

Työpaperi 26/2022

Sote-rahoituksen tarvevakiointi

Päivitys 2022

Holster Tuukka, Haula Taru, Korajoki Merja

Sote-uudistuksessa vastuu sosiaali- ja terveydenhuollon palveluiden järjestämisestä siirretään hyvinvointialueille sekä Helsingin kaupungille ja alueiden rahoitus perustetaan pitkälti tarveperusteiseen valtionrahoitukseen. Tässä työpaperissa kuvataan hyvinvointialueiden sote-rahoituksen tarvevakiointiin ehdotetut tilastolliset mallit, niiden kehitystyö sekä esitetään niiden tuottamat tulokset herkkyystarkasteluineen.

Esipuhe

Sote-rahoituksen tarvevakiointi -tutkimuksen päivitystyö on ollut pitkäjänteinen prosessi, johon on osallistunut mittava joukko asiantuntijoita. Haluamme kiittää kaikkia tärkeää palautetta antaneita sekä kehitystyöhön osallistuneita tahoja.

Työn tekeminen on ajoittunut poikkeukselliseen pandemia-ajanjaksoon, mikä on vaatinut omaksumaan uudenlaisia työskentelytapoja. Korona-aika on kuormittanut voimakkaasti sekä sosiaali- ja terveydenhuollon palvelujärjestelmää että siellä toimivia ammattilaisia; Tämänkin vuoksi olemme erityisen kiitollisia saamastamme merkittävästä asiantuntijatuesta erityisesti sairausluokituksen kehittämistyöhön osallistuneilta erikoissairaanhoidon ja perusterveydenhuollon ammattilaisilta eri sairaanhoitopiireistä ja kunnista. Erityisesti kiitämme tutkimusprofessori emeritus Unto Häkkistä kommentteista.

Tiivistelmä

Tuukka Holster, Taru Haula, Merja Korajoki. Sote-rahoituksen tarvevakiointi, päivitys 2022. Terveyden ja hyvinvoinnin laitos (THL). Työpaperi 26/2022. 51 sivua. Helsinki 2022. ISBN 978-952-343-876-7 (verkkojulkaisu)

Sote-uudistuksessa vastuu sosiaali- ja terveydenhuollon palveluiden järjestämisestä siirretään kuntia suuremmille hyvinvointialueille (21 kpl) sekä Helsingin kaupungille ja alueiden rahoitus perustetaan pääosin tarpeperusteiseen valtionrahoitukseen. Rahoituksen tarvevakiointi perustuu THL:n tutkimukseen sosiaali- ja terveydenhuollon palvelujen käyttöön yhteydessä olevista yksilötason tekijöistä väestössä.

Hyvinvointialueiden rahoituslain lausuntokierroksilla rahoitusmalliin kohdistui kritiikkiä, erityisesti liittyen puutteisiin sairastavuutta rekisteriaineistosta määrittävässä sairausluokituksessa, kustannustiedoissa ja tilastollisten mallinnusten herkkyytarkasteluissa. THL:ssa on jatkettu sote-rahoituksen tarvevakioinnin kehitystyötä. Sairausluokitusta on tarkennettu ja laajennettu yhteistyössä erikoissairaanhoidon ja perusterveydenhuollon asiantuntijoiden kanssa. Myös palvelujen käytön laskennallisten kustannusten määrittämisessä käytettyä palveluluokitusta ja yksikkökustannuksia on tarkennettu. Tilastollisia malleja on päivitetty muutujien osalta erityisesti sairausluokituksen kehitystyön myötä.

Tässä työssä kuvataan päivitettyt tarvevakiointimallit, sairausluokitus ja palveluluokitus sekä käytetyt yksikkökustannukset. Lisäksi avataan tarkemmin myös tilastollisten mallien estimointiprosessia, sekä esitetään tuloksia herkkyytarkasteluineen. Tulokset osoittavat, että kehitystyön myötä väestön palvelutarpeet tulevat paremmin huomioiduksi sote-rahoituksessa.

Avainsanat: terveydenhuolto, sosiaalihuolto, palvelut, soterahoitus, rahoitusjärjestelmät, tarvekriteerit, tarvetekijät, tarvearviot, sairastavuus, kustannusarviot, tilastollinen mallintaminen

Sisällys

Esipuhe.....	2
Tiivistelmä.....	3
Sisällys.....	4
1 Johdanto.....	5
2 Teoreettinen tausta.....	7
2.1 Tarveperusteinen kapitaatorahoitus.....	7
2.2 Mallinnus ja tarpeen arviointi.....	7
3 Aineistot ja menetelmät.....	10
3.1 Aineistot.....	10
3.1.1 Kustannusten määrittäminen.....	10
3.1.2 Sairastavuuden määrittely.....	11
3.2 Osamallien valinta ja kustannusaineisto.....	12
3.3 Mallien validointi ja muuttujavalinta.....	12
3.4 Estimointi- ja arviointiotokset ja tarvekerrointen laskenta.....	13
3.5 Malliperheen valinta.....	14
3.6 Muuttujat.....	15
3.6.1 Kustannukset ja niiden typistys.....	15
3.6.2 Sairausluokitus.....	15
3.6.3 Synnytykset ja keskoset.....	17
3.6.4 Ikä ja sukupuoli.....	17
3.6.5 Taustamaa.....	18
3.6.6 Sosioekonomiset muuttujat.....	18
3.6.7 Työkyvyttömyyseläkkeet.....	18
3.6.8 Julkisten ja yksityisten palveluiden saatavuus.....	18
3.6.9 Interaktiot.....	19
3.6.10 Monisairastavuus.....	20
3.6.11 Perhetilanne ja siviilisäätö.....	20
4 Tulokset.....	21
4.1 Ehdotetut mallit.....	21
4.2 Mallien arviointi.....	30
4.3 Tarvekertoimet.....	35
5 Johtopäätökset.....	39
Liitteet.....	41
Lähteet.....	50

1 Johdanto

Pääministeri Marinin hallituksen sote-uudistuksessa vastuu sosiaali- ja terveydenhuollon palveluiden järjestämisestä siirretään kuntia suuremmille hyvinvointialueille ja Helsingin kaupungille. Laissa hyvinvointialueiden rahoituksesta (617/2021) hyvinvointialueiden rahoitus määrättiin (pääosin) tarveperusteiseen valtionrahoitukseen perustuvaksi. Rahoituksen tarvevakiointi perustuu THL:n tutkimukseen tarvevakioinnista (Häkkinen ym. 2020), missä kansainvälistä kirjallisuutta seuraten tarkasteltiin tilastollisesti sosiaali- ja terveydenhuollon palveluiden käyttöön yhteydessä olevia tekijöitä. Kuitenkin jo lakiesityksen perusteluissa todettiin, että tutkimusta tulee päivittää ennen uuden sote-järjestelmän suunniteltua käynnistymistä vuonna 2023. Myös lain lausuntokierroksella rahoitusmalliin kohdistettiin kritiikkiä. Lisäksi lain 14 §:n 4 momentissa vaaditaan tarkistamaan rahoituksen perustana olevia malleja vähintään neljän vuoden välein.

THL:n tarvevakiointitutkimusta on kritisoitu erityisesti puutteista sairausluokituksessa ja herkkyystarkasteluissa. Näihin kiinnitettiin huomiota myös HUS:in Tempo Leconilta tilaamassa arviointiraportissa (Malmi ym. 2021). Tämän raportin tarkoituksena on kuvata päivitetty tarvevakiointimallit ja sairausluokitus, avata tarkemmin mallien estimointiprosessia sekä esitellä tuloksia ja niiden herkkyystarkasteluja.

Kritiikin kohteena on ollut myös se, etteivät tarvevakiointimallit huomioi palvelunjärjestäjistä riippumattomia palvelujärjestämisen yksikkökustannuksiin vaikuttavia alueellisia eroja olosuhteissa. Olosuhteita huomioidaan muutamissa rahoitusmalliin sisältyvissä määräytymistekijöissä (mm. kaksikielisyys ja saaristisuus), mutta niille annetut painot eivät pohjaudu tutkimukseen. THL on toteuttanut erillisen tutkimuksen olosuhdetekijöiden huomioimisesta rahoituksessa yhdessä Valtion taloudellisen tutkimuskeskuksen (VATT) kanssa (Häkkinen ym. 2022; Lyytikäinen ym. 2022), eikä niihin keskitytä tässä raportissa.

Verrattuna Häkkinen ym. (2020) raporttiin, tarvevakiointimalleihin on tehty seuraavia teknisiä muutoksia:

- Mallit on estimoitu terveyden- ja vanhustenhuollon osalta käyttäen uudempaa, vuoden 2019 palvelukäyttöaineistoa.¹
- Malleissa sairastavuuden määrittämisessä käytettyä sairausluokitusta on päivitetty siten, että vanhaan luokitukseen on lisätty diagnooseja joko vanhoihin sairausluokkiin tai lisätty kokonaan uusia sairausluokkia, ja muutamia vanhaan luokitukseen sisältyneitä sairausluokkia on jaettu pienemmiksi luokiksi. Nykyinen luokitus kattaa 123 sairausluokkaa, kun vanha sisälsi 52 luokkaa. Sairausluokituksen aikaisempaan versioon verrattuna päivitetty luokitus perustuu vahvemmin käytännön työssä toimivien kliinisten asiantuntijoiden suositukseen, sillä siihen saatiin laajasti kommentteja erikoissairaanhoidon ja perusterveydenhuollon toimijoilta.
- Näkövamma huomioidaan nyt WHO:n näkövammaluokituksen mukaisilla muuttujilla henkilöille, joille tieto on saatavilla.
- Kaikissa sairastavuustiedoissa poiminta-ajanjakson pituudeksi on asetettu yksi vuosi, kun Häkkinen ym. (2020) työssä useimpien sairausluokkien tiedot poimittiin kahden vuoden ajalta.
- Malleissa palvelujen käytön kustannusten määrittämisessä käytettyä palveluluokitusta ja yksikkökustannuksia on tarkennettu muun muassa huomioiden laajemmin eri ammattiryhmien yhteystapoja eri palveluissa, sekä JHS-järjestelmän lakattua pyrkien mukailemaan AURA-käsikirjan automaattisen talousraportoinnin palveluluokitusta.
- Mallien määrittelyä on muutettu siten, että mallien estimoinnissa kaikki ennustemuuttujat poimitaan synnytyksiä ja keskosuuksia lukuun ottamatta palvelunkäytön vuotta edeltäviltä vuosilta. Käyttäen riskivakioinnin tavallista terminologiaa, malli on tällöin puhtaasti ns. *prospective*. Käytännön etuna tässä on, että tällöin tarvekertoimiin saadaan aina kokonaisuudessaan käyttöön uusin rekistereistä saatavilla oleva tieto sairastavuudesta.
- Malleihin on muutamia muita muuttujia koskien tehty muutoksia, joista keskeiset ovat:

¹ Sosiaalihuollon avopalveluista ei ole toistaiseksi kansallista rekisteritiedonkeruuta eikä raportissa Häkkinen ym. (2020) kerättyä alueellista aineistoa pystytty hankkimaan tähän raporttiin tuoreemmalta vuodelta, joten sosiaalihuollon avopalvelujen osalta on käytetty edelleen vuoden 2017 palvelukäyttöaineistoa.

- Asuntokunnan käytettävissä olevat tulot kulutusyksikköä kohden on logaritmoitu nyt myös terveydenhuollon mallissa.
- Henkilön taustamaa kuvataan nyt yksinkertaisemmin yhdellä muuttujalla aikaisemman viiden sijaan.
- Siviilisäätö on sisällytetty kolmena muuttujana kaikkiin malleihin.
- Terveyden- ja vanhustenhuollon malleissa ikä- ja sukupuoli on huomioituna aikaisempaa hienojakoisemmalla joukolla muuttujia.
- Poistettu muuttujia, jotka kytkeytyvät vahvasti alueen toimintatapoihin ja hoitopäätöksiin ja voivat siksi periaatteessa sisältää hankalia kannustinvaikutuksia: vanhustenhuollon mallissa muuttuja yksin tai laitoksessa asumiselle ja terveydenhuollon mallissa Kelan korvaamien avohoidon lääkkeiden kustannukset.
- Suoraan monikanavarahoitukseen liittyvät, Kelan sairausvakuutuksen korvaamien kustannusten muuttujat on poistettu tulevan monikanavarahoituksen osittaisen purkamisen vuoksi.

Toteutettuja herkkyystarkasteluja ovat:

- Sairausluokituksen vaihtoehtoiset määrittelyt:
 - Vertailu aikaisemman (Häkkinen ym. 2020) ja tässä raportissa esitellyn sairausluokituksen välillä.
 - Hoitajakäyntien ICD-10 ja ICPC-2 diagnoosien huomioiminen sairauksien määrittelyssä.
 - Englannin NHS:n General & Acute -mallissa käytetyn sairausluokituksen (Chaplin ym. 2016) kokeileminen.
- Mallien estimointivuoden valinnan vaikutus (vuosien 2017 ja vuosien 2018 aineistoilla estimoitujen mallien vertailu vuoden 2019 aineistolla estimoituihin malleihin).
- Hyvinvointialueindikaattorien sisällyttäminen tarjontatekijöiden kontrolloimiseksi.
- Muuttujille taustamaa ja matka-aika erikoissairaanhoidon päivystykseen tehtyjen ns. neutralisointien vaikutus.

Kaikissa herkkyyssanalyseissa tarkastelun kohteena oleva muuttuja on hyvinvointialueen tarvekerroin, joka kertoo alueen asukkaiden keskimääräisen julkisten palveluiden tarpeen suhteessa maan keskimääräiseen palvelutarpeeseen. Herkkyystarkasteluissa verrataan tämän raportin perusmallien perusteella laskettuja tarvekertoimia yllä esitettyihin vaihtoehtoisiiin mallimäärittelyihin perustuviin tarvekertoimiin. Periaatteessa mallivalinnassa on tietysti pyritty siihen, että perusmalleissa tehdyt päätökset ovat teoreettisesti ja asiantuntijankemysten mukaan mahdollisimman perusteltuja ja toisaalta vastaavat empiirisen mallivalinnan kriteereitä. Joissain kohdin mallivalinnassa päädytään kuitenkin tilanteisiin, joissa jää jossain määrin epäselväksi, mikä olisiärkevin määrittely. Herkkyystarkastelut kertovat, missä määrin palvelutarpeen estimointi ja siten rahoitus on vakaata määrittelymuutosten suhteen.

Herkkyyssanalyysien tuloksena huomataan, että tarvekertoimet ovat suhteellisen robusteja vaihtoehtoisille mallivalinnoille. Kuitenkin huomataan, että sairausluokituksen valinnalla on vaikutusta tarvekertoimiin. Lisäksi nähdään, että taustamaan neutralisointia koskevalla normatiivisella valinnalla on suurehko merkitys tuloksiin.

Kuten Häkkinen ym. (2020) toteavat, tarveperusteisen rahoitusjärjestelmän kehittäminen on pitkäjänteistä työtä. Tässä raportissa esitetyt mallit ovat selkeä edistysaskel ja toteutetuilla mallivalinnan raportoinneilla ja herkkyystarkasteluilla on nähdäksemme huomattavasti parannettu rahoitusmallin uskottavuutta. Tilaa jatkokäytölle jää kuitenkin edelleen ja johtopäätöksissä käsittelemme lyhyesti mahdollisia kehityssuuntia.

2 Teoreettinen tausta

2.1 Tarveperusteinen kapitaatorahoitus

Smith (2003, 2007) esittelee kaavaperusteisen rahoituksen (*formula funding*) talusteoreettisesta näkökulmasta. Valtio haluaa asettaa palveluiden järjestämisen alemman hallintotason vastuulle, jakaa rahoituksen niille kriteereillä, jotka vastaavat yhteiskunnassa yleisesti hyväksytyjä normatiivisia käsityksiä, mutta rajoittaa sosiaali- ja terveystalouden kokonaisuuden. Käytännössä tarveperusteisella rahoituksella pyritään turvaamaan kaikille kansalaisille kohtuullisen yhtäläinen pääsy sosiaali- ja terveystalouteen, kuitenkin ehdollisena sille, että samalla kustannuskontrollin kannustimet eivät liikaa heikkenisi. Kustannuskontrollin heikkeneminen voisi johtaa tehotonta palvelutuotantoa ja ylihoitoa suhteessa yhteiskunnan tavoitteisiin. Jos valtiolla olisi täydellinen informaatio parhaista toimintatavoista, tällaisia järjestelyjä ei tarvittaisi: valtio yksinkertaisesti korvaisi optimaalisista toimintatavoista niiden tuotantokustannukset, eikä mistään muusta.²

Tässä tutkimuksessa seurataan kirjallisuudessa tavanomaista tarpeen määrittelyä, missä menneen palvelukäytön (kustannusten) ajatellaan heijastavan yhteiskunnallisia arvostuksia palvelutarvetta koskien. Henkilötasolla toteutuneet kustannukset ovat jonkinlainen funktio $c_i = f(x_i, z_i)$, missä x_i ovat *tarvetekijät* ja z_i ei-kompensoitavat tekijät. Muuttujat z_i sisältävät sekä tarjontatekijät (etäisyys palveluiden piiriin, jonotusajat ja niin edelleen) että muita mahdollisia tekijöitä, joiden tosiasiallisen vaikutuksen palveluiden käyttöön ei nähdä olevan moraalisesti hyväksyttävää vaihtelua tarpeessa. Mahdollinen esimerkki viimeksi mainituista on taustamaa, jos maahanmuuttajat hakeutuvat palveluiden piiriin heikommin kulttuurillista syistä. Vastaavasti myös muita hoidon intensiteetin preferensseihin liittyviä tekijöitä voisi sisällyttää näihin, jos niistä olisi saatavilla tietoa (esimerkiksi jotkut potilaat vaativat parempaa hoitoa ja enemmän tutkimuksia suhteessa muihin vastaaviin tapauksiin). Näissä on kysymys ns. *tyydyttymättömän tarpeen* ongelmista, missä toteutunut palvelukäyttöaineisto ei täysin heijasta yhteiskunnan arvostuksia koskien palveluiden kohdentumista.

