten kanssa. Tutkimusluvuissa rajattiin tarkasti ne suoritteet, joita potilasjärjestelmistä oli aiheellista kerätä, ja lupien ulkopuolelle rajautui kaikista arkaluonteisimmat tiedot, mukaan lukien potilaskertomukset. Erikoissairaanhoidon laskuaineiston käyttöön tutkimuslupaa ei erikseen haettu, sillä se koostuu kunnan omistamasta, Mehiläiselle lähetetyistä erikoissairaanhoidon läheteiden mukaan koostetuista laskutiedoista, ja käyttöluva siihen saatiin suoraan Mehiläiseltä.

Koska tutkielman toteutuksen kannalta oli olennaista yhdistää dataa kolmesta eri tietolähteestä, muodostui niistä uusi henkilörekisteri. Tästä on laadittu Tietosuojalain, EU:n tietosuoja-asetuksen ja Julkisuuslain mukainen tietosuojaseloste, jossa selvennetään rekisterinpitäjän ja tutkimuksen vastuullisen johtajan henkilöllisyydet, henkilötietojen käyttöperuste sekä se, mistä osatiedoista uusi henkilörekisteri koostuu, mihin sitä on tarkoitus käyttää, miten sitä säilytetään, miten henkilötietojen turvaaminen ja tietosuoja järjestetään ja milloin tutkimusrekisteri hävitetään.

Myös aineistonkeruu, muokkaus ja säilytys on hoidettu eettisten periaatteiden mukaisesti. Aineistonkeruussa käytettiin kunkin potilastietojärjestelmän ylläpitäjän valmiita teknisiä rajapintoja. Perusterveydenhuollon aineiston (Effic) tiedonkeruun hoiti järjestelmän ylläpitäjä Tieto Oyj, sosiaalihuollon aineiston (ProConsona) osalta CGI. Aineistonkeruun tilausvaiheessa sovittiin aineiston salauksesta, jakamisesta ja säilytyksestä. Erikoissairaanhoidon aineisto saatiin tutkittavan kunnan Business

Controllerilta, ja se toimitettiin Mehiläisen sovellusasiantuntijalle muistitikun avulla, jotta aineisto saatiin salattua samalla salausavaimella kuin perusterveydenhuollon ja sosiaalihuollon aineistot. Kaikki aineistot sisälsivät henkilötunnustietoja, joten ne kryptattiin ennen luovuttamista käyttämällä samaa salausavainta, jotta ne olisivat myöhemmin yhdistettävissä. Salausavain ei ole missään vaiheessa ollut tutkijan tiedossa. Yksittäisistä potilaista säilytettiin käyntitilastojen ja diagnoosien lisäksi pelkästään sukupuoli- ja ikätiedot.

Kerättyjen, ja valmiiksi kryptattujen, aineistojen siirto tutkielman tekijälle järjestettiin salatun sähköpostin avulla. Kukin lähetetty tiedosto oli lisäksi salattu salasanalla, joka lähetettiin tutkijan ilmoittamaan puhelinnumeroon tekstiviestitse. Kaikki tutkimusmateriaali, aineistot mukaan lukien, säilytettiin turvallisella, Mehiläisen ylläpitämällä palvelimella, johon kolmansilla osapuolilla ei ole pääsyä.

## 7 Johtopäätökset

Tämän tutkielman ensisijaisena tavoitteena oli selvittää, mitkä yksittäiset tekijät selittävät tutkittavan kunnan terveydenhuollon potilaan päätymistä kaikista kalleimpaan 10 % potilaiden ryhmään. Tämän lisäksi pyrittiin selvittämään, lisääkö sosiaalihuollon asiakkuus potilaan todennäköisyyttä kuulua kalleimpien potilaiden ryhmään.