Tarvetekijät ovat yhteiskunnassa yleisesti hyväksytyjä palvelutarvetta aiheuttavia tekijöitä, joita julkisen terveydenhuollon tapauksessa ovat erityisesti sairaudet ja toimintakyky, sekä mahdolliset muut tekijät, jotka saattavat hyväksytysti vaikuttaa julkisten palveluiden tarpeeseen (esimerkiksi pääsy samoihin palveluihin toisella sektorilla voi vähentää tarvetta julkiselle terveydenhuollolle). Vahvasti palveluiden tuottajan tai järjestäjän toimintaan liittyvät muuttujat, kuten aikaisempi palvelukäyttö, eivät yleensä ole tarvetekijöitä.³ Vaikka hyvinvointialueiden rahoitusjärjestelmässä kokonaisbudjetti onkin lukittu, voisivat tällaiset muuttujat periaatteessa heikentää kustannuskontrollia ja aiheuttaa eräänlaisen kilpailun alueiden välillä (Barrow, 1986; Smith, 2007). Tällaisessa tilanteessa valtion saattaisi olla pidemmän päälle vaikeaa pitää kiinni kovin tiukasta budjettirajoitteesta.

2.2 Mallinnus ja tarpeen arviointi

Mallinnuksessa halutaan ennustaa tuleva suhteellinen palvelutarve siten, että erotetaan toteutuneesta palvelukäytöstä tarvetekijöistä aiheutunut osuus muista palvelukäyttöön vaikuttavista tekijöistä. Ei-kompensoitava tekijä tulee sisällyttää kontrollitekijäksi, jos se korreloi tarvetekijöiden kanssa. Esimerkiksi jos joillakin alueilla on menneeseen rahoitukseen liittyvistä syistä huomattavasti muuta maata korkeampi palvelutarjonta, voi alueella yleisemmin esiintyvän sairauden kerroin vääristyä ylöspäin. Tällöin syntyy käytännössä tilanne, missä mennyt rahoitus vaikuttaa myös tulevaan rahoitukseen.

Tarvevakioinnissa tarkastellaan siis lineaarista mallia tyyppiä

$$c_i = \alpha + \sum_{k=1}^K \beta_k x_{i,k} + \sum_{j=1}^J \gamma_j z_{i,j} + \varepsilon_i.$$

² Optimaalisuudella tässä tarkoitetaan, että toimintatavat maksimoivat yhteiskunnallista hyötyfunktiota, eli ovat kustannusvaikuttavia. Todellisuudessa valtiolta tietysti puuttuu tarkka tieto paitsi kustannusvaikuttavista toimenpiteistä, myös tarkasta yhteiskunnallisesta hyötyfunktiosta.

³ Esimerkiksi Hollannissa kuitenkin käytetään suoraan mennyttä palvelukäyttöä kuvaavia tekijöitä (van Kleef ym. 2018).

missä henkilön i palvelukäyttöä selitetään tarvetekijöillä ja ei-kompensoitavilla tekijöillä. Mallin estimoinnissa saadaan mallin kertoimille estimaatit $\hat{\alpha}$, $\hat{\beta}_k$ ja $\hat{\gamma}_j$. Regressiokerroin $\hat{\beta}_k$ kertoo, miten paljon korkeampaan palvelukäyttöön tarvetekijä x_k on yhteydessä.⁴ Vakiotermin α on kaikille henkilöille sama ja se voidaan tulkita kiinnitetyksi asukaskohtaiseksi eräksi (Smith, 2007).⁵

Mallin antama ennuste henkilön i palvelukäytölle on

$$\hat{c}_i = \hat{\alpha} + \sum_{k=1}^K \hat{\beta}_k x_{i,k} + \sum_{j=1}^J \hat{\gamma}_j z_{i,j}.$$

Koska ei-kompensoitavista tekijöistä johtuvaa vaihtelua palvelukäytössä ei haluta kompensoida, henkilön tarve määritellään kuitenkin

$$\tau_i = \sum_{k=1}^K \hat{\beta}_k x_{i,k} + \sum_{j=1}^J \hat{\gamma}_j \bar{z}_j,$$

missä muuttujat z on *neutralisoitu* asettamalla ne koko maan keskiarvoonsa (Schokkaert ym. 2004).⁶ Vastavasti alueen m tarve on summa yli alueen asukkaiden tarpeen:

$$\tau_m = \sum_i^{n_m} (\sum_{k=1}^K \hat{\beta}_k x_{i,k} + \sum_{j=1}^J \hat{\gamma}_j \bar{z}_j),$$

missä n_m on alueen m väkiluku. Alueen keskimääräinen tarve voidaan luonnollisesti laskea myös käyttäen aluetason tietoja, kun regressiokerroimet on estimoitu:

$$\frac{\tau_m}{n_m} = \sum_{k=1}^K \hat{\beta}_k \bar{x}_{m,k} + \sum_{j=1}^J \hat{\gamma}_j \bar{z}_j,$$

missä $\bar{x}_{m,k}$ on alueen m keskiarvo tarvetekijälle $x_{i,k}$.

Näin määritellyn tarpeen voi nähdä henkilön ja alueen absoluuttisena palvelutarpeena suhteessa kyseisen vuoden palvelujärjestelmään. Keskeisenä mielenkiinnon kohteena on kuitenkin suhteellinen tarve, mikä voidaan määrittellä *tarvekertoimina*:

$$\vartheta_{m,t} = \frac{\tau_{m,t}/n_{m,t}}{\tau_t/n_t},$$

mikä kertoo alueen m asukkaiden keskimääräisen palvelutarpeen suhteessa maan keskimääräiseen vuonna t .⁷ Kiinnitettyä budjettia B voidaan jakaa alueittaisiksi budjeteiksi B_m tarvekerrointen perusteella seuraavasti:

$$B_{m,t} = \vartheta_{m,t} n_{m,t} \frac{B_t}{n_t},$$

⁴ Yksinkertaisuuden vuoksi tästä on tiputettu indeksointi ajanhetkeen (vuoteen) t . On kuitenkin hyvä muistaa, että kyseessä ovat vuoden t palvelukäyttö $c_{i,t}$ ja siihen perustuvat, vuoden t rahassa mitatut regressiokerroimet $\hat{\beta}_{k,t}$.

⁵ Vakiotermin α tarkoittaa, mitä palvelukäytön ennustettaisiin olevan henkilöllä, jolla kaikkien selittävien muuttujien arvot ovat nolla. Esimerkiksi tässä raportissa suositellussa terveydenhuollon mallissa se vastaa karkeasti puhuen työtöntä tai työvoiman ulkopuolella olevaa, matalasti koulutettua ja taustaltaan suomalaista, naimisissa olevaa keski-ikäistä miestä, jolla ei ole tuloja eikä kirjattuna sairauksia. Jos mallissa olisi vain sairauksia selittävinä muuttujina, voisi vakiotermin tulkita olevan (akuuteista sairauksista, seuloista yms. johtuva) palvelukäytön perustaso, mikä osaltaan mittaisi mallin kykyä selittää toteutunutta palvelukäyttöä. Muiden kuin sairausmuuttujien sisällyttäminen malliin monimutkaistaa tulkintaa.

⁶ Tai johonkin muuhun, kaikille samaan arvoon. Keskiarvo on vakuutusperusteisissa järjestelmissä tavanomainen ja luonteva valinta, sillä se takaa budjetin pysymisen vakiona (ennustettu keskimääräinen kustannus ei muutu).

⁷ Tarvekertoimet voidaan laskea myös käyttäen ns. painotettuja kertoimia. Muuttujan k painotettu kerroin \hat{p}_k suhteuttaa regressiokerroimen koko maan keskiarvoiseen tarpeeseen:

$$\hat{p}_k = \frac{\hat{\beta}_k}{\frac{1}{n} \sum_{i=1}^n (\hat{\alpha} + \sum_{k=1}^K \hat{\beta}_k x_{i,k} + \sum_{j=1}^J \hat{\gamma}_j \bar{z}_j)} = \frac{\hat{\beta}_k}{\tau_t/n_t},$$

joten se kertoo, kuinka paljon korkeampiin lisäkustannuksiin (yli vakiotermin) tarvetekijä k on yhteydessä suhteessa maan keskiarvoiseen tarpeeseen. Ajamalla regressio ilman vakiotermin $\hat{\alpha}$ saadaan painokerroin jokaiselle ikäryhmälle. Alueen m tarvekerroin eli asukkaiden keskimääräinen tarve suhteutettuna maan keskiarvoon on painotettuja kertoimia käyttäen esitettyä siis

$$\vartheta_{m,t} = \sum_{k=1}^K \hat{p}_k \bar{x}_{m,k,t}.$$

Painotetut kertoimet on esitettyinä liitteessä.

eli alueen rahoitus on tarvekerroin kertaa väkiluku kertaa koko maassa sosiaali- ja terveydenhuoltoon jaettava rahoitus per henkilö. Tällöin kiinnitetty budjetti B määritellään muilla keinoin perustuen arvioihin koko maan tarpeen tulevasta kehityksestä ja yllä esitetystä tarvemallinnuksessa pyritään vain mittaamaan suhteellista tarvetta yhden vuoden poikkileikkauksessa.⁸

Yllä olevassa tarpeen määrittelyssä tarvetekijän kerroin $\hat{\beta}_k$ on sama kaikilla alueilla. Sairauden tapauksessa regressiokertoimen tulisi heijastaa maan keskimääräisellä laatutasolla odotettavia lisäkustannuksia kyseisen sairauden hoidosta, jolloin kannustin olisi diagnosoida ja hoitaa potilas keskimääräisellä laatutasolla. Alueellisesta järjestäjästä riippumattomista syistä korvaus $\hat{\beta}_k$ ei kuitenkaan kaikilla alueilla turvaa mahdollisuutta samaan palvelutasoon: alueelliset olosuhteet voivat vaikuttaa sosiaali- ja terveystalouden kustannusfunktioon.

Palvelutuotannon yksikkökustannukset voivat olla alueella korkeammat esimerkiksi hankalampien työmarkkinaolosuhteiden vuoksi. Jollain alueilla voidaan myös joutua operoimaan väistämättä matalammilla skaalaeduilla, jos palvelut halutaan kuitenkin väestölle myös tällä seudulla tarjota. Englannissa on käytössä ns. pienten sairaaloiden väistämättömään tehottomuuteen liittyvä korvaus (ACRA, 2015). Ongelmana näiden tekijöiden huomioimisessa on tietystikin vaikeus tunnistaa, missä määrin tehottomuus on väistämätöntä skaalaedun puutetta ja missä määrin alueen heikkoja toimintatapoja.⁹ Olosuhdetekijöiden huomioimisesta sosiaali- ja terveydenhuollon rahoituksesta löytyy tarkemmin raporteista Häkkinen ym. (2022) ja Lyytikäinen ym. (2022).

Vastaavia tilastollisia malleja käytetään laajalti myös terveystalouden markkinoiden regulaatiossa tasamaan vakuutusmaksujen välistä riskiä ja vähentämään kannustimia asiakkaiden valikointiin (Ellis ym. 2017; Juhnke ym. 2016). Motivaatio terveystalouden markkinoiden riskivakiointimalleissa on sikäli hyvin samankaltainen julkisten järjestäjien rahoituksen kanssa, että malleilla pyritään turvaamaan kaikille pääsy terveystalouden palvelujen piiriin. Koska julkinen järjestäjä ei kuitenkaan voi valikoida asiakkaitaan, olennaiseksi eroksi jää se, ettei julkisten järjestäjien rahoituksessa yksittäisten potilasryhmien yli- tai alikompensaatiosta aiheutuvat valikoitumiskannustimet nouse kysymykseksi. Käytännössä eri maissa käytössä olevat tarvevakioinnin menetelmät vaihtelevat jo maiden järjestelmien eroavaisuuksien vuoksi, ja erityisesti sisällytetyssä sairastavuustiedossa on suuria eroja (Juhnke ym. 2016). Raportista Häkkinen ym. (2020) löytyy tarkempaa esittelyä muutamien maiden rahoitusmalleista.

⁸ Huomioiden ennustetut muutokset erityisesti väestörakenteessa. Kts. Some-malli.

⁹ Eli onko kyse siitä, että samalta tuotantofunktiolta joudutaan poimimaan matalamman tuotos-panos-suhteen piste, vai siitä, että tehottomista toimintatavoista johtuen alueella tuotantofunktio on matalammalla kuin muualla maassa.

3 Aineistot ja menetelmät

3.1 Aineistot

Tutkimuksen tiedonkeruuseen on kuulunut laaja joukko valtakunnallisia rekistereitä, jotka sisältävät yksilötason palvelunkäyttö- ja kustannustietoja. Lisäksi niiltä osin, joille rekisteritietoa ei ole saatavilla, on tietopohjaa laajennettu täydentävillä aineistoilla, kuten Tilastokeskuksen tarjoamilla FOLK-moduuleilla, ja erillistiedonkeruilla, kuten sosiaalihuollon avopalveluaineistot suoraan alueilta tai kunnilta.

Tutkimuksessa käytetyt rekisterit ja aineistokokonaisuudet ovat:

- THL
 - Sosiaali- ja terveydenhuollon hoitoilmoitusrekisteri
 - Syntyneiden lasten rekisteri
 - Toimeentulotukirekisteri
 - Syöpärekisteri
 - Näkövammarekisteri
 - Lastensuojelurekisteri
- Kela
 - Yksityislääkärissäkäyntien kustannukset ja korvaukset
 - Yksityisen hammashoidon kustannukset ja korvaukset
 - Yksityisen tutkimuksen ja hoidon kustannukset ja korvaukset
 - Korvattujen lääkkeiden kustannukset ja korvaukset ATC-ryhmissä
 - Matkojen ja yöpymisten kustannukset ja korvaukset
 - Toimeentulotuen määrä
 - Vammaistukea saaneiden diagnoosit
 - Sairauspäivärahaajaksojen diagnoosit
 - Tarkasteluaikana voimassa olevat työkyvyttömyyseläkkeet ja diagnoosit
 - Henkilön tarkasteluaikana voimassa olevat lääkekorvausoikeudet
- DVV: Suomessa asuvien henkilöiden syntymäpäivä, postinumero, siviilisäätö ja kuolinpäivä
- Eläketurvakeskus: Tarkasteluaikavälillä voimassa olleet työkyvyttömyyseläkkeet ja diagnoosit
- Tilastokeskus:
 - FOLK-valmisaineistokokonaisuudet: Perustieto, Työssäkäynti, Tutkinnot, Tulotieto, Asuntokunta, Perhe
 - Joukko räätälöityjä muuttujia liittyen mm. henkilön kansalaisuuteen, koulutukseen sekä asuntokunnan tuloihin
- Sosiaalihuollon avopalveluista ei ole koko väestön kattavaa rekisteritietoa. Siksi niiden osalta tehtiin suora tiedonkeruu seuraavilta alueilta ja kunnilta:
 - Oulun kaupunki
 - Eksote
 - Päijät-Hämeen hyvinvointiyhtymä
 - Helsinki

3.1.1 Kustannusten määrittäminen

Palvelujen käytön todellisista kustannuksista on rekistereissä vain vähän yksilötason tietoa, joten suuri osa kustannuksista perustuu laskennallisiin kustannusarvioihin. Terveys- ja sosiaalihuollon palvelujen käytön osalta kustannusarvot muodostetaan ryhmittelemällä yksilötason rekisteriaineisto palveluluokkaryhmiin, joille puolestaan on määritetty kustannus erillisten liitetaulukkojen 1–3 mukaisesti.

Häkkinen ym. (2020) työssä kustannusaineisto mukaili JHS 200 Kuntien ja kuntayhtymien palveluluokista kuntien vastuulla olevista terveys- ja sosiaalipalveluista. JHS-järjestelmä on sittemmin lakannut, minkä

vuoksi työssä käytettyä palveluluokitusta päivitettiin kuntien ja kuntayhtymien automatisoidun talousraportoinnin käsikirjan (AURA) palveluluokituksen vuoden 2021 version mukaiseksi. Palveluluokitukseen tehtiin edelleen joitain täsmennyksiä: esimerkiksi avohoidon palvelujen osalta palveluokitusta jaettiin yksityiskoh- taisemmaksi lisäämällä tasoja ammattiryhmille ja yhteystavoille, jolloin kustannsluokittelu tarkentui, ja pe- rusterveys- ja terveydenhuollon vuodeosastohoitoon osalta tarkastelu jaettiin yksityisiin ja julkisiin hoitojaksoihin kar- kean palveluntuottajatunnisteen perusteella tehdyn jaon avulla. Huomionarvoista on, että AURA-palvelu- luokitusta kehitetään edelleen ja vuoden 2022 versioon on ehdotettu muutoksia, jotka poikkeavat tässä työssä käytetystä esimerkiksi somaattisen erikoissairaanhoidon, suun terveydenhuollon ja ikääntyneiden palve- luissa.