Potilastietojärjestelmistä kerätyllä tilastoaineistolla saadut tulokset vahvistavat suurelta osin aikaisempien paljon palveluita käyttäviä ja kaikista kalleimpia potilaita selvittäneiden tutkimusten tuloksia. Potilaiden kallein kymmenes aikaansai tutkittavan kunnan terveydenhuollon menoista lähes 90 %. Tulos on samansuuntainen aikaisempien Suomessa tehtyjen tutkimusten kanssa (ks. esim. Leskelä et al. 2013; Kapiainen et al. 2010). Analyysin havainnot vahvistavat myös aikaisempaa käsitystä siitä, että yksittäisen potilaan kuulumista kaikista kalleimpien potilaiden ryhmään ennustaa vahvasti masennukseen ja muihin mielenterveyden ja käyttäytymisen häiriöihin ja tuki- ja liikuntaelinsairauksiin liittyvät perusterveydenhuollon käynnit sekä aikuistyypin diabetesdiagnoosi (ks. Kapiainen et al. 2010; Reho et al. 2018; Ash et al. 2001; Reid et al. 2003; Rais et al. 2013; Chechulin et al. 2014). Tulokset vahvistivat myös suomalaisen tutkimuksen tulokset siitä, että sosiaalipalveluiden ja vanhuspalveluiden asiakkuus lisää merkittävästi potilaan todennäköisyyttä kuulua kalleimpien potilaiden ryhmän (Leskelä et al. 2013). Tutkielma antaa viitteitä myös eteisvärinään liittyvien perusterveydenhuollon käyntien, kotihoidon ja kehitysvammahuollon tapahtumien sekä perusterveydenhuollon ja erikoissairaanhoidon vuodeosastojaksojen yhteydestä korkeisiin potilaskohtaisiin sotekustannuksiin.

Sosiaali- ja terveystalouden ennustaminen ja paljon terveystaloudellisesti käyttävien ja tarvitsevien potilaiden tunnistaminen ovat ajankohtaisia aiheita, joilla on suuri yhteiskunnallinen ja kansantaloudellinen merkitys. Mikäli paljon palveluita käyttävät potilaat ja heidän hoidontarpeensa kyettäisiin tunnistamaan paremmin, olisi sekä julkisten että yksityisten palveluntuottajien mahdollista luoda heille yksilöidympiä



hoitopolkuja, parantaa hoidon vaikuttavuutta, keskittyä terveysongelmien ehkäisyyn sairauksien parantamisen sijaan, ja mahdollisesti myös hidastaa sote-kustannusten kasvua. Tutkimusaiheen kansantaloudellinen merkitys on tunnustettu myös Suomen ylimpiä hallinnollisia toimielimiä myöten. Valtion tarkastusviraston lausunnon (2017) mukaan Sosiaali- ja terveysministeriön tulisi jatkossa ohjata resursseja paljon palveluita käyttävien ja tarvitsevien potilaiden sekä heidän tarpeittensa tunnistamiseen. Tämän tutkielman tulosten perusteella etenkin mielenterveys- ja sosiaalipalveluihin, ikääntyneiden palveluihin ja diabeteksen hoitoon sekä näitä palveluita käyttävien potilaiden ja asiakkaiden hoitopolkujen kehittämiseen tulisi kiinnittää nykyistä enemmän huomiota.

Kansainväliset tutkimukset osoittavat, että paljon palveluita käyttävät tai kalliit potilaat eivät ole homogeeninen ryhmä, vaan heidät voidaan jakaa moneen eli alaryhmään, joilla kaikilla on eri syyt palveluiden suurenkäytölle (Lee et al. 2017). Todennäköisesti jotkin näistä ryhmistä ovat sellaisia, joiden aikaansaamia kustannuksia on vaikea karsia terveyden edistämisen keinoin, eivätkä heidän sairautensa ole helposti ennaltaehkäistävissä koska ne eivät välttämättä johdu suoraan yksilön elämäntavoista. Olisikin tärkeää koittaa jatkossa identifioida näitä eri paljon palveluita käyttävien ryhmiä myös Suomessa, ja selvittää, mitä eroja ja yhtäläisyyksiä heillä on. Myös palveluiden suurenkäyttöä Suomessa käsittelevän pitkittäistutkimuksen toteuttaminen olisi aiheellista, koska poikkileikkaustutkimus ei anna tietoa esimerkiksi siitä, pysyvätkö potilaat samassa paljon palveluita käyttävien ryhmässä vuodesta toiseen, vai tapahtuuko liikkumista ryhmien välillä tai niistä pois.

Yksi mielenkiintoinen tutkimuskohde on myös yksityisen työterveyshuollon asiakkaat, joiden palveluiden käyttöä ovat Suomessa jo tutkineet esimerkiksi Reho et al. (2018). Työterveyshuollon asiakkaat eroavat demografisilta tekijöiltään julkisen terveydenhuollon potilaista, koska he kaikki ovat työllistettyjä ja työikäisiä. Tämä väestönosa jää pelkästään julkisten terveystalouden palveluiden käyttöä tarkastelevissa tutkimuksissa tutkimusasetelman ulkopuolelle, koska suuri osa heistä ei käytä julkisia perusterveydenhuollon palveluita. Jatkossa olisikin tärkeää selvittää, miten

työterveyshuollon paljon palveluita käyttävät asiakkaat ja julkisen perusterveydenhuollon paljon palveluita käyttävät potilaat eroavat toisistaan.

Koska sosiaalihuollon asiakkuudella on useiden diagnoosien ja sairausryhmien ohella todettu olevan vahva yhteys korkeisiin potilaskohtaisiin sotekustannuksiin sekä palveluiden suurkankeeseen, tulisi mahdollisuudet sosiaali- ja terveydenhuollon potilas- ja asiakasrekisterien yhteiskäyttöön selvittää. Tietojen yhteiskäytön mahdollistaminen esitetään yhtenä Juha Sipilän terveyspalveluiden asiakaslähtöisyyttä koskevan kärkihankkeen tavoitteiden saavuttamisen keinona (Valtioneuvoston kanslia 2018). Järjestelmien yhdistäminen antaisi entistä paremmat työkalut paljon palveluita käyttävien ja kalliiden potilaiden sekä heidän tarpeittensa tunnistamiseen ja hoidon kustannusvaikuttavuuden parantamiseen. Yhtenä mahdollisena ratkaisuna kustannusten kiihtyvään kasvuun on esitetty asiakasvastaava-toimintamallia (case management), jonka keskiössä toimisivat täydennyskoulutetut sairaanhoitajat. Toimintamallin pilotoinneista on saatu positiivisia tuloksia, joiden mukaan mallin käyttö vähentää sekä potilaiden että hoitohenkilökunnan stressiä, parantaa potilaiden itse arvioitua terveydentilaa sekä vähentää avohoito- ja vuodeosastojaksojen määrää ja terveydenhuollon kustannuksia (Hudon et al. 2018; Reinius et al. 2013).

## Lähteet

Andersson, S. O., Lynöe, N., Hallgren, C. G., & Nilsson, M. (2004). Is frequent attendance a persistent characteristic of a patient? Repeat studies of attendance pattern at the family practitioner. *Scandinavian journal of primary health care*, 22(2), 91–94.

Ash A.S., Zhao Y., Ellis R.P. & Kramer M.S. (2001). Finding Future High-Cost Cases: Comparing Prior Cost versus Diagnosis-Based Methods. *Health Services Research*, 36(6 Part II), 194–206.

Berk M.L. & Monheit A.C. (2001). The Concentration of Health Care Expenditures, Revisited. *Health Affairs*, 20(2), 9–18.

Bland, J. M., & Altman, D. G. (2000). The odds ratio. *Bmj*, 320(7247), 1468.

Brilleman, S. L., Gravelle, H., Hollinghurst, S., Purdy, S., Salisbury, C., & Windmeijer, F. (2014). Keep it simple? Predicting primary health care costs with clinical morbidity measures. *Journal of Health Economics*, 35(100), 109–122.

Calver, J., Brameld, K.J., Preen, D.B., Alexia, S.J., Boldy, D.P. & McCaul, K.A. (2006). High-Cost Users of Hospital Beds in Western Australia: A Population-Based Record Linkage Study. *Medical Journal of Australia*, 184, 393–97.

Chechulin, Y., Nazerian, A., Rais, S. & Malikov, K. (2014). Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada). *Health Policy*, 9(3), 68–79.

Deber, R. B., Kraetschmer, N., Urowitz, S., & Sharpe, N. (2005). Patient, consumer, client, or customer: what do people want to be called? *Health Expectations*, 8(4), 345–351.

Fleishman H.A. & Cohen J.W. (2010). Using Information on Clinical Conditions to Predict High-Cost Patients. *Health Services Research*, 45(2), 532–52.

Hudon, C., Chouinard, M. C., Dubois, M. F., Roberge, P., Loignon, C., Tchouaket, É., ... & Bouliane, D. (2018). Case management in primary care for frequent users of health care services: a mixed methods Study. *The Annals of Family Medicine*, 16(3), 232–239.

Kapiainen, S., Seppälä, TT., Häkkinen, U., Lauharanta, J., Roine, R. & Korppi-Tommola, M. (2010). Pääkaupunkiseudun erittäin kalliit potilaat. THL Avauksia 3/2010.

Kapiainen, S., Väisänen, A. & Haula, T. (2014). Terveyden- ja sosiaalihuollon yksikkökustannukset Suomessa vuonna 2011. THL raportti 3/2014.

Karppinen, K. (2018). Potilas- ja asiakas -sanojen käyttö terveydenhuollon organisaatioiden verkkosivuolla. Pro gradu -tutkielma. Oulun yliopisto.

Kivelä, K., Elo, S., & Kääriäinen, M. (2018). Frequent Attenders in Primary Health Care: A Concept Analysis. *International journal of nursing studies*, 86, 115–124.