Palvelujen käytöstä yksilötasolla muodostuva laskennallinen vuosikustannus perustuu pääasiassa keski- määräisiin kansallisiin yksikkökustannuksiin (Mäklin & Kokko, 2021). Käytetyt yksikkökustannukset on listattuna liitetaulukoissa 1–3. Somaattisen erikoissairaanhoidon palvelujen osalta on käytetty Sairaaloiden tuottavuus -tilaston DRG-ryhmittäisiä kustannuksia.

3.1.2 Sairastavuuden määrittely

Sairastavuuden määrittämisessä käytetään laajasti eri yksilötason rekisteritietolähteitä. Tällä pyritään varmis- tamaan, että henkilön sairastavuus tulee havaituksi, vaikka tällä ei olisi määrättyä palvelujenkäyttöä tarkas- teluajavälillä. Tiedot poimitaan yksilötasolla eri tietolähdekokonaisuuksista, jotka on esitetty alla olevassa taulukossa.

Taulukko 1. Sairastavuuden määrittelyssä käytetyt tietolähteet

Tietolähde	Tietolähteen selite (rekisteri tai rekisterin osa, josta tiedot poimittu)
TL1a	THL: Terveydenhuollon hoitoilmoitusrekisteri, laitoshoidot
TL1b	THL: Sosiaalihuollon hoitoilmoitusrekisteri
TL2	Kela ja ETK: Vammaustukijaksot ja kuntoutus sekä työkyvyttömyyseläkkeet
TL3	THL: Terveydenhuollon hoitoilmoitusrekisteri, erikoissairaanhoidon avohoitokäynnit
TL4	THL: Perusterveydenhuollon hoitoilmoitusrekisteri, käynnit lääkärin vastaanotolla
TL5	Kela: Sairaspäiväraha- ja sairauslompakot
TL6	Kela: Lääkkeiden korvausoikeudet
TL7	THL: Syöpärekisteri
TL8	THL: Syntymärekisteri
TL9	THL: Näkövammarekisteri
TL11	Kela: Korvatut lääkkeet

Kriteerit, joilla henkilöllä määritellään olevan tietty sairausluokituksen sairaus, vaihtelevat käytettävän tietolähteen ominaisuuksien mukaan. Esimerkiksi hoitoilmoitusrekisteriin liittyvistä tietolähteistä sairasta- vuus poimitaan tyypillisesti ICD-10-diagnoosin perusteella. Diagnoositiedoista huomioidaan sekä pää- että sivudiagnoosit, lisäksi myös pitkäaikaisdiagnoosit. Syy- ja oirediagnoosit huomioidaan itsenäisinä koodeina eli ICD-10-luokituksessa lueteltuja oire-syy-pareja ei vaadita siinä muodossa; oire- tai syykoodina ilmoitettu diagnoosi määrittää sairauden. ICD-9-koodit tarkistetaan ainoastaan Kelan työkyvyttömyyseläkeaineistosta, jossa vanhaa luokitusta esiintyy verrattain paljon pitkien (ennen ICD-10-luokituksen käyttöönottoa alkaneiden) eläkejaksoiden vuoksi. Diagnoosikoodien lisäksi sairausluokitus huomioi muita eri rekistereille tyypilli- siä luokituksia, kuten Kelan lääkekorvausoikeusnumeroita ja lääkkeiden ATC-koodeja.

Eri lähteistä poimittua sairaustietoa kohdellaan siinä mielessä yhdenvertaisena, että jos henkilöltä löytyy sairausluokkaan sisältyvä diagnoosi mistä tahansa sen määrittelyssä käytetystä lähteestä, todetaan sairaus

olemassa olevaksi ja kyseisen sairauden sisältävä sairausluokkaa kuvaava muuttuja saa henkilöllä arvon 1. Tarkat sairausluokkiin sisältyvien sairauksien poimintaehdot ja käytetyt tietolähteet on esitetty erillisessä liitetaulukossa 4.

3.2 Osamallien valinta ja kustannusaineisto

Tässä raportissa on noudatettu aikaisempaa käytäntöä (Häkkinen ym. 2020), missä kokonaistarpeen arviointi on jaettu kolmeen osamalliin, terveydenhuoltoon, vanhustenhuoltoon ja sosiaalihuoltoon (liitetaulukoissa 1–3). Tämä on osittain tarpeellista aineistorajoitteiden vuoksi: koska sosiaalihuollon avopalveluiden käytöstä ei ole kansallista tiedonkeruuta, sitä ei voida sisällyttää koko väestön kattavaan mallinnukseen. Vastaavasti on toimittu esimerkiksi Englannissa, kun eri palvelukokonaisuuksien mallinnuksessa käytettävissä olevat aineistot (selitettävä muuttuja tai ennustemuuttujat) ovat olleet hyvin erilaisia, tai kun valtionosuuksia on jaettu eri organisaatioitasoille (Clinical Commissioning Groupsille, CCG:ille tai kunnille).

Periaatteessa malleja voisi jakaa osamalleihin myös perustuen väestön osajoukkoihin. Chaplinin ym. (2016) G&A-malli on ikästratifikoitu (0–14-vuotiaille, 15–64-vuotiaille ja yli 64-vuotiaille omat mallinsa) perustuen näin saatavaan parempaan ennustekykyyneen. Aikarajoitteista johtuen tässä tutkimuksessa ei kokeiltu vastaavia stratifikaatioita. Lisäksi niiden hintana on tarveperusteisen rahoituksen heikentyvä läpinäkyvyys.

3.3 Mallien validointi ja muuttujavalinta

Tarvemalleja ei voida pitää puhtaina ennustemalleina, joissa kiinnostuksen kohteena olisi yksinkertaisesti ennustaa parhaalla mahdollisella tavalla henkilöiden (tai alueiden) tulevaa palvelukäyttöä. Puhtaassa ennustemallissa ei pyritä asettamaan yksittäisten selittävien muuttujien ja vastemuuttujan väliselle tilastolliselle yhteydelle vahvaa kausalitulkintaa (Shmueli, 2010).

Tavallinen menettely ennustemallin luomisessa on jakaa käytettävissä oleva otos estimointi- ja validointiotoksiin (Kuhn ja Johnson, 2013). Estimointiotoksessa estimoidaan mallin parametrit ja validointiotoksessa arvioidaan, miten hyvin estimoitu malli ennustaa. Menettelyä käytetään, sillä ennustemallille tarvitaan arvio ennustevirheestä eli mallin antaman ennusteen keskimääräisestä poikkeamasta ennustettavan muuttujan todellisesta arvosta. Ennustevirhe täytyy arvioida mallin estimointiotoksen ulkopuolisesta aineistosta (*out-of-sample error*), koska kun on saatavilla suuri määrä potentiaalisia ennustemuuttujia, on malli helppo ylisovittaa (*overfit*) siihen otokseen, missä malli estimoidaan. Tällöin malli kuvaa hyvin estimointiotoksen sisäistä vaihtelua vastemuuttujassa mutta heikosti sen ulkopuolella (mihin mallia lopulta halutaan käyttää). Riippuen mallinnusprosessin monimutkaisuudesta, voi siihen sisältyä moniakin eri estimointi- ja validointiotoksia.¹⁰ Vaihtoehtoisia tapoja arvioida ennustevirhettä erillisessä validointiotoksessa on erityisesti niin sanottu ristiinvalidointi, missä otos jaetaan useisiin osaotoksiin, joista vuorollaan kukin on kerran validointiotos (*k-fold cross-validation*) sekä yksinkertainen satunnaistettu jako estimointi- ja validointiotoksiin.

Vastaavasti on toimittu yleisesti myös riski- ja tarvevakiointimalleissa, esimerkiksi Englannissa (Chaplin ym. 2016; Anselmi ym. 2019). Koska malleilla pyritään ennustamaan tulevaa palvelutarvetta, tulee mallien toimivuutta arvioida erillisessä validointiotoksessa lasketulla ennustevirheellä.¹¹ Tässä raportissa esitetyjen mallien estimoinnissa on seurattu samaa menettelyä. Koska havaintomäärä on hyvin suuri (terveydenhuollon mallissa havaintoja on noin 5,6 miljoonaa, sosiaali- ja vanhustenhuollon malleissa molemmissa yli miljoona), on perusteltua käyttää laskennallisesti kevyempää yksinkertaista jakoa estimointi- ja validointiotoksiin (Kuhn ja Johnson, 2013).

Kuhn ja Johnson (2020) suosittavat, että ennustemallien muuttujavalinnan tulisi pohjautua sekä sisällölliseen asiantuntemukseen että empiirisiin kriteereihin. Lineaarisen mallin ennustekyky on pienellä otoskoolalla herkkä ylimääräisille ennustemuuttujille, joilla ei ole yhteyttä vastemuuttujaan (Kuhn & Johnson, 2020). Tässä tutkimuksessa ongelma ei muodostune huomattavaksi, sillä havaintomäärä on hyvin suuri suhteessa

¹⁰ Koneoppimismalleissa tarvitaan yleisesti ottaen niin sanottujen hyperparametrien valintaan oma jakonsa estimointi- ja validointiotoksiinsa, minkä lisäksi vaihtoehtoisten mallien toimivuutta testataan erillisessä testiotoksessa.

¹¹ Oletuksena on, että validointiotos kuvaa hyvin sen ajanhetken todellisuutta, jota mallilla pyritään ennustamaan (eli että palvelukäytön ja ennustemuuttujien yhteisjakauma on vakio yli ajan).

muuttujien määrään. Muuttujavalinta on väistämättä jossain määrin subjektiivista ja pohjautuu ennakoarvioihin tekijöistä, jotka vaikuttavat tulevaan palvelukäyttöön. Lisäksi muuttujien funktionaalinen yhteys vasteeseen voi olla monimutkainen.

Käytännössä mallivalinnassa noudatettiin seuraavaa menettelyä. Lähtökohdaksi valittiin teoreettisesti, kirjallisuuden ja asiantuntijanäkemyksen perusteella mallinnuksen tavoitteiden kannalta mielekkäimmät muuttujat, jotka on esitetty perusteluineen alla osiossa 3.6. Asiantuntijaperusteinen tieto voi silti jättää tarkan funktionaalisen muodon varsin epäselväksi (tulisiko tulot huomioida logaritmisella skaalalla vai polynomina?). Siksi malleja vertaillaan keskeisten muuttujien vaihtoehtojen määritysten suhteen empiirisesti. Vaihtoehtoihin muuttujamäärityksiin pohjautuvia malleja verrataan ennustekyvyn perusteella osiossa 4.2. Lisäksi raportoidaan myös tarvekerrointen herkkyysohjelmille keskeisille vaihtoehtoisille määrityksille osiossa 4.1.

Muuttujavalinnan prosessi eroaa jossain määrin NHS:n G&A-mallin mallinnuksessa käytetystä menettelystä. Chaplin ym. (2016) käyttävät yksilötason validointiotosta vain laskemaan sovittelevyyden kriteereitä muutamien muuttujien esivalikointiin (esim. monisairastavuusmuuttujan monimutkaisuus, käyntitiedosta poimittujen diagnoosien määrä). Lisäksi Chaplin ym. sisällyttävät aluetason tarve- ja tarjontamuuttujia,¹² jotka valitaan tilastollisella mallivalintaprosessilla (tilastolliseen merkitsevyyteen perustuen askelittain tiputtaen). Lopullisten mallivaihtoehtojen välillä Chaplin ym. valitsevat perustuen ennustevirheisiin yleislääkärin vastaanottotason (GP-practice) validointiotoksessa. Tässä raportissa haluttiin pysyä yksinkertaisemmassa ja läpinäkyvämmässä prosessissa.

Oman ongelmansa tuottavat muuttujat, joiden regressiokerroin poikkeaa vahvasti ennako-oletuksesta. Erityisesti sairauksien odotetusti olevan yhteydessä korkeampaan tulevaan terveydenhuollon palvelujen käyttöön ja kustannuksiin, siis saavan positiivisen kertoimen terveydenhuollon mallissa. Sairauden negatiivisen kertoimen tilanteessa voi olla kysymys tyydyttymättömän tarpeen ongelmasta, eli henkilöt eivät saa palveluita. Toisaalta joidenkin sairauksien negatiiviset kertoimet ovat yksinkertaisesti seurausta siitä, että palvelukäyttö on jaettu kolmeen osamalliin: käytännössä jotain sairautta sairastava henkilö voi saada kaiken tarvitsemansa palvelun yhdellä sektorilla (esim. vammaisten asumispalveluissa) jolloin hän ei tarvitse vastaavia palveluja toisella sektorilla (esim. ikääntyneiden asumispalveluissa). Tässä tutkimuksessa tyydyttiin jättämään malleista pois sairaudet, jotka saivat negatiivisen regressiokertoimen. Herkkyystarkasteluissa kuitenkin raportoidaan tarvekertoimet myös koko sairausluokituksen sisältävistä täysmalleista.

3.4 Estimointi- ja arviointiotokset ja tarvekerrointen laskenta

Mallien estimoinnissa ja arvioinnissa käytetyt otokset ovat perustapauksessa¹³ seuraavat.

1. Terveydenhuolto. Vuoden 2019 palvelukäyttö ja edeltävien vuosien tarvetekijät (poikkeuksena synnytykset, jotka ovat vuodelta 2019) henkilöille, jotka asuivat vakinaisesti Suomessa vuoden 2018 lopussa + vuonna 2019 syntyneet keskoset.
 - 1.1. Estimointiotos, satunnainen 80 prosenttia.
 - 1.2. Validointiotos, satunnainen 20 prosenttia.
2. Sosiaalihuolto. Vuoden 2017 palvelukäyttö ja edeltävien vuosien tarvetekijät henkilöille, jotka asuivat vakinaisesti Suomessa vuoden 2016 lopussa.
 - 2.1. Estimointiotos, satunnainen 80 prosenttia.
 - 2.2. Validointiotos, satunnainen 20 prosenttia.
3. Vanhustenhuolto. Vuoden 2019 palvelukäyttö ja edeltävien vuosien tarvetekijät 65 vuotta täyttäneille henkilöille, jotka asuivat vakinaisesti Suomessa vuoden 2018 lopussa.
 - 3.1. Estimointiotos, satunnainen 80 prosenttia.
 - 3.2. Validointiotos, satunnainen 20 prosenttia.

¹² Aluetason tarve- ja tarjontamuuttujia Suomeen verrattuna heikommasta yksilötason aineistoistaan johtuen.

¹³ Lisäksi esitetään herkkyystarkasteluna tarvekertoimet, joissa on käytetty terveyden- ja vanhustenhuollon osalta vuosien 2017 ja 2018 palvelukäyttöaineistoja (ja vastaavasti aikaisempien vuosien tarvetekijöitä).

Mallien vertailua varten mallit estimoidaan estimointiotoksessa ja niiden ennustevirheet lasketaan validointiotoksessa (tulokset osiossa 4.2).¹⁴

Tarvekerrointen laskentaa varten mallit estimoidaan uudestaan mutta käyttäen otoksia 1, 2 ja 3 kokonaisuudessaan (rahoitukseen ehdotettujen perusmallien osalta tulokset osiossa 4.1). Näin estimoitujen mallien regressiokertoimia käyttäen voidaan arvioida palvelutarve myös muiden vuosien aineistoilla. Tuoreimpana kokonaisuudessaan saatavilla olevana aineistona tässä raportissa keskitytään esittämään tarvekertoimet käyttäen vuoden 2019 lopun väestöä ja muuttujatietoja vuodelta 2019 (tulokset osiossa 4.3). Tarvekerrointen laskennassa myös synnytysten ja keskosten tiedot ovat vuodelta 2019. Vertailun vuoksi esitetään myös tarvekertoimet käyttäen vuosien 2017 ja 2018 väestöjä.

Riskivakioinnin kirjallisuudessa puhutaan yleisesti ns. *prospective* ja *concurrent* -malleista. Oletetaan, että ollaan kiinnostuneita jakamaan rahoitusta vuodelle $t+1$ ja tuoreimmat saatavilla olevat tiedot ovat vuodelta t . Oletetaan, että mallin estimoinnissa käytetään vuoden t palvelunkäyttötietoja. Vuosi t on tuorein, jolta aineisto on saatavilla ja siten oletettavasti lähimpänä vuoden $t+1$ palvelukäyttöä, tai se voi muutoin olla tärkeimmäksi koettu vertailuvuosi (periaatteessa muukin valinta olisi mahdollinen). *Prospective*-mallissa estimointivaiheessa ennustemuuttujissa käytetään vain vuoden $t-1$ ja sitä edeltäviä tietoja. Kyseessä on tällöin määrittelyltään sikäli puhdas ennustemalli, että tulevaa palvelutarvetta ennustetaan vain nykyhetkessä saatavilla olevilla tiedoilla. *Concurrent*-mallissa myös selittävien muuttujien¹⁵ tiedot ovat vuodelta t , eli samalta ajanjaksolta kuin palvelukäyttö. *Concurrent*-malleja käytetään yleisesti tapauksissa, joissa kompensoidaan jälkikäteen vuonna t tarvekiöistä aiheutuneista kustannuksista.

Tämän tutkimuksen malleissa on kysymys lähes puhtaasti *prospective*-mallinnuksesta, missä nykyisellä sairastavuudella ennustetaan tulevaa palvelutarvetta. Kuitenkin mallinnuksessa on myös *concurrent*-mallin elementtejä: synnytysten ja keskossuukien suhteen tehdään edelleen se poikkeus, että ne sisällytetään mallinnukseen samalta vuodelta palvelukäytön kanssa. Avoimeksi ongelmaksi mallissa jää, että saatavilla olevien tarvetekijätietojen (tuoreimpana kokonaisuudessaan vuoden 2019 aineisto) ja vuoden, jolle suhteellinen tarve tulisi ennustaa (2023), on huomattava ajallinen ero. Lisäksi tämän tyyppisellä tarvevakiointimallilla ei pystytä ennustamaan sitä uutta sairastavuutta, mikä tulevaisuudessa aiheuttaa palvelutarvetta.

Vuoden 2020 palvelukäyttötieto (sosiaalihuollon avopalveluita lukuun ottamatta) olisi ollut saatavilla tätä tutkimusta varten. Sitä päätettiin kuitenkin olla käyttämättä siitä syystä, että Covid-19 pandemia on vaikuttanut sosiaali- ja terveydenhuollon palvelujen käyttöön ja voimavarojen kohdentamiseen palvelujen välillä ja niiden sisällä (Kestilä ym. 2022; Kestilä ym. 2021; Rissanen ym. 2020). Pandemian hoito, siihen varautuminen, jäljitys- eristys- ja rokotustoimet sekä ihmisten oma varovaisuus on aiheuttanut häiriöitä muiden sairauksien ja tilojen hoitoon. Jos tarvetekijöiden regressiokertoimet estimoitaisiin tällaisella aineistolla, se heijastaisi tulevaan rahoitukseen tämän poikkeuksellisen ajanjakson palvelukäytön (van Kleef ym. 2021). Siksi käytettiin vuotta 2019 tuoreimpana saatavilla olevana, ”normaalin palvelukäytön” vuotena.

3.5 Malliperheen valinta

Kuhn ja Johnson (2013) esittävät, että ennustemallia kehittäessä tulisi etsiä vertailukohdaksi paras malli, joka kyseessä olevaan ongelmaan pystytään olemassa olevalla datalla kehittämään. Ennustekyvyltään paras mahdollinen malli on luultavasti hyvin monimutkainen, epälineaarinen malli. Tämä voi olla useissa sovelluksissa ongelma, sillä ennustemallilta voidaan vaatia myös tulkinnallista helppoutta. Kun tunnetaan ennustekyvyltään paras mahdollinen malli, pystytään valistuneesti valitsemaan yksinkertaisin mahdollinen malli, joka saavuttaa riittävän hyväksi koetun ennustekyvyn.

Koneoppimismenetelmien hyödyntämistä on tarkasteltu erillisessä tutkimuksessa (Holster, Ji & Marttinen, 2022). Tutkimuksessa mukailtiin raportissa Häkkinen ym. (2020) tehtyjä mallispesifikaatio- ja muuttujavalintoja, tietyin poikkeuksin. Siinä havaittiin, että *random forest* ja *gradient boosting machine* pystyvät

¹⁴ Tarvevakiointimallin virhe on $e_i = \hat{c}_i - c_i = \sum_{k=1}^K \hat{\beta}_k x_{i,k} + \sum_{j=1}^J \hat{\gamma}_j z_{i,j} - c_i$ ja tämä voidaan laskea regressiokertoimia $\hat{\beta}_k$ käyttäen sekä mallin estimointiotoksessa että havainnoille, joita ei käytetty estimoinnissa.

¹⁵ *Concurrent*-mallin tapauksessa on hieman luontevampaa puhua selittävästä muuttujista kuin ennustemuuttujista, kun taas *prospective*-mallissa ennustemuuttuja on yleisemmin käytetty termi.

selkeästi korkeampaan ennustetarkkuuteen verrattuna tavanomaiseen lineaariseen malliin. *Random forest* saavutti terveydenhuollon mallissa noin 0,17 selitysasteen validointiotoksessa, kun se lineaarisella regressiolla oli noin 0,136.¹⁶ Koneoppimismalleja on muissa maissa tarkasteltu muutamissa riskivakiointitutkimuksissa, ja niissä on vastaavalla tavoin huomattu koneoppimismallien parempi ennustekyky (Rose, 2016; Shrestha, 2018).

Koneoppimismallien osalta yksittäisten ennustemuuttujien yhteyttä kustannuksiin ei voida kuvata lineaarisen regression tapaan yksittäisellä parametriarvolla eikä hyvinvointialueiden rahoituslaskelmia voida siksi hajottaa eri tarvetekijöistä johtuviin osiinsa.¹⁷ Toisaalta myös lineaarisesta regressiosta voidaan tehdä hyvin monimutkainen sisällyttämällä valtava määrä muuttujia ja niiden välisiä interaktioita. Jos lineaarisessa mallissa on tuhansia ennustemuuttujia, sitä on vaikea pitää olennaisesti selkeämpänä verrattuna epälineaariin koneoppimismalleihin. Mallin ennustekyvyn ja läpinäkyvyyden välillä on siis vaihtosuhde: jos rahoituksen halutaan perustuvan yksinkertaisiin malleihin, joudutaan myös hyväksymään suuremmat virheet rahoituksen kohdentumisessa.

Koneoppimismallit tarjoavat mielenkiintoisen jatkotutkimuksen aiheen. Kansainvälisiä käytäntöjä seuraten tässä raportissa keskitytään suhteellisen yksinkertaisiin lineaarisiin malleihin.

3.6 Muuttujat

Terveydenhuollon, vanhustenhuollon ja sosiaalihuollon mallit sisältävät hieman eri muuttujia (taulukot 2–4). Mallien muuttujat ovat määräytyneet aiemman kirjallisuuden, sekä sisällöllisten ja tilastollisten näkökulmien perusteella. Suurin osa muuttujista on indikaattorimuuttujia (tai ns. dummy-muuttujia), jotka saavat arvon nolla tai yksi, kuvaten esimerkiksi onko henkilö opiskelija vai ei. Muuttujat, niiden valinnan perustelut sekä niiden osalta tehtävät herkkyystarkastelut kuvataan tässä luvussa muuttujatyypeittäin.

3.6.1 Kustannukset ja niiden tyypistys

Mallien selitettävä muuttuja on sosiaali- ja/tai terveydenhuollon palvelujen käytön keskimääräiset laskennalliset kustannukset per henkilö (ks. 3.1.1). Kustannusten jakauma on yksilötasolla hyvin vinoutunut: suurimmalla osalla väestöstä on matalahkot kustannukset, mutta pienellä prosentilla vuosittaiset kustannukset nousevat satoihin tuhansiin tai miljooniin euroihin. Äärihavainnot ovat potentiaalisesti ongelma lineaariselle regressiolle (Kuhn & Johnson, 2013). Äärihavaintojen vaikutusta voidaan vähentää asettamalla yksilön kustannuksille maksimiarvo, johon sitä korkeammat arvot asetetaan. Esimerkiksi Englannin G&A-mallissa yksilön kustannukset on rajattu (mielivaltaisesti asetettuun) 100 000 puntaan (Chaplin ym. 2016). Myös Häkkinen ym. (2020) työssä yksilön terveydenhuollon kustannukset rajattiin (200 000 €).

Herkkyystarkasteluissa vertaillaan malleja ilman rajausta malleihin, joissa rajausta on asetettu a) 200 000 euroon, b) 100 000 euroon.

3.6.2 Sairausluokitus

Sairastavuutta koskeva raakadata vaatii huomattavaa esikäsittelyä, jolla se saatetaan yksinkertaisempaan muotoon tilastollisia analyysejä varten. ICD-10-luokittelu sisältää yli 14 000 koodia, näiden välillä on yhteisvaikutuksia ja lisäksi terveydentilan kehityksen ajallinen elementti voi olla tärkeä (eli milloin diagnosoitu mitä, sairauksien luonnollinen kulku, hoito- ja seurantakäytännöt). Tilastollinen analyysi olisi laskennallisesti hyvin vaativaa ja läpinäkymätöntä, jos esikäsittelyä ei tehtäisi. Maailmalla käytetyissä riskivakiointimalleissa ongelma on ratkaistu luokittelemalla diagnooseja lääketieteelliseen asiantuntijatietoon ja tilastolliseen analyysiin perustuen suuremmiksi sairausluokiksi (Juhnke ym. 2016; Chaplin ym. 2016). Eri maissa kehitysprosessin lopputuloksena on saatu vaihtelevan kokoisia sairausluokituksia, joiden sisällöissä ja tarkkuuksissa on suuria eroja (Juhnke 2016). Lähtökohtana luokittelussa on jakaa diagnooseja luokkiin, joiden sisällä odotettava tuleva palvelukäyttö olisi mahdollisimman homogeenista (Ellis ym. 2017; Pope ym. 2004).

¹⁶ Vertailutuloksia ei ole vielä tuotettu päivitettyjä sairaus- ja palveluluokituksia käyttäen.

¹⁷ Osiossa 2.2 kuvatuilla tavoin.

Häkkinen ym. (2020) työssä sote-rahoituksen tarvevakiointia varten alettiin THL:ssa vuoden 2018 lopulla kehittämään Suomeen sopivaa sairausluokitusta. Luokituksen kehittäminen aloitettiin mallista, joka on ollut käytössä sosiaali- ja terveydenhuollon valtionosuuskriteereissä. Kyseisessä luokituksessa sairastavuus oli määritelty hyödyntäen Kela-korvattujen lääkkeiden erityiskorvausoikeustietoja suppeassa joukossa sairauksia (12 erityiskorvausoikeutta). Muodostuneeseen 52 sairausluokan luokitukseen tuli sekä tarkemmalla että karkeammalla tasolla olevia sairausluokkia. Häkkinen ym. tutkimuksessa (2020) sairastavuus selitti huomattavan osuuden yksilötason kustannusvaihtelusta, ja alueiden välillä oli merkittäviä eroja sairastavuudessa, mikä heijastui tarvekertoimiin ja rahoitukseen. Sairausluokituksen kehitystyötä jatkettiin sote-uudistuksen rahoitusmallista saadun palautteenkin myötä. Tammikuussa 2021 pyydettiin sairaanhoitopiirejä ja suurimpia kuntia ja kuntayhtymiä kommentoimaan luokitusta, sekä osallistumaan helmikuussa 2021 THL:sta pidettyihin Teams -etätyöpajoihin, joihin osallistui yhteensä 63 eri erikoisalujen asiantuntijaa. Kommenttien pohjalta joidenkin luokkien määrittelyjä tarkennettiin, joitakin laajempia luokkia jaettiin pienemmiksi luokiksi ja tehtiin täysin uusia luokkia. Luokituksessa olevien luokkien määrä kasvoi 123 luokaksi (erillinen Liitetaulukko 4).

Sairausluokituksen määrittelyssä on pyritty huomioimaan kattavasti sairastavuustietoa sisältäviä rekisterilähteitä (taulukko 1). Tällä pyritään varmistamaan, että tilastollisiin malleihin sisällytetty sairastavuustieto on mahdollisimman kattavasti kerätty. Lisäksi näin pyritään välttämään mahdollisimman pitkälti vaikuttamien hoidon kannusteisiin. Jos esimerkiksi sairastavuustieto poimittaisiin vain vuodeosastohoitojaksojen diagnooseista, kannustettaisiin rahoituksella vuodeosastohoitoon avohoidon sijasta. Samaan aikaan joudutaan kuitenkin tasapainoilemaan rekisteritietojen kattavuuden ja laadun kanssa (Sund 2012; Laatikainen ym. 2020).

Asiantuntijasuositusta seuraten sairastavuustietoa on perusterveydenhuollon osalta poimittu vain lääkkeiden vastaanottokäyntien tiedoista. Perusteluna on, että lääkärin vastaanottokäynnillä kirjaamaa diagnoositietoa pidetään luotettavana ja valtakunnallisesti kattavana verrattuna muihin yhteystapoihin ja ammattiryhmiin liittyvään diagnoositietoon. Asiantuntijanäkemykseen perustuen on joidenkin sairauksien osalta tehty poikkeavia ratkaisuja myös siitä mistä tietolähteistä ja miltä ajanjaksolta kyseistä sairastavuustietoa huomioidaan. Esimerkiksi syöpäsairauksissa rajoitetaan palvelujen käyttötietojen osalta erikoissairaanhoidon diagnooseihin. Häkkinen ym. (2020) raportissa muutamia poikkeuksia lukuun ottamatta (esim. syöpä, dementia) sairaudet poimittiin kahdelta tarkasteltavaa palvelujen käytön vuotta edeltävältä vuodelta.

Häkkinen ym. (2020) työstä hiukan poikkeavasti päädyttiin huomioimaan vuodeosastojaksoihin, perusterveydenhuollon lääkärin vastaanottokäynteihin ja erikoissairaanhoidon avohoitokäynteihin liittyvät päädiagnoosit, sivudiagnoosit sekä mahdolliset pitkäaikaisdiagnoosit. Englannissa ennustekyvyn havaittiin taantuvan kuuden diagnoosin poiminnan jälkeen, joskaan se ei huomattavasti laajemmilla poiminnoilla heikennytkään (Chaplin ym. 2016). Epävarmat diagnoosit lisäävät mallinnukseen kohinaa ja voivat heikentää ennustekykyä. Toisaalta myös vakavammassa sairauksissa voi nykyään olla soveltuvalta osin enenevässä määrin etävastaanottoihin ja muiden ammattiryhmien käynteihin perustuvia hoitoja, eikä rahoitusjärjestelmä saisi rangaista mahdollisesti tehokkaampien toimintatapojen käyttöönotosta. Tästä syystä herkkyytarkasteluna sisällytetään mallinnukseen myös hoitajien vastaanottokäyntien diagnoosit.

Raportissa Häkkinen ym. (2020) sairausmuuttujiin poimittiin diagnoositieto kahdelta vuodelta. Vastavasti on toimittu myös Englannin NHS:n G&A-mallissa (Chaplin ym. 2016), mutta ilman mitään tilastollisia testejä sen perusteeksi. Poiminta-aikaa pidentämällä voidaan periaatteessa saada sairastavuustietoa kerättyä kattavammin. Toisaalta kauempaa menneisyydestä poimittu sairastavuustieto on vielä vähemmän ajankohdasta kuin edeltävän vuoden tieto (siihen oletettavasti liittyisi muutoin tuoreempaa palvelukäyttöä) tai suorastaan vanhentunutta, jos sairaus on parantunut. Näin ollen ennustekyky kasvaa johonkin pisteeseen asti poiminta-ajan pidetessä ja laskee sen jälkeen. Tässä tutkimuksessa sairauksia poimitaan 1) yhden edeltävän vuoden ajalta, 2) kahden edeltävän vuoden ajalta ja 3) kolmen edeltävän vuoden ajalta.¹⁸

Oletettavasti sairauden diagnosoinnin ajankohdalla on myös merkitystä: sairaus voi aiheuttaa palvelukäyttöä vielä vuosienkin päästä ensidiagnoosista, mutta mahdollisesti vähemmän kuin edeltävänä vuonna tehty diagnoosi. Tästä syystä malleissa kokeillaan herkkyytarkasteluna sisällyttää edeltävän vuoden diagnooseihin perustuvien sairausmuuttujat ohella myös 1) edeltävää edeltävän vuoden diagnooseihin perustuvat

¹⁸ Tässä tehdään yksinkertaistus: periaatteessa voisi olla sairauskohtaisia eroja poiminta-ajassa.

sairausmuuttujat, jolloin sairausmuuttujien määrä kaksinkertaistuu, 2) edeltävää edeltävän vuoden ja sitä edeltäneen vuoden diagnooseihin perustuvat sairausmuuttujat, jolloin sairausmuuttujien määrä kolminkertaistuu. Näissä tarkasteluissa sairaudet poimittiin aina vain yhden, kyseisen vuoden diagnooseihin perustuen.

Herkkyystarkasteluna kokeillaan myös Englannin NHS:n G&A-mallissa käytettyä sairausluokitusta (Chaplin ym. 2016), jossa ICD-10-tautiluokitus on jaettu 152 pääluokkaansa.¹⁹ Malleihin sisällytettiin näitä luokkia vastaavat indikaattorimuuttujat. Lähtökohtaisesti Englannin sairausluokittelun vahvuutena voisi ajatella olevan se, että se käyttää lähes kaiken diagnoositiedon. Toisaalta koska luokitteluun ei sisälly mitään asiantuntijaperusteista etukäteisarviointia siitä, mitkä diagnoosit muodostaisivat odotettavalta tulevalta palvelukäytöltään homogenisia ryhmiä, on luokittelu oletettavasti vähemmän kuin optimaalinen.