Koskela, T. H., Ryyänen, O. P., & Soini, E. J. (2010). Risk factors for persistent frequent use of the primary health care services among frequent attenders: a Bayesian approach. *Scandinavian journal of primary health care*, 28(1), 55–61.

Kujanpää, T. S., Jokelainen, J., Auvinen, J. P., & Timonen, M. J. (2017). The association of generalized anxiety disorder and Somatic Symptoms with frequent attendance to health care services: A cross-sectional study from the Northern Finland Birth Cohort 1966. *The International Journal of Psychiatry in Medicine*, 52(2), 147–159.

Kuntaliitto (2017). ICPC-2 - Perusterveydenhuollon kansainvälinen luokitus. <https://www.kuntaliitto.fi/asiantuntijapalvelut/sosiaali-ja-terveysasiat/icpc-2-perusterveydenhuollon-kansainvalinen-luokitus> [Luettu 22.2.2019].

Kuntaliitto (2018). Kuntien väestö ja ikärakenne 31.12.2017. <https://www.kuntaliitto.fi/asiantuntijapalvelut/talous/kuntien-vaesto-ja-ikarakenne> [Luettu 11.2.2019].

Laki potilaan asemasta ja oikeuksista 785/1992.

Laki sosiaalihuollon asiakkaan asemasta ja oikeuksista 812/2000.

Lappalainen, M. (2015). Constructing a framework to manage high utilizers in social and health care. Department of Information and Service Economy. Aalto University School of Business.

Lee, N. S., Whitman, N., Vakharia, N., & Rothberg, M. B. (2017). High-cost patients: Hot-spotters don't explain the half of it. *Journal of general internal medicine*, 32(1), 28–34.

Leskelä, R-L., Komssi, V., Sandström, S., Pikkujämsä, S., Haverinen, A., Olli, S-L. & Ylitalo-Katajisto, K. (2013). Paljon sosiaali- ja terveystalvveluja käyttävät asukkaat Oulussa. *Suomen Lääkärilehti* 48/2013 v sk 68.

Lääkärilehti (2019). Asiakas syrjäyttää potilaan.  
<https://www.laakarilehti.fi/blogi/hannu-ollikainen/asiakas-syrjayttaa-potilaan/?public=c44b02cb10f02ba45162607675d40d41> [Luettu 11.2.2019].

Metsämuuronen, J. (2006). *Tutkimuksen tekemisen perusteet ihmistieteissä* 2 (4. painos). Vaajakoski: Gummerrus.

Moturu, S.T., Johnson, W.G., & Liu, H. (2010). Predictive Risk Modelling for Forecasting High-Cost Patients: A Real-World Application Using Medicaid Data. *International Journal of Biomedical Engineering and Technology*, 3(1), 114–32.

Potilaan lääkärilehti (2018). Potilas vai asiakas?  
<http://www.potilaanlaakarilehti.fi/uutiset/potilas-vai-asiakas/> [Luettu 11.2.2019].

Powers, C.A., Meyer, C.M., Roebuck, M.C. & Vaziri, B. (2005). Predictive modeling of total healthcare costs using pharmacy claims data. *Medical Care*, 43(11), 1065–1072.

Province of British Columbia (2017). Expanded Diagnosis Clusters.  
<http://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/expandeddiagnostic-cluster-dictionary.pdf> [Luettu 5.10.2017].

Rais, S., Nazerian, A., Ardal, S., Chechulin, Y., Bains, N. & Malikov, K. (2013). High Cost Users of Ontario's Healthcare Services. *Healthcare Policy*, 9(1), 44–51.

Reho, T. T., Atkins, S. A., Talola, N., Viljamaa, M., Sumanen, M. P., & Uitti, J. (2018). Frequent attenders in occupational health primary care: A cross-sectional study. *Scandinavian journal of public health*, 47(1), 28–36.

Reid, R., Evans, R., Barer, M., Sheps, S., Kerluke, K., McGrail, K., Hertzman, C. & Pagliccia, N. (2003). Conspicuous Consumption: Characterizing High Users of Physician Services in One Canadian Province. *Journal of Health Services Research & Policy*, 8(4), 215–24.

Reinius, P., Johansson, M., Fjellner, A., Werr, J., Öhlén, G., & Edgren, G. (2013). A telephone-based case-management intervention reduces healthcare utilization for frequent emergency department visitors. *European Journal of Emergency Medicine*, 20(5), 327–334.

Rita, H. (2004). Vetosuhde (*odds ratio*) ei ole todennäköisyyksien suhde. *Metsätieteen aikakauskirja* 2/2004, 207–212.