Tässä raportissa otettiin käyttöön myös WHO:n näkövammaluokittelu. Lisäksi käytetään indikaattorimuuttujia henkilöille, joille löytyi sairausluokituksessa listattu näkövammadiagnoosi mutta ei WHO:n näkövammaluokittelun tietoa.

Ensivaiheessa sairausluokittelu on sisällytetty malleihin kokonaisuudessaan. Muiden muuttujien osalta lopulliseksi todetusta mallista on sitten karsittu pois sairausmuuttujat, joiden kertoimet ovat negatiivisia tai joiden ei oletettaisi aiheuttavan palvelukäyttöä kyseisessä palveluluokassa. Karsinta toteutettiin tarpeen vaatiessa vaiheittain, jos ensimmäisten poistamisen johdosta jonkin toisen sairausmuuttujan kerroin sai negatiivisen etumerkin. Herkkyystarkasteluissa on kuitenkin raportoitu myös tulokset täysmalleista, joista sairausmuuttujia ei ole karsittu.

3.6.3 Synnytykset ja keskoset

Terveydenhuollon mallissa huomioitiin synnytykset ja keskoset vastaavilla tavoin kuin raportissa Häkkinen ym. (2020). Palvelunkäyttövuoden aikana (2019) synnyttämiseksi on sisällytetty indikaattori ja keskoset on huomioitu sisällyttämällä heidät mallinnuksen perusjoukkoon (ainoina v. 2019 uusina havaintoina) ja sisällyttämällä malliin synnytyksen enneaikaisuus päivissä toisen asteen jatkuvana muuttujana.

3.6.4 Ikä ja sukupuoli

Sosiaali- ja terveydenhuollon palvelujen tarve vaihtelee iän ja sukupuolen mukaan, eikä ikärakenne ole jakaumaltaan samanlainen joka puolella maata (Chaplin ym. 2016). Terveydenhuollon palvelujen tarve on suuri erityisesti pienillä lapsilla ja iäkkäillä, ja sosiaalihuollon palvelujen tarve korostuu lapsilla, nuorilla ja iäkkäillä. Lakien ja asetusten velvoittamia eri ikä- tai vuosiluokkaan tai muihin ryhmiin kuuluville järjestettäviä palveluja kohdistuu suuremmissa määrin naisille kuin miehille, esimerkiksi raskaana olevien seurantaan ja synnytykseen liittyvät palvelut, sekä valtakunnalliset seulontaohjelmat.

Terveyden- ja sosiaalihuollon malleissa eri ikävaiheita kuvaavia indikaattorimuuttujia ryhmiteltiin hiukan Häkkinen ym. raportista (2020) poikkeavasti lakien ja asetusten²⁰ velvoittamia eri ikä- ja vuosiluokkiin kohdennettuja terveystarkastuksia jäljitellen: alle 1-vuotiaat, 1–6-vuotiaat, 7–12-vuotiaat, 13–18-vuotiaat, 19–25-vuotiaat ja siitä eteenpäin viiden vuoden ikäryhmissä. Nuoruusvaihe (13–18-vuotiaat) erotettiin omaksi muuttujaksi myös siksi, että eri ikäryhmissä painottuvat hieman eri palvelut (Haula & Peltola 2020) ja tällä on merkitystä myös sosiaalihuollon mallissa. Vanhustenhuollon mallissa viisi ikävuotta sisältäviä muuttujia on kuusi 65-vuoden iästä lähtien ja lisäksi muuttuja yli 95-vuoden iälle.

Yleisesti ottaen iän kaltaisen jatkuvan muuttujan muuntaminen joukoksi indikaattoreita ei ole suositeltavaa (Kuhn & Johnson, 2020). Tässä tapauksessa kuitenkin tiedetään, että erityisesti perusterveydenhuollon palvelujen käyttöä kohdistuu eri ikäluokkiin muun muassa terveystarkastusten ja seulontojen mukaan. Ikä ja

¹⁹ Luokitus siis jaetaan yksinkertaisesti osiin ”ylhäältä alas”.

²⁰ Terveydenhuoltolaki 1326/2010, <https://www.finlex.fi/fi/laki/ajantasa/2010/20101326#L2P13>;

Laki sosiaali- ja terveydenhuollon järjestämisestä 612/2021, <https://www.finlex.fi/fi/laki/alkup/2021/20210612#Pidm45237816815632>;

Laki korkeakouluopiskelijoiden opiskeluterveydenhuollosta 695/2019, <https://www.finlex.fi/fi/laki/alkup/2019/20190695>;

Oppilas- ja opiskeluhuoltolaki 1287/2013, <https://www.finlex.fi/fi/laki/alkup/2013/20131287>;

Valtioneuvoston asetus neuvolatoiminnasta, koulu- ja opiskeluterveydenhuollosta sekä lasten ja nuorten ehkäisevästä suun terveydenhuollosta 338/2011, <https://www.finlex.fi/fi/laki/alkup/2011/20110338>;

Valtioneuvoston asetus seulonnoista 339/2011, <https://www.finlex.fi/fi/laki/alkup/2011/20110339>

sukupuoli huomioidaan vastaaventyypisillä jaotteluilla indikaattorimuuttujiin yleisesti myös muissa riskintausmalleissa (esim. Chaplin ym. 2016; van Kleef ym. 2017). Tässä tutkimuksessa kokeillaan 1) indikaattorimuuttujaa naissukupuolelle ja iälle indikaattorimuuttujia yllä olevan luokittelun mukaisesti, 2) sukupuolittaisia indikaattorimuuttujia iälle yllä olevan luokittelun mukaisesti, 3) indikaattoria naissukupuolelle ja ikää kolmannen asteen jatkuvana muuttujana.

3.6.5 Taustamaa

Tutkimusten perusteella Suomessa ulkomaalaistaustaisten henkilöiden sosiaali- ja terveydenhuollon palvelujen käytössä on eroja muuhun väestöön verrattuna (Kuusio ym. 2020). Häkkinen ym. (2020) työssä taustamaa huomioitiin terveydenhuollon mallissa viitenä indikaattorina (OECD-maat, Lähi-Itä, Itä-Eurooppa, Aasia ja muu maailma). Tässä raportissa kokeillaan vastaavan jaottelun vaihtoehtona sisällyttää yksi indikaattori kaikille, joiden taustamaa ei ole Suomi.

3.6.6 Sosioekonomiset muuttujat

Sosioekonomisen tilanteen osalta on käytetty hyvin vastaavia tietoja kuin raportissa Häkkinen ym. (2020). Henkilön tulotaso on mitattu asuntokunnan käyttötuloilla kulutusyksikköä kohden. Lisäksi kokeiltiin indikaattoria toimeentulotuen perusosan saajille. Koulutustaso huomioidaan kahdella indikaattorilla (toisen asteen koulutus, korkea-asteen koulutus). Pääasiallinen toiminta sisällytetään muuttujilla opiskelija/koululainen, varusmies/siviilipalvelusmies ja työllinen. Pääasiallisen toiminnan määrittelyssä käytetään vuoden viimeisen viikon työsuhteen mukaista tietoa (Työvoimakäsite, TVM).

Sosioekonomiset muuttujia tarvitaan kuvaamaan taipumusta ja mahdollisuutta käyttää yksityisen terveydenhuollon ja työterveyshuollon palveluita. Tämä on tärkeää erityisesti työterveyshuollon kannalta, minkä saatavuudesta ei ole ollut hyvää rekisteritietoa vielä tämän tutkimuksen aineiston ajankohtana. Holster ym. (2022) raportoivat tulotason olevan työtilanteen ohella tärkeä työterveyshuollon käytön determinantti.

3.6.7 Työkyvyttömyyseläkkeet

Terveyden- ja sosiaalihuollon malleissa käytettiin myös muuttujina oikeutta työkyvyttömyyseläkkeeseen. Työkyvyttömyyseläke voi indikoida korkeampaa palvelutarvetta, mitä sairausluokitus ei täysin tavoita.²¹ Tässä raportissa tieto työkyvyttömyyseläkkeistä on poimittu kaikista saatavilla olleista lähteistä koko käytettävissä olevilta ajanjaksoilta²², riippumatta siitä, onko henkilö siirtynyt sittemmin työkyvyttömyyseläkkeeltä vanhuuseläkkeelle.

Työkyvyttömyyseläketiedon voisi sisällyttää yksinkertaisesti indikaattorimuuttujana. Häkkinen ym. (2020) sisällyttävät terveydenhuollon malliin työkyvyttömyyseläkkeen kahdella indikaattorilla, omansa 54-vuotiaille ja sitä nuoremmille ja toisen 55-vuotiaille ja sitä vanhemmille. Erityisesti nuoremman ikäluokan saama regressiokerroin on huomattavan korkea.

Tässä tutkimuksessa tarkastellaan kahta vaihtoehtoista tapaa määrittellä työkyvyttömyyseläkkeet malliin: 1) työkyvyttömyyseläke huomioidaan yksinkertaisena indikaattorimuuttujana, 2) työkyvyttömyyseläke sisällytetään viitenä ikäryhmittäisenä muuttujana: työkyvyttömyyseläkeläiset jaetaan viiteen ikäkvintiiliin ja jokaiselle kvintiilille asetetaan oma indikaattorimuuttujansa.

3.6.8 Julkisten ja yksityisten palveluiden saatavuus

Häkkinen ym. (2020) kontrolloivat yksityisten palveluiden käyttöä Kelan (osittain) korvaamien yksityislääkärien tutkimuksen ja hoidon,²³ suun terveydenhuollon ja avohoidon reseptilääkkeiden kustannuksilla.

²¹ Tässä on huomattava, että työkyvyttömyyksiin liittyvää diagnoositietoa käytettiin lisäksi sairastavuustiedon poiminnassa.

²² Käytössä on Kelan ja ETK:n tiedot voimassa olleista työkyvyttömyyseläkkeistä vuodesta 2015 alkaen ja Tilastokeskuksen tiedot vuodesta 2011 alkaen. Koska aineistossa tieto henkilöiden työkyvyttömyyseläkkeestä katoaa henkilön siirtyessä vanhuuseläkkeelle, tietoa työkyvyttömyyseläketäustasta ei ole estimointiaineistossa (tiedot vuosilta 2011–2018) juurikaan yli 71-vuotiaille. Tämä ratkaisu eroaa raportista Häkkinen ym. (2020), jossa käytettiin vain kahden vuoden tietoja.

²³ Sisältäen lääkärin, erikoislääkärin ja muu yksityinen tutkimus ja hoito.

Tarvekertoimissa nämä muuttujat on neutralisoitu. Tämän voi ajatella kontrolloivan samaan aikaan sekä yksityisten palveluiden saatavuutta että taipumusta käyttää niitä (on huomattava, että sosioekonominen aseman vaikutus mahdollisuuksiin käyttää yksityisiä palveluita on erikseen vakioituna, kuten myös sairastavuus). Tulkintaa kuitenkin vaikeuttaa se, että korkeampi yksityisten palveluiden käyttö voi johtua myös heikomasta julkisten palveluiden saatavuudesta.

Oletetaan, että julkisten palveluiden kustannusten ja yksityisten palvelujen käytön välillä huomataan negatiivinen yhteys. Jos se johtuu siitä, että samalla julkisten palveluiden tasolla jotkut käyttävät enemmän yksityisiä palveluita niiden paremman alueellisen saatavuuden vuoksi, on luontevaa ajatella, että paremman yksityisten palveluiden saatavuuden alueen rahoitusta tässä suhteessa myös vähennetään. Muutoin alueelle jaetaan julkisten palveluiden järjestämiseen enemmän rahaa suhteessa käyttöön, kuin maassa keskimäärin.²⁴ Jos taas havaittu negatiivinen yhteys johtuu heikomasta julkisten palveluiden saatavuudesta, ei samalla julkisten palveluiden tasolla käytetä enempää yksityisiä palveluita, vaan vaikutusta ajaa joidenkin alueiden heikko julkisten palveluiden saatavuus. Tällöin olisi luontevaa ajatella, että yksityisten ja julkisten palveluiden käytön välinen yhteys neutralisoitaisiin, jotta heikon saatavuuden alueita ei entisestään rangaistaisi matalammalla rahoituksella.²⁵

Siksi samaan aikaan tulisi pystyä kontrolloimaan julkisten palveluiden saatavuus. Häkkinen ym. (2020) käyttävät matka-aikaa ympärivuorokautista erikoissairaanhoidon päivystystä tarjoavaan sairaalaan mittaamaan palveluiden saatavuutta. Chaplin ym. (2016) kontrolloivat NHS:n G&A-mallissa tarjontapuolen tekijöinä julkisten palveluiden saatavuutta etäisyyspainotetuilla hoitopaikoilla mitattuna. Lisäksi he sisällyttävät CCG-tason alueindikaattorit tarjontatekijöiksi. Tarve-ennusteita laskiessa Chaplin ym. neutralisoivat nämä tarjontamuuttujat tavallisen käytännön mukaisesti. Yksityisten terveystalouden käyttöä mittaamaan he sisällyttävät indikaattorimuuttujan sille, onko henkilö käyttänyt edellisen vuoden aikana yksityisiä terveystalouksia.

Koska monikanavarahoituksen purku poistanee lähivuosina Kelan korvaaman yksityisen palvelukäytön aineistolähteen, tiputettiin tästä päivitysversiona pois Kelan korvaamiin kustannuksiin liittyvät muuttujat. Yksityisten palveluiden käyttöä pyritään tavoittamaan yllä mainituilla sosioekonomisilla muuttujilla. Matka-aika erikoissairaanhoidon ympärivuorokautiseen päivystykseen sisällytettiin edelleen terveydenhuollon malliin.

Alueindikaattorien käytössä ongelmana on, että tällöin kontrolloitua tulee kaikki se aluetason vaihtelu palvelukäytössä, mikä ei tule muilla sisällytetyillä muuttujilla huomioitua. Osa tästä vaihtelusta johtuu luultavasti havaitsemattomasta tarpeesta, mitä ei haluttaisi vakioida pois. Tästä syystä alueindikaattoreita ei myöskään voi missään nimessä sisällyttää tarvetekijöiden joukkoon (mutta mahdollisesti neutralisoitavaksi tarjontakontrolliksi): tällöin vain jaettaisiin rahoitusta menneen palvelukäytön perusteella, käytännössä olettaen alueellisten palvelukäyttöerojen johtuneen kokonaisuudessaan tarve-eroista.

Lisäksi toteutettiin kuitenkin herkkyystarkastelu, missä malleihin sisällytettiin myös tarjontakontroleiksi hyvinvointialueille indikaattorimuuttujat (sosiaalihuollossa Eksotelle, Päijät-Hämeen hyvinvointikuntayhtymälle ja Oulun ja Helsingin kaupungeille). Hyvinvointialueiden käyttöä voi kritisoida siitä, etteivät ne ole aineistovuonna olleet palveluiden järjestämisvastuussa. Kuitenkin hyvinvointialueet seuraavat kohtuudella sairaanhoitopiirirajoja, ja niiden sisällä voisi ainakin erikoissairaanhoidon palvelutarjonnan olettaa olevan suhteellisen homogeenista.

3.6.9 Interaktiot

Regressiopuu-pohjaiset mallit, kuten *random forest* ja *gradient boosting machine* huomioivat luontaisesti muuttujien yhteisvaikutukset eli interaktiot (Kuhn & Johnson, 2020), mikä osaltaan selittää myös niiden hyvää ennustekykä. Interaktioiden huomioiminen lineaarisessa regressiossa vaatii niiden eksplisiittistä määrittelyä ja sisällyttämistä selittäviksi muuttujiksi. Tällöin tulisi olla jokin varsinaista mallinnusta edeltävä menettely ennustekyvälle merkittävien interaktioiden tunnistamiseen, esimerkiksi regressiopuiden käyttö

²⁴ Tässä kohdin erottelu henkilön julkisten palveluiden tarpeeseen ja terveystalouden kokonaistarpeeseen on kriittinen.

²⁵ Tietysti on mahdollista, että huono julkisten palveluiden saatavuus johtuu tehottomuudesta, mistä ei luultavasti haluttaisi kompensoida. Jos aikaisempi rahoitus pohjautuisi täysin (täydellisesti toteutettuun) olosuhdekorjattuun tarveperusteisuuteen, näin väistämättä olisi.

(Kuhn & Johnson, 2020). Kuitenkin Buchner ym. (2016) ja van Veen ym. (2018) havaitsivat, että regressiopoilla tunnistettujen interaktioiden vaikutus lineaarisen mallin ennustekykyyneen on varsin vaatimaton. Chaplin ym. (2016) sisällyttävät Englannin G&A-malliin ICD10:n 21 pääluokan välisiä parittaisia interaktioita.

Sairausryhmien välisten interaktioiden tarkastelu rajattiin tämän raportin ulkopuolelle aikarajoitteista joutuena. Yllä mainittuun kirjallisuuteen perustuen näyttää myös, että sairausluokituksen muu edelleenkehittäminen voi olla tehokkaampi tapa parantaa malleja,

3.6.10 Monisairastavuus

Englannin NHS:n rahoitusjärjestelmässä Chaplin ym. (2016) sisällyttävät malliin myös joukon indikaattoreita sairauksien määrälle, joista valitaan sisällytettävien indikaattorien määrä varsinaisten sairausluokamuuttujien tilastollisen merkitsevyyden ja etumerkin perusteella (halutaan saada mahdollisimman paljon merkitseviä ja etumerkiltään positiivisia muuttujia). Chaplin ym. raportoivat sairauksien määrän indikaattorien vaikutuksen myös sovitteen hyvyteen, mutta eivät kuitenkaan käytä sovitteen hyvyttä spesifikaatiossa. Nämä mallispesifikaatiot suoritetaan käyttäen siis vain estimointiotoksen tietoja.

Tässä tutkimuksessa kokeiltiin sisällyttää malleihin sairauksien määrä jatkuvana muuttujana. Muuttuja saa arvon nolla, kun henkilöllä on yksi tai nolla sairausluokitukseen kuuluvaa sairautta ja arvon 1, kun henkilöllä on kahdessa luokituksen luokassa sairaus, arvon 2, kun henkilöllä on kolmessa luokituksen luokassa sairaus ja niin edelleen.

3.6.11 Perhetilanne ja siviilisäätty

Raporttia Häkkinen ym. (2020) seuraten, yksin asuminen sisällytettiin terveydenhuollon malliin ikäluokittaisina indikaattoreina (alle 75-vuotias, 75–84-vuotias, 85–89-vuotias ja yli 90-vuotias). Kaikkiin malleihin sisällytettiin myös siviilisäätty jaoteltuna indikaattorimuuttujiksi naimattomille, eronneille ja leskille. Häkkinen ym. (2020) sisällyttivät lisäksi vanhustenhuollon malliin muuttujan yksin tai laitoksessa asumiselle.

Perhetilanne ja läheisiltä saatu apu on oletettavasti tärkeää erityisesti ikääntyneiden palvelujen tarpeen suhteen. Tämä on kuitenkin mielekkäämpää huomioida siviilisäädyn kaltaisten muuttujien kautta eikä asumistilannetta kuvaavilla muuttujilla, sillä asumistilanne riippuu suoraan alueen tekemistä hoitopäätöksistä. Terveydenhuoltoon ei sisälly varsinaisia asumispalveluita, joten terveydenhuollon mallissa asumistilanteen huomioiminen on ongelmattomampaa.

4 Tulokset

4.1 Ehdotetut mallit

Taulukoissa 2–4 esitetään hyvinvointialueiden järjestämän sote-palveluiden rahoitukseen ehdotetut terveyden-, vanhusten- ja sosiaalihuollon mallit (sektorit th, vh, sh) ja niiden estimointitulokset. Seuraavassa aluvuossa käsitellään tarkemmin näiden mallien suorituskykyä verrattuna vaihtoehtoisiin mallivalintoihin. Malleja tai niiden tuloksia ei voi verrata suoraan raportissa Häkkinen ym. (2020) esitettyihin malleihin sairausluokituksessa ja palvelukäyttöaineistossa tapahtuneista muutoksista johtuen. Erityisesti on huomioitava, että terveyden- ja vanhustenhoidon mallien huonompi selitysaste verrattuna Häkkisen ym. raportointiin johtuu tässä raportissa käytetystä, tuoremasta ja tarkemmasta palvelukäyttöaineistosta. Esimerkiksi terveydenhuollon palvelujen käytön laskennallisissa vuosikustannuksissa näkyy huomattavaa kasvua yksilötason äärihavainnoissa, kun verrataan vuotta 2019 vuoteen 2017: korkein havainto kasvaa muutamista sadoista tuhansista yli kahteen miljoonaan euroon. Näillä äärihavainnoilla on huomattava vaikutus selitysasteeseen.

Malleissa tärkeitä sosiodemografisia ja -ekonomisia palvelujen käytön kustannuksia lisääviä tekijöitä ovat työkyvyttömyys (th, sh), työttömyys (th, sh), matala koulutustaso (th, sh, vh), matalat asuntokunnan käytävissä olevat rahatulot (vh, sh, th) ja terveydenhuollon mallissa hyvin iäkkäiden yksinasuminen (taulukot 2–4). Ikäluokissa suurimpia regressiokertoimia on vanhustenhoidon mallissa yli 85-vuotiailla, terveydenhuollon mallissa elinkaaren ääripäässä (alle yksivuotiailla ja iäkkäillä) ja 13–18-vuotiaiden tyttöjen ikäryhmässä, ja sosiaalihuollon mallissa 13–18-vuotiaiden ikäluokassa. Sosioekonomisten muuttujien etumerkit vastaavat ennako-oletuksia: tulo- ja koulutustason laskiessa julkisia palveluita käytetään enemmän, sillä pääsy yksityisten palveluiden ja työterveyshuollon piiriin on heikompaa. Vastaavasti työssäkäyvät, varusmiehet ja opiskelijat käyttävät vähemmän erityisesti terveydenhuollon palveluita, mikä sopii siihen, että heillä on pääsy vaihtoehtoisille palvelutarjoajille.²⁶

Sairastavuusmuuttujien osalta suurimpia eroja Häkkinen ym. (2020) työssä raportoituun ovat terveydenhuollon mallissa mm. syöpäluokan jakautuminen yhdestä luokasta 20:een eri syöpiä sisältävään luokkaan ja Psykoosisairaudet ja kaksisuuntainen mielialahäiriö -luokan lisäkustannusten lähes kolminkertaistuminen (luokan määrittelyä tarkennettiin). Uusissa sairausluokissa suurimpia lisäkustannuksia on neurologisten sairauksien ja suun sairauksien luokissa. Vanhojen laajojen luokkien jakamisen myötä nähdään eroja luokkiin sisältyneiden sairauksien lisäkustannuksissa, esimerkiksi em. syöpien lisäksi ruuansulatuselinten sairauksien ja naistentautien osalta. Terveydenhuollon mallissa suurimpia regressiokertoimia esiintyy luokissa Munuaisten vajaatoiminta, Laihuushäiriö, Opidoidiriippuvuus, Psykoosisairaudet ja kaksisuuntainen mielialahäiriö, Krooniset haavat ja Imukudoksen, verta muodostavien kudosten ja lähisukuisten kudosten pahanlaatuiset kasvaimet. Terveydenhuollon mallista jätettiin pois luokat Muistisairaudet ja Alzheimerin tauti, Älyllinen kehitysvammaisuus ja CP-oireyhtymä. Näistä älyllinen kehitysvammaisuus sai huomattavan negatiivisen kertoimen (noin 2000, tulosta ei esitetty).

Muistisairaudet ja Alzheimerin tauti -luokan lisäkustannukset olivat suurimmat ja kasvoivat vanhustenhoidon mallissa (luokan määrittelyä tarkennettiin). Hengityshalvaus-luokan lisäkustannukset kaksinkertaistuivat vanhustenhoidon mallissa; myös Psykoosisairaudet ja kaksisuuntainen mielialahäiriö -luokan lisäkustannukset kasvoivat. Muita vanhustenhoidon mallissa suurimpien lisäkustannusten luokkia olivat Neuroimmunologiset sairaudet, Hydrokefalus, Aivoverenkiertohäiriöt ja WHO:n vakavimmat näkövammaluokat. Sosiaalihuollon mallissa suurimmat lisäkustannukset olivat edelleen luokissa Älyllinen kehitysvammaisuus, Hengityshalvaus, CP-oireyhtymä ja Tarkkaavaisuus- ja käyttöhäiriöt. Uusissa luokissa lisäkustannukset olivat suurimmat WHO:n näkövammaisuusluokittelun, Opidoidiriippuvuuden ja neurologisten sairauksien luokissa.

²⁶ Vanhustenhoidon mallissa työssäkäynnin kerroin oli positiivinen siinä vaihtoehtoisessa mallimäärittelyssä, jossa sen sisällyttämistä kokeiltiin. Tulosta ei ole helppo tulkita eikä toisaalta vanhustenhoidossa työllisyystilanteen sisällyttäminen vaikuta teoreettisesti olennaiselta. Koska sen vaikutus mallien ennustekykyyneen oli hyvin vähäinen (työllisiä on vähän yli 64-vuotiaissa), ei sitä sisällytetty lopulliseen malliin.

Taulukko 2. Terveysthuollon mallin estimointitulokset: regressiokertoimet ja keskivirheet.

	Kerroin	Keskivirhe
Sukupuoli, ikäryhmä		
Nainen, ikä 0	1016,8***	77,80
Nainen, ikä 1–6	34,12	37,01
Nainen, ikä 7–12	226,0***	37,12
Nainen, ikä 13–18	641,7***	38,44
Nainen, ikä 19–25	259,5***	35,92
Nainen, ikä 26–30	191,7***	37,73
Nainen, ikä 31–35	144,3***	38,24
Nainen, ikä 36–40	98,27*	39,56
Nainen, ikä 41–45	28,10	37,44
Nainen, ikä 46–50	118,0*	46,25
Nainen, ikä 51–55	220,2***	39,63
Nainen, ikä 55–60	153,4***	35,24
Nainen, ikä 61–65	180,3***	40,83
Nainen, ikä 66–70	354,2***	44,79
Nainen, ikä 71–75	663,0***	47,43
Nainen, ikä 76–80	701,0***	46,24
Nainen, ikä 81–85	939,0***	54,25
Nainen, ikä 86–90	446,2***	73,53
Nainen, ikä 91–95	-407,9***	100,5
Nainen, yli 95 vuotta	-888,7***	152,9
Mies, ikä 0	1315,4***	134,7
Mies, ikä 1–6	211,2***	37,47
Mies, ikä 7–12	251,4***	37,50
Mies, ikä 13–18	92,71*	37,46
Mies, ikä 19–25	80,12	46,00
Mies, ikä 26–30	97,03*	46,84
Mies, ikä 31–35	-5,415	39,01
Mies, ikä 36–40	34,61	46,67
Mies, ikä 46–50	25,24	42,03
Mies, ikä 51–55	137,2***	37,24
Mies, ikä 55–60	172,3***	35,90
Mies, ikä 61–65	243,3***	38,21
Mies, ikä 66–70	473,6***	40,33
Mies, ikä 71–75	834,9***	45,68
Mies, ikä 76–80	1087,4***	57,96
Mies, ikä 81–85	1440,9***	62,85
Mies, ikä 86–90	1493,5***	85,56
Mies, ikä 91–95	1092,6***	147,2
Mies, yli 95 vuotta	1078,7**	393,6
Sairastavuus		
Tuberkuloosi	1460,7**	538,2
HIV, C-hepatiitti	3246,3***	433,4
B-hepatiitti	942,5*	383,3

4 Tulokset

	Kerroin	Keskivirhe
Ei-tuberkuloottiset mykobakteerit	2634,2**	814,4
Ruusut	1839,9***	118,2
Veneeriset syylät	381,5*	153,4
Huulen, suun ja nielun pahanlaatuiset kasvaimet	2772,9***	242,9
Ruuansulatuselinten pahanlaatuiset kasvaimet	4571,2***	127,3
Hengityselinten ja rintaontelon elinten pahanlaatuiset kasvaimet	4136,4***	205,0
Luun ja nivelruston pahanlaatuiset kasvaimet	5576,0***	1048,3
Ihon melanooma ja muut pahanlaatuiset ihokasvaimet	680,6***	69,08
Muut pehmytkudoksen pahanlaatuiset kasvaimet	4419,2***	382,5
Rintasyöpä	1817,9***	58,43
Naisen sukupuolielinten pahanlaatuiset kasvaimet	3211,0***	155,5
Miehen sukupuolielinten pahanlaatuiset kasvaimet	1464,4***	89,73
Virtsaelinten pahanlaatuiset kasvaimet	2097,3***	145,7
Silmän, keskushermoston ja aivohermojen pahanlaatuiset kasvaimet	3806,8***	281,3
Kilpirauhasen ja muiden umpirauhasten pahanlaatuiset kasvaimet	1680,1***	468,4
Pahanlaatuiset kasvaimet, joiden sijaintipaikka on epäselvä, sekundaarinen tai määrittämätön	5918,3***	153,9
Imukudoksen, verta muodostavien kudosten ja lähisukuisten kudosten pahanlaatuiset kasvaimet	6234,0***	164,0
Pintasyövät	311,9	203,4
Keskushermostokalvojen ja aivohermojen hyvänlaatuiset kasvaimet	1280,2***	146,0
Kasvu- ja leviämistäipumukseltaan epäselvät tai tuntemattomat kasvaimet	3096,8***	147,6
Krooniset hankinnaiset ja perinnölliset anemiat, hyytymyshäiriöt, neutropenia	2858,3***	132,5
Immuunipuutokset / immunol. Häiriöt	3462,6***	424,8
Amyloidoosit	4606,8***	802,1
Kilpirauhasen vajaatoiminta	87,71***	23,18
Diabetes	706,7***	29,98
Hypertyreoosi	1068,0***	90,62
Struuma	766,6*	320,8
Hyperparatyreoosi	248,8	204,3
Lihavuus	842,8***	90,77
Päihde- ja riippuvuushäiriöt (pl. Opioidiriippuvuus)	2744,4***	120,7
Opioidiriippuvuus	8004,6***	606,8
Tupakoinnin aiheuttamat haitat	135,8*	60,01
Psykoosisairaudet ja kaksisuuntainen mielialahäiriö	6085,2***	186,1
Masennus- ja ahdistuneisuushäiriöt	911,3***	40,43
Dissosiaatio- ja somatisaatiohäiriöt	1980,5***	231,0
Laihuushäiriö	7779,7***	555,5
Syömishäiriöt (pl. Laihuushäiriö)	3055,9***	498,9
Unihäiriöt	328,6***	52,78
Persoonallisuushäiriöt	938,3***	213,9
Oppimiskyvyn vaikeudet yhdistettynä muihin kuin laaja-alaisiin kehityshäiriöihin	948,0***	72,51
Laaja-alaiset kehityshäiriöt ("autismispektri")	498,3*	252,6
Tarkkaavaisuus- ja käytöshäiriöt	1771,5***	85,09
Sekavuustilat ja elimelliset aivo-oireyhtymät	910,6***	151,2
Transsukupuolisuus tai muu syntymässä määriteltyä (juridista) sukupuolta vastaamaton	2552,4**	807,3

	Kerroin	Keskivirhe
sukupuoli-identiteetti		
Määrittämätön mielenterveyden häiriö	2889,8***	310,6
Hengityshalvaus	4829,3***	532,6
Parkinson ja muut rappeuttavat liikehäiriösairaudet	2187,2***	155,4
Epilepsia	639,4***	65,90
Uniapnea	354,8***	38,92
Neuroimmunologiset sairaudet	1255,9***	81,81
Keskushermoston ja ääreishermoston tulehdus/tulehdukselliset sairaudet	2312,9***	162,6
Migreeni ja muut päänsärkysairaudet, muu krooninen kipu	576,3***	50,98
Hydrokefalus	3013,0***	327,4
Neuromuskulaarisairaudet	1944,9***	215,7
Allerginen silmätulehdus + allerginen nuha	61,04***	10,10
Silmien rappeumataudit	1214,9***	63,81
Glaukooma	315,7***	36,59
Näkövammaisuus (ne, joilla ei tietoa WHO-luokituksesta)	283,5***	63,73
Silmien sarveiskalvosairaudet	408,2***	49,36
Silmien verkkokalvoirtaumat ja verisuonitukokset	745,4***	90,01
Silmien taittovirheet	568,2***	47,13
Silmien lasiaissairaudet	505,5***	95,77
Korvakirurgia	742,7***	102,6
Huimaus/Korvan tasapainoelimen häiriöt	318,6***	61,15
Johtumistyyppinen ja sensorineuraalinen kuulonalenema	276,6***	37,51
Verenpainetauti	350,2***	21,45
Sepelvaltimotauti	668,2***	28,62
Eteisvärinä	1068,1***	32,99
Sydämen vajaatoiminta	2228,9***	72,06
Aivoverenkiertohäiriöt	1270,3***	56,65
Ateroskleroosi	4417,9***	138,8
Sydämen läppäsairaudet	1694,7***	74,13
Kardiomyopatiat	967,1***	165,0
Sydämen johtumishäiriöt	790,7***	135,2
Keuhkoveritulppa	1574,3***	174,6
Rinta-aortan aneurysmat	1460,3***	162,8
Keuhkokuume	2079,0***	101,0
COPD	1541,4***	108,7
Astma	427,0***	23,98
Keuhkokudoksen sairaudet	2306,3***	287,1
Hengityselinten krooninen toimintavajaus	4602,3***	517,5
Hammaskaries ja hammasytimen ja hampaanjuuren kärkeä ympäröivien kudosten sairaudet	2252,8***	148,0
Purentaelimen poikkeavuudet	1542,1***	80,38
Hampaan kiinnityskudosten sairaudet	2306,6***	214,2
Suun protetiikka	823,4***	173,2
Refluksi ja ulcus	677,2***	17,87
Divertikkelit ja ärtyvä suoli	678,4***	68,47
Sappirakon ja sappiteiden sairaudet	1210,8***	123,5