Smits, F. T., Brouwer, H. J., ter Riet, G., & van Weert, H. C. (2009). Epidemiology of frequent attenders: a 3-year historic cohort study comparing attendance, morbidity and prescriptions of one-year and persistent frequent attenders. *BMC Public Health*, 9(1), 36.

Sosiaali- ja terveysministeriö (2016). Sote-uudistuksen ja aluehallintouudistuksen tilannekatsaus. <https://alueuudistus.fi/documents/1477425/1892966/Sote-uudistus+tilannekatsaus+18.4.2016.pdf/9188e75c-9494-4d26-88e2-4ff70b808120> [Luettu 11.2.2019].

Szumilas M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3), 227-9.

Terveyden ja hyvinvoinnin laitos (2017). Yleistietoa kansantaudeista. <https://www.thl.fi/fi/web/kansantaudit/yleistietoa-kansantaudeista> [Luettu 6.10.2017].

Terveyden ja hyvinvoinnin laitos (2018). Terveydenhuollon menot ja rahoitus 2016. <https://thl.fi/fi/tilastot-ja-data/tilastot-aiheittain/sosiaali-ja-terveydenhuollon-talous/terveydenhuollon-menot-ja-rahoitus> [Luettu 11.2.2019].

Tilastokeskus (2018a). Nuorten määrä uhkaa vähentyä huomattavasti syntyvyyden laskusta johtuen. [https://www.stat.fi/til/vaenn/2018/vaenn\\_2018\\_2018-11-16\\_tie\\_001\\_fi.html?fbclid=IwAR2y8T81QaPN7AtsUwGtqJ\\_qv7Let1Nqw\\_DMfi4\\_kx7M-pWh720h0oiKkeo](https://www.stat.fi/til/vaenn/2018/vaenn_2018_2018-11-16_tie_001_fi.html?fbclid=IwAR2y8T81QaPN7AtsUwGtqJ_qv7Let1Nqw_DMfi4_kx7M-pWh720h0oiKkeo) [Luettu 11.2.2019].

Tilastokeskus (2018b). Väestörakenne väestö iän (1-v.) ja sukupuolen mukaan alueittain 1972 – 2017. Tilastokeskuksen PX-Web-tietokannat. [http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/StatFin\\_\\_vrm\\_\\_vaerak/statfin\\_vaerak\\_pxt\\_004.px/?rxid=a69e3f94-d860-4851-b521-923f78603f4c](http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/StatFin__vrm__vaerak/statfin_vaerak_pxt_004.px/?rxid=a69e3f94-d860-4851-b521-923f78603f4c) [Luettu 23.2.2019].

Tilastokeskus (2018c). Kuntien sosiaali- ja terveystoiminnan vuoden 2017 nettokustannukset pienemmät kuin edellisvuonna. Tilastokeskuksen Kuntatalous 2017 - tiedote. [http://www.stat.fi/til/kta/2017/kta\\_2017\\_2018-09-21\\_tie\\_001\\_fi.html](http://www.stat.fi/til/kta/2017/kta_2017_2018-09-21_tie_001_fi.html) [Luettu 17.2.2019].

Tilastokeskus (2018d). Väestöllisiä tunnuslukuja alueittain 1990 – 2017. Tilastokeskuksen PX-Web-tietokannat. [http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/StatFin\\_\\_vrm\\_\\_vaerak/statfin\\_vaerak\\_pxt\\_007.px/table/tableViewLayout2/?rxid=a4c3d956-5e52-4643-abd3-f8a246c9352f](http://pxnet2.stat.fi/PXWeb/pxweb/fi/StatFin/StatFin__vrm__vaerak/statfin_vaerak_pxt_007.px/table/tableViewLayout2/?rxid=a4c3d956-5e52-4643-abd3-f8a246c9352f) [Luettu 17.2.2019].

Valtioneuvoston kanslia (2018). Ratkaisujen Suomi: Hallituksen toimintasuunnitelma vuosille 2018–2019. Valtioneuvoston julkaisusarja 27/2018. <http://julkaisut.valtioneuvosto.fi/handle/10024/160963>.

Valtion tarkastusvirasto (2017). Paljon palveluita käyttävät ja tarvitsevat asiakkaat perusterveydenhuollossa. <https://www.vtv.fi/app/uploads/2018/05/22105012/paljon-palveluja-tarvitsevat.pdf> [Luettu 11.2.2019].

Zhao, Y., Ash, A.S., Ellis, R.P., Ayanian, J.Z., Pope, G.C., Bowen, B. & Weyuker, L. (2005). Predicting pharmacy costs and other medical costs using diagnoses and drug claims. *Medical Care*, 43(1), 34–43.