	Kerroin	Keskivirhe
Maksan tulehdussairaudet ja vajaatoiminta	2270,8***	188,3
Crohnin tauti ja haavainen koliitti	1212,5***	46,11
Keliakia	525,5***	103,9
Atooppinen ekseema	354,6***	34,01
Psoriaasi	595,6***	62,09
Krooniset haavat	5936,4***	240,4
Allerginen kosketusihottuma	522,5***	156,4
Aktiininen keratoosi	193,8*	82,14
Nivelreuma	1140,2***	39,11
Artroosisairaudet	920,9***	26,83
Olkapään vaivat	191,5***	31,25
Muualla luokittelemattomat muut nivelsairaudet	478,2***	36,77
Selkärangan sairaudet	617,9***	21,63
Luukato	1186,0***	109,7
Polven sisäiset viat	207,2***	49,86
Varpaiden hankinnaiset epämuotoisuudet	906,7***	152,1
Raajakipu	672,0***	40,34
Munuaisten vajaatoiminta	8340,3***	188,9
Munuaissairaudet	1899,4***	103,2
Virtsakivet ja virtsaushäiriöt	845,8***	58,27
Endometrioosi	1296,7***	91,61
Kohdunkaulan dysplasia	267,4***	55,99
Kuukautisvuotohäiriöt	726,9***	115,6
Hedelmättömyys	1566,8***	56,79
Miehen sukupuolielinten sairaudet	552,6***	98,04
Vammat ja myrkytykset	571,5***	18,24
Lonkkamurtuma	1498,1***	164,5
WHO:n näkövammaluokitus		
1. aste	483,7***	130,6
2. aste	524,3	305,9
3. aste	91,20	250,0
4. aste	453,0	324,5
5. aste	1845,4*	823,9
9. aste (määrittelemätön)	1384,2***	414,9
Raskauden ennenaikaisuus päivinä	-327,7***	68,48
Raskauden ennenaikaisuus päivinä²	19,88***	1,606
Synnytys	5462,7***	23,82
Työkyvytön, ikäkvantiili		
1. ikäkvantiili	3099,9***	187,1
2. ikäkvantiili	530,0***	121,0
3. ikäkvantiili	-368,9***	64,55
4. ikäkvantiili	280,4***	78,93
5. ikäkvantiili	595,2***	90,35
Pääasiallinen toiminta		
Työllinen	-240,7***	16,41

	Kerroin	Keskivirhe
Opiskelija	-112,5***	28,96
Varusmies	-414,1***	30,23
Koulutus		
Toinen aste	-68,09***	19,33
Korkeakoulu	-100,1***	18,82
Asuntokunnan käyttötulo per kulutusyksiköt, luonnollinen logaritmi	-111,6***	13,00
Siviilisääty		
Naimaton	41,48**	14,63
Eronnut	89,03***	17,79
Leski	-108,2***	25,50
Perhemuoto		
Yhden aikuisen perhe	92,21***	11,12
Yksinasuja, alle 75-vuotias	-123,9***	18,52
Yksinasuja, 75–84-vuotias	747,4***	38,87
Yksinasuja, 85–89-vuotias	1850,2***	73,54
Yksinasuja, yli 89-vuotias	3456,0***	106,7
Taustamaa muu kuin Suomi	-205,3***	18,85
Matka-aika minuutteina	-1,640***	0,288
Matka-aika minuutteina²	0,00540***	0,00156
Vakiotermi (asukaskohtainen erä)	1703,9***	130,9
Havaintoja	5 490 245	
R²	0,064	

Taulukko 3. Sosiaalihuollon mallin estimointitulokset: regressiokertoimet ja keskivirheet.

	Kerroin	Keskivirhe
Nainen	4,975	9,467
Ikäryhmä		
Ikä 0	21,43	47,51
Ikä 1–6	36,78	31,82
Ikä 7–12	59,66	38,45
Ikä 13–18	1187,1***	57,69
Ikä 19–25	-135,6***	23,52
Ikä 26–30	-130,1***	19,58
Ikä 31–35	-31,71	19,16
Ikä 36–40	16,24	19,28
Ikä 46–50	45,28*	20,70
Ikä 51–55	87,62***	18,51
Ikä 56–60	125,2***	20,58
Ikä 61–65	102,5***	19,94
Ikä 66–70	32,49	22,19
Ikä 71–75	45,83	25,25
Ikä 76–80	-76,45**	29,17
Ikä 81–85	-187,5***	32,91
Ikä 86–90	-427,1***	43,40
Ikä 91–95	-756,0***	61,79
Yli 95 vuotta	-1063,7***	131,0
Sairastavuus		
HIV, C-hepatiitti	1096,0***	191,6
Ruusut	651,8***	117,2
Päihde- ja riippuvuushäiriöt (pl. Opioidiriippuvuus)	2415,5***	112,8
Opioidiriippuvuus	3879,0***	465,2
Psykoosisairaudet ja kaksisuuntainen mielialahäiriö	1678,1***	97,14
Masennus- ja ahdistuneisuushäiriöt	53,67	50,36
Älyllinen kehitysvammaisuus	27300,7***	479,2
Laaja-alaiset kehityshäiriöt ("autismispektri")	4941,4***	543,3
Tarkkaavaisuus- ja käytöshäiriöt	7196,6***	399,1
Muistisairaudet ja Alzheimerin tauti	236,8***	66,03
Hengityshalvaus	12986,2***	1164,1
Parkinson ja muut rappeuttavat liikehäiriösairaudet	942,4***	113,3
Epilepsia	1981,6***	125,2
CP-oireyhtymä	11700,7***	902,1
Neuroimmunologiset sairaudet	3458,9***	285,5
Keskushermoston ja ääreishermoston tulehdus/tulehdukselliset sairaudet	706,6***	210,4
Hydrokefalus	1928,3**	733,9
Neuromuskulaarisairaudet	2502,8***	484,0
Aivoverenkiertohäiriöt	687,1***	63,05
Keuhkokuume	619,2***	83,66
COPD	180,8*	78,90
Hengityselinten krooninen toimintavajaus	3983,3***	764,4

	Kerroin	Keskivirhe
Hammaskaries ja hammasytimen ja hampaanjuuren kärkeä ympäröivien kudosten sairaudet	652,0*	263,1
Hampaan kiinnityskudosten sairaudet	613,8	341,2
Krooniset haavat	1129,1***	195,3
Nivelreuma	128,2**	47,62
Raajakipu	156,8**	53,98
Munuaissairaudet	654,8***	126,0
Hedelmättömyys	186,1***	27,55
Vammat ja myrkytykset	413,8***	26,09
WHO:n näkövammaluokitus		
1. aste	1799,6***	179,6
2. aste	3580,7***	589,3
3. aste	4442,7***	608,0
4. aste	5890,2***	703,9
5. aste	14458,3***	2533,4
9. aste (määrittelemätön)	1520,6*	698,4
Pääasiallinen toiminta		
Työllinen	-11,52	14,71
Opiskelija	-324,0***	33,68
Varusmies	28,27	39,62
Koulutus		
Toinen aste	-291,6***	15,37
Korkeakoulu	-132,2***	13,31
Työkyvytön, ikäkvantiili		
1. ikäkvantiili	3013,0***	161,1
2. ikäkvantiili	2069,2***	118,4
3. ikäkvantiili	946,6***	98,75
4. ikäkvantiili	626,3***	77,51
5. ikäkvantiili	732,6***	85,49
Asuntokunnan käyttötulo per kulutusyksiköt, luonnollinen logaritmi	-423,5***	8,366
Siviilisääty		
Naimaton	150,0***	10,29
Eronnut	-72,21***	11,79
Leski	-112,9***	19,29
Yhden aikuisen perhe	272,5***	28,96
Taustamaa ei Suomi	-479,1***	18,70
Vakiotermi (asukaskohtainen erä)	4399,0***	81,23
Havaintoja	1 039 910	
R²	0,258	

Taulukko 4. Vanhustenhuollon mallin estimointitulokset: regressiokertoimet ja keskivirheet.

	Kerroin	Keskivirhe
Sukupuoli, ikäryhmä		
Nainen, ikä 65–70	-80,01**	24,80
Nainen, ikä 71–75	151,5***	33,36
Nainen, ikä 76–80	967,3***	50,96
Nainen, ikä 81–85	3571,2***	80,26
Nainen, ikä 86–90	8317,5***	127,5
Nainen, ikä 91–95	13401,1***	230,6
Nainen, ikä yli 96 vuotta	16713,4***	505,0
Ikäryhmä, mies		
Mies, ikä 71–75	193,3***	34,30
Mies, ikä 76–80	388,4***	50,47
Mies, ikä 81–85	1687,9***	84,41
Mies, ikä 86–90	4896,3***	151,1
Mies, ikä 91–95	8934,5***	330,1
Mies, ikä yli 96 vuotta	10587,5***	908,1
Sairastavuus		
Tuberkuloosi	3325,0**	1175,9
Ruusut	1418,8***	200,0
Krooniset hankinnaiset ja perinnölliset anemiat, hyytymyshäiriöt, neutropenia	565,8*	236,3
Diabetes	982,9***	42,50
Lihavuus	2159,0***	216,5
Päihde- ja riippuvuushäiriöt (pl. Opioidiriippuvuus)	4147,2***	238,9
Psykoosisairaudet ja kaksisuuntainen mielialahäiriö	7453,0***	222,9
Masennus- ja ahdistuneisuushäiriöt	3553,0***	151,0
Dissosiaatio- ja somatisaatiohäiriöt	637,8	562,5
Sekavuustilat ja elimelliset aivo-oireyhtymät	3695,6***	350,5
Muistisairaudet ja Alzheimerin tauti	19316,9***	120,6
Hengityshalvaus	11592,3***	1171,2
Parkinson ja muut rappeuttavat liikehäiriösairaudet	5067,0***	181,8
Epilepsia	3116,2***	173,4
CP-oireyhtymä	4191,1*	1681,0
Neuroimmunologiset sairaudet	9259,2***	699,9
Keskushermoston ja ääreishermoston tulehdus/tulehdukselliset sairaudet	1050,1***	295,6
Hydrokefalus	7833,4***	777,0
Neuromuskulaarisairaudet	5469,3***	576,9
Eteisvärinä	160,5**	59,60
Sydämen vajaatoiminta	2655,1***	126,4
Aivoverenkiertohäiriöt	7026,5***	126,9
Ateroskleroosi	925,6***	175,2
Keuhkoveritulppa	1510,8***	333,9
Keuhkokuume	2198,4***	157,7
Hengityselinten krooninen toimintavajaus	1715,7**	532,0
Hammaskaries ja hammasytimen ja hampaanjuuren kärkeä ympäröivien kudosten sairaudet	582,7*	292,2
Refluksi ja ulcus	575,1***	37,29

	Kerroin	Keskivirhe
Krooniset haavat	4787,9***	366,4
Nivelreuma	665,0***	105,0
Artroosisairaudet	607,7***	62,70
Luukato	3598,9***	218,6
Munuaissairaudet	4954,6***	254,6
Virtsakivet ja virtsaushäiriöt	668,4***	92,65
Vammat ja myrkytykset	1750,4***	70,27
Lonkkamurtuma	5677,5***	367,9
WHO:n näkövammaluokitus		
1. aste	2698,1***	279,6
2. aste	5188,5***	750,9
3. aste	4918,2***	752,0
4. aste	8150,7***	1030,2
5. aste	11163,7**	3506,2
9. aste (määrittelemätön)	1024,6	861,2
Koulutus		
Toinen aste	-280,3***	33,19
Korkeakoulu	-147,8***	31,62
Asuntokunnan käyttötulo per kulutusyksiköt, luonnollinen logaritmi	-1257,7***	15,02
Siviilisäät		
Naimaton	2264,7***	56,98
Eronnut	1030,0***	35,84
Leski	1758,5***	47,46
Taustamaa muu kuin Suomi	-1306,0***	78,23
Vakiotermi (asukaskohtainen erä)	11410,4***	156,2
Havaintoja	1 198 119	
R²	0,299	

4.2 Mallien arviointi

Taulukot 5–7 esittävät terveydenhuollon, sosiaalihuollon ja vanhustenhuollon vaihtoehtoisten mallien ennustekyvyn mittarit estimointi- ja validointiotoksissa. Mallien ennustekykyä vertaillaan käyttäen mittareina selitysstetta R^2 ja keskineliövirheen neliöjuurta (RMSE). Lopullisen mallin tulokset on kursivoitu. Ensimmäisenä esitetään tulokset demografisista malleista, joissa palvelukäyttöä on selitetty käyttäen ainoastaan ikä- ja sukupuolimuuttujia. Seuraavissa malleissa on vaiheittain lisätty sairausmuuttujat, sosioekonomiset muuttujat ja kokeiltu vaihtoehtoisia määrittelyjä muutamille muuttujille. Taulukkojen viimeisissä sarakkeissa on esitetty tulokset malleista, joissa sairastavuus määritellään vaihtoehtoisilla tavoin. Tilan säästämiseksi taulukoissa ei esitetä täydellistä listausta estimoiduista malleista, mutta ne kertovat olennaisen mallinnusprosessista ja mallien suhteellisista ennustekyvyyistä.

Nähtyjen tulosten perusteella voi yleisenä huomiona todeta, että vain yhden vuoden diagnoositiedon käyttäminen sairausmuuttujissa tuottaa parhaan ennustekyvyn kaikkien sektorien osalta.²⁷ Hoitajakäyntien diagnoositiedon käyttämisen osalta tulokset ovat epäselvät: terveydenhuollon malliin tällä ei käytännössä ole merkitystä, sosiaalihuollon malleissa ennustekyky paranee, mutta vanhustenhuollon osalta hoitajadiagnoosien käyttäminen heikentää ennustekykyä. Uuden sairausluokituksen käyttäminen parantaa selkeästi terveydenhuollon mallia. Vanhustenhuollon mallissa sairausluokituksen päivittämisen tuoma parannus on pienehkö

²⁷ Lisäksi kokeiltiin sisällyttää sairausmuuttujiin kolmen edeltävän vuoden diagnoositieto (tuloksia ei esitetty). Mallien ennustekyky heikentyi poimintavuosien lisääntyessä.

ja sosiaalihuollossa mitätön. Vanhusten- ja sosiaalihuollossa palvelukäyttö aiheutunee terveydenhuoltoon verrattuna pienemmästä määrästä sairauksia, ja selitysasteet ovat jo valmiiksi olennaisesti terveydenhuoltoa korkeammat. Lisäksi tuloksista nähdään, ettei sairauksien karsiminen heikennä huomattavasti ennustekykyä.

Englannin lähes koko ICD-10-tautiluokituksen sisältävän sairausluokituksen käyttäminen parantaisi ennustekykyä selkeähkösti terveydenhuollon mallissa, mutta heikentäisi vanhustenhuollon ja erityisesti sosiaalihuollon mallin selityskykyä. Tälle tulokselle on luonteva tulkinta: terveydenhuollon mallissa kattava sairastavuustiedon sisällyttäminen on merkityksellistä, kun taas vanhusten- ja sosiaalihuollon malleissa tärkeimpien, tarkasti määriteltyjen sairauksien (esimerkiksi dementia ja älyllinen kehitysvammaisuus) sekoittuminen näille palveluluokille vähemmän merkitykselliseen sairaustietoon heikentää olennaisesti mallien ennustekykyä. Tulos osoittaa, ettei sairausluokituksen kärsivällinen asiantuntijatietämykseen perustuva kehittäminen ole turhaa, mutta erityisesti terveydenhuollon mallin kannalta luokitus ei vielä nykyiselläänkään ole tarpeeksi kattava.

Muuten kuin sairausluokituksen osalta voidaan todeta, että lisähyöty mallien monimutkaistamisesta ei ole suuri. Palvelunkäytön selittämiselle keskeistä on sairauksiin ja toimintakykyyn liittyvät tiedot, eikä esimerkiksi sosioekonomista taustaa koskevien tietojen monimutkaistaminen paranna malleja. Esimerkiksi tarkemmalla taustamaaluokittelulla, käyttämällä korkeamman asteen polynomeja jatkuvista muuttujista tai lisäämällä monisairastavuusmuuttujia, ei saavuteta juurikaan parempia malleja. Tarkemman ikä- ja sukupuoliryhmittelyn käyttäminen ei myöskään paranna minkään palvelusektorin mallia huomattavasti. Selitettävien kustannusmuuttujien tyypistäminen Chaplinin ym. (2016) ja Häkkisen ym. (2020) tapaan ei myöskään vaikuta näiden tulosten valossa järkevältä. Taulukosta 5 myös nähdään, että myös terveydenhuollon mallissa asuntokunnan käyttötulo kannattaa huomioida luonnollisena logaritmina.

Lisäksi nähdään, ettei näin suurilla aineistoilla jaotellulla estimointi- ja validointiotoksiin ole erityisen suurta merkitystä, vaan molemmista otoksista lasketut kriteerit ohjaavat samoihin valintoihin. Mallien ennustekyky ei kovin helposti lähde heikkenemään ylimääräisen monimutkaisuuden vuoksi, mikä nähdään esimerkiksi siitä, että sekä terveyden-, vanhusten- että sosiaalihuollon tapauksessa sairausindikaattorien lisääminen myös toiselle aikaisemmalle vuodelle parantaisi ennustekykyä.²⁸

²⁸ Myös kolmannen edeltävän vuoden sairausindikaattorien lisääminen parantaisi ennustekykyä, mutta suhteellisesti huomattavasti vähemmän (tuloksia ei esitetty).

Taulukko 5. Terveydenhuollon mallit, ennustekyky.

	Estimointiotos		Validointiotos	
	R2	RMSE	R2	RMSE
Uusi sairausluokitus, 1 v poiminta-aika				
Demografia1	0,0120	10786,6	0,0139	9964,8
Demografia1 + täysi sairastavuus + synnytys1	0,0608	10516,7	0,0692	9681,1
Demografia1 + täysi sairastavuus + synnytys1 + taustamuuttujat1	0,0628	10505,7	0,0715	9669,2
Demografia1 + täysi sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1	0,0628	10505,7	0,0715	9669,2
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1	0,0629	10504,9	0,0717	9668,4
Demografia3 + täysi sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1	0,0626	10506,8	0,0713	9670,4
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat2 + tarjonta1	0,0627	10506,3	0,0715	9669,4
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat3 + tarjonta1	0,0629	10505,0	0,0716	9668,6
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat4 + tarjonta1	0,0621	10509,4	0,0706	9674,2
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat5 + tarjonta1	0,0625	10507,4	0,0713	9670,5
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1 + monisairastavuus	0,0631	10503,8	0,0718	9667,7
Demografia2 + täysi sairastavuus + synnytys2 + taustamuuttujat1 + tarjonta2 + monisairastavuus	0,0631	10503,8	0,0718	9667,7
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1 + Kelan korvaamat	0,0632	10503,2	0,0719	9667,3
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1, kustannukset typistetty 200 000 euroon	0,0578	10533,6	0,0675	9689,9
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1, kustannukset typistetty 100 000 euroon	0,0564	10541,5	0,0661	9697,6
<i>Demografia2 + karsittu sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1</i>	<i>0,0627</i>	<i>10506,1</i>	<i>0,0714</i>	<i>9670,0</i>
Uusi sairausluokitus, 2 v poiminta-aika				
Demografia2 + karsittu sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1	0,0569	10538,8	0,0639	9708,7
Uusi sairausluokitus, 1 v poiminta-aika, sis. hoitajakäynnit				
Demografia2 + karsittu sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1	0,0627	10506,0	0,0714	9669,7
Vanha sairausluokitus, 1 v poiminta-aika				
Demografia2 + karsittu sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1	0,0572	10536,9	0,0650	9703,0
Uusi sairausluokitus, 1 v poiminta-aika, sairausindikaattorit kahdelle vuodelle				
Demografia2 + karsittu sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1	0,0642	10498,1	0,0727	9662,8
Englannin sairausluokitus, 1 v poiminta-aika				
Demografia2 + täysi sairastavuus + synnytys1 + taustamuuttujat1 + tarjonta1	0,0676	10478,6	0,0766	9642,7

Demografia1: indikaattorit naissukupuolelle ja 19 ikäryhmälle taulukossa 2 raportoitua mallia vastaavasti, mutta ilman interaktiota sukupuolen kanssa. Demografia2: indikaattorit taulukossa 2 raportoitua mallia vastaavasti (19 ikäryhmää miehille ja 20 naisille). Demografia3: indikaattorit naissukupuolelle ja iälle (3. asteen polynomi). Täysi sairastavuus: koko päivitetty sairausluokitus, mukaan lukien WHO:n näkövammaluokitus. Karsittu sairastavuus: tiputettu pois sairauksia, jotka saavat negatiivisen regressiokertoimen tai joiden vaikutus tulevaan palvelukäyttöön ei ole selkeästi perusteltavissa. Synnytys1: indikaattorit synnytykselle ja keskosuuden enenaikaisuudelle (päivissä, 2. asteen polynomi). Synnytys2: muutoin kuin synnytys1, mutta keskosuuden enenaikaisuus 3. asteen polynomina. Taustamuuttujat1: indikaattorit työkyvyttömyyseläkkeelle (viisi ikäkvintiliä), pääasialliselle toiminnalle (työssäkäyvä, opiskelija, varusmies), koulutustasolle (toinen aste, korkeakoulutus), siviilisäädylle (naimaton, eronnut, leski), perhetilanteelle (yhden aikuisen perhe, yksinasuva alle 75-vuotias, yksinasuva 75–84-vuotias, yksinasuva 85–89-vuotias, yksinasuva 90-vuotias) ja ulkomaa-laistautustalle, sekä luonnollisen logaritmin asuntokunnan käyttötuloista kulutusyksikköä kohden. Taustamuuttujat2: muutoin vastaava, mutta käyttötulot 3. asteen polynomina. Taustamuuttujat3: kuten taustamuuttujat1, mutta taustamaa vastaavina indikaattoreina kuin Häkkinen ym. (2020). Taustamuuttujat4: kuten taustamuuttujat1, mutta työkyvyttömyyseläke yhtenä indikaattorina, ilman ikäjaottelua. Taustamuuttujat5: kuten taustamuuttujat1, mutta yksinasuva yhtenä indikaattorina, ilman ikäjaottelua. Tarjonta1: etäisyys erikoissairaanhoidon päivystykseen 2. asteen polynomina; Tarjonta2: etäisyys erikoissairaanhoidon päivystykseen 3. asteen polynomina. Monisairastavuus: sairauksien määrä (sairausluokitukseen sisältyviä sairauksia yli yhden sairauden) 3. asteen polynomina. Kelan korvaamat: sairausvakuutuksen korvaamien avohoidon lääkkeiden, yksityislääkärikäyntien (ja tutkimuksen ja hoidon) ja yksityisen suun terveydenhuollon kustannukset erillisinä muuttujina, kuten raportissa Häkkinen ym. (2020). Valittu malli kursivoituna.

Taulukko 6. Sosiaalihuollon mallit, ennustekyky.

	Estimointiotos		Validointiotos	
	R2	RMSE	R2	RMSE
Uusi sairausluokitus, 1 v poiminta-aika				
Demografia1	0,0042	5255,7	0,0044	5332,3
Demografia1 + täysi sairastavuus	0,1961	4722,3	0,2099	4750,0
Demografia1 + täysi sairastavuus + taustamuuttajat1	0,2164	4662,2	0,2320	4683,1
Demografia2 + täysi sairastavuus + taustamuuttajat1	0,2165	4662,1	0,2320	4683,1
Demografia3 + täysi sairastavuus + taustamuuttajat1	0,2145	4667,8	0,2301	4688,9
Demografia1 + täysi sairastavuus + taustamuuttajat2	0,2025	4703,5	0,2163	4730,7
Demografia1 + täysi sairastavuus + taustamuuttajat3	0,2159	4663,7	0,2316	4684,5
Demografia1 + täysi sairastavuus + taustamuuttajat4	0,2152	4665,8	0,2313	4685,4
Demografia1 + täysi sairastavuus + taustamuuttajat5	0,2169	4660,6	0,2325	4681,7
Demografia1 + täysi sairastavuus + taustamuuttajat1 + yksin/laitoksessa	0,2164	4662,2	0,2320	4683,1
Demografia1 + täysi sairastavuus + taustamuuttajat1 + monisairastavuus	0,2164	4662,1	0,2321	4683,0
Demografia1 + täysi sairastavuus + taustamuuttajat1 + toimeentulotuki	0,2165	4662,0	0,2321	4682,8
Demografia1 + täysi sairastavuus + taustamuuttajat1, kustannukset typistetty 200 000 euroon	0,2164	4662,2	0,2320	4683,1
Demografia1 + täysi sairastavuus + taustamuuttajat1, kustannukset typistetty 100 000 euroon	0,2160	4663,5	0,2310	4686,2
<i>Demografia1 + karsittu sairastavuus + taustamuuttajat1</i>	<i>0,2157</i>	<i>4664,4</i>	<i>0,2315</i>	<i>4684,6</i>
Uusi sairausluokitus, 2 v poiminta-aika				
Demografia1 + karsittu sairastavuus + taustamuuttajat1	0,1971	4719,4	0,2120	4743,9
Uusi sairausluokitus, 1 v poiminta-aika, sis. hoitajakäynnit				
Demografia1 + karsittu sairastavuus + taustamuuttajat1	0,2158	4664,0	0,2320	4683,2
Vanha sairausluokitus, 1 v poiminta-aika				
Demografia1 + karsittu sairastavuus + taustamuuttajat1	0,2139	4669,6	0,2306	4687,6
Uusi sairausluokitus, 1 v poiminta-aika, sairausindikaattorit kahdelle vuodelle				
Demografia1 + karsittu sairastavuus + taustamuuttajat1	0,2212	4648,1	0,2359	4671,3
Englannin sairausluokitus, 1 v poiminta-aika				
Demografia1 + täysi sairastavuus + taustamuuttajat1	0,1903	4739,4	0,1986	4783,9

Demografia1: indikaattorit naissukupuolelle ja 19 ikäryhmälle taulukossa 3 raportoituja mallia vastaavasti. Demografia2: indikaattorit taulukossa 3 raportoituja mallia vastaavasti (19 ikäryhmää miehille ja 20 naisille). Demografia3: indikaattorit naissukupuolelle ja iälle (3. asteen polynomi). Täysi sairastavuus: koko päivitetty sairausluokitus, mukaan lukien WHO:n näkövammaisuus. Karsittu sairastavuus: tiputettu pois sairauksia, jotka saavat negatiivisen regressiokertoimen tai joiden vaikutus tulevaan palvelukäyttöön ei ole selkeästi perusteltavissa. Taustamuuttajat1: indikaattorit työkyvyttömyyseläkkeelle (viisi ikäkvintiiliä), pääasialliselle toiminnalle (työssäkäyvä, opiskelija, varusmies), koulutustasolle (toinen aste, korkeakoulutus), siviilisäädylle (naimaton, eronnut, leski), perhe-tilanteelle (yhden aikuisen perhe) ja ulkomaalaistaustalle, sekä luonnollisen logaritmin asutokunnan käyttötuloista kulutusyksikköä kohden. Taustamuuttajat2: muutoin vastaava, mutta käyttötulot 3. asteen polynomina. Taustamuuttajat3: kuten taustamuuttajat1, mutta taustamaa vastaavina indikaattoreina kuin Häkkinen ym. (2020). Taustamuuttajat4: kuten taustamuuttajat1, mutta työkyvyttömyyseläke yhtenä indikaattorina, ilman ikäjaottelua. Taustamuuttajat5: kuten taustamuuttajat1, mutta yksinasumisen indikaattoreilla (yksinasuva ikäryhmissä alle 75-vuotias, 75–84-vuotias, 85–89-vuotias, yli 90-vuotias). Yksin/laitoksessa: Häkkinen ym. raporttia (2020) seuraten indikaattori henkilöille, jotka asuvat yksin tai laitoksessa. Monisairastavuus: lisäsairauksien määrä (sairausluokitukseen sisältyviä sairauksia yli yhden sairauden) 3. asteen polynomina. Toimeentulotuki: indikaattori henkilöille, jotka saaneet perustoimeentulotukea. Valittu malli kursivoituna.

Taulukko 7. Vanhustenhuollon mallit, ennustekyky.

	Estimointiotos		Validointiotos	
	R2	RMSE	R2	RMSE
Uusi sairausluokitus, 1 v poiminta-aika				
Demografia1	0,1201	17005,9	0,1242	17047,6
Demografia1 + täysi sairastavuus	0,2790	15394,0	0,2821	15434,2
Demografia1 + täysi sairastavuus + taustamuuttajat1	0,2988	15180,2	0,3021	15217,6
Demografia2 + täysi sairastavuus + taustamuuttajat1	0,2998	15170,1	0,3031	15206,6
Demografia3 + täysi sairastavuus + taustamuuttajat1	0,2993	15175,2	0,3023	15215,4
Demografia2 + täysi sairastavuus + taustamuuttajat2	0,2853	15326,6	0,2884	15366,4
Demografia2 + täysi sairastavuus + taustamuuttajat3	0,2997	15171,1	0,3030	15207,8
Demografia2 + täysi sairastavuus + taustamuuttajat1 + monisairastavuus	0,2999	15169,2	0,3032	15205,4
Demografia2 + täysi sairastavuus + taustamuuttajat1 + yksin/laitoksessa + toimeentulotuki	0,3054	15109,0	0,3088	15144,6
Demografia2 + täysi sairastavuus + taustamuuttajat4	0,2999	15168,3	0,3033	15204,8
Demografia2 + täysi sairastavuus + taustamuuttajat1, kustannukset typistetty 200 000 euroon	0,2998	15170,4	0,3031	15207,3
Demografia2 + täysi sairastavuus + taustamuuttajat1, kustannukset typistetty 100 000 euroon	0,2984	15185,5	0,3015	15224,9
<i>Demografia2 + karsittu sairastavuus + taustamuuttajat1</i>	<i>0,2985</i>	<i>15183,9</i>	<i>0,3020</i>	<i>15218,6</i>
Uusi sairausluokitus, 2 v poiminta-aika				
Demografia2 + karsittu sairastavuus + taustamuuttajat1	0,2921	15252,9	0,2953	15292,3
Uusi sairausluokitus, 1 v poiminta-aika, sis. hoitajakäynnit				
Demografia2 + karsittu sairastavuus + taustamuuttajat1	0,2970	15200,4	0,3004	15236,2
Vanha sairausluokitus, 1 v poiminta-aika				
Demografia2 + karsittu sairastavuus + taustamuuttajat1	0,2925	15248,9	0,2967	15276,8
Uusi sairausluokitus, 1 v poiminta-aika, sairausindikaattorit kahdelle vuodelle				
Demografia2 + karsittu sairastavuus + taustamuuttajat1	0,3057	15105,4	0,3096	15136,3
Englannin sairausluokitus, 1 v poiminta-aika				
Demografia2 + täysi sairastavuus + taustamuuttajat1	0,2957	15214,0	0,2971	15272,0

Demografia1: indikaattorit naissukupuolelle ja 7 ikäryhmälle taulukossa 4 raportoituja mallia vastaavasti, mutta ilman interaktiota sukupuolen kanssa. Demografia2: indikaattorit taulukossa 4 raportoidun mallin mukaisesti (6 ikäryhmää miehille ja 7 naisille). Demografia3: indikaattorit naissukupuolelle ja iälle (3. asteen polynomina). Täysi sairastavuus: koko päivitetty sairausluokitus, mukaan lukien WHO:n näkövammaluokitus. Karsittu sairastavuus: tiputettu sairauksia, jotka saavat negatiivisen regressiokertoimen tai joiden vaikutus tulevaan palvelukäyttöön ei ole selkeästi perusteltavissa. Taustamuuttajat1: indikaattorit koulutustasolle (toinen aste, korkeakoulu), siviilisäädyille (naimaton, eronnut, leski) ja ulkomaalaistaustalle, sekä luonnollinen logaritmi asutokunnan käyttötuloista kulutusyksikköä kohden. Taustamuuttajat2: muutoin edellistä vastaava, mutta käyttötulot 3. asteen polynomina. Taustamuuttajat3: kuten taustamuuttajat1, mutta taustamaa vastaavina indikaattoreina kuin Häkkinen ym. (2020). Taustamuuttajat4: kuten taustamuuttajat1, mutta lisätty indikaattorit työssäkäynnille ja työkyvyttömyyseläkkeelle. Monisairastavuus: lisäsairauksien määrä (sairausluokitukseen sisältyviä sairauksia yli yhden sairauden) 3. asteen polynomina. Yksin/laitoksessa: Häkkinen ym. raporttia (2020) seuraten indikaattori henkilöille, jotka asuvat yksin tai laitoksessa. Toimeentulotuki: indikaattori henkilöille, jotka saaneet perustoimeentulotukea. Valittu malli kursivoituna.

4.3 Tarvekertoimet

Sektorikohtaiset tarvekertoimet ja niiden painotettuna keskiarvona laskettu kokonaistarvekerroin on esitetynä taulukossa 8 vuosille 2017–2019. Vuoden 2019 tarvekerroin on ehdotettu osaksi hyvinvointialuiden rahoitusta vuodelle 2023, mutta hyvinvointialueiden rahoituslain mukaisesti seuraavien vuosien rahoituksessa käytettäisiin vuosittain päivittyviä tietoja. Vertailun vuoksi taulukossa esitetään myös vuosien 2017 ja 2018 väestö- ja tarvetekijätietoa käyttävät tarvekertoimet, joiden osalta kyseessä on ”taaksepäin ennustaminen”. Muutamilla pienillä hyvinvointialueilla, erityisesti Keski-Pohjanmaan hyvinvointialueella, tarvekerroin muuttuu huomattavasti yli ajan.

Taulukko 8. Terveyden-, sosiaali- ja vanhustenhuollon tarvekertoimet sekä kokonaistarvekerroin hyvinvointialueittain vuosina 2017–2019.

	Hyvinvointialue	Vuosi	Terveydenhuollon tarvekerroin	Sosiaalihuollon tarvekerroin	Vanhustenhuollon tarvekerroin	Kokonaistarvekerroin
1	Itä-Uusimaa	2017	0,981	0,870	0,907	0,943
		2018	0,990	0,874	0,925	0,953
		2019	0,983	0,867	0,914	0,945
2	Keski-Uusimaa	2017	0,946	0,868	0,736	0,888
		2018	0,946	0,871	0,755	0,892
		2019	0,939	0,866	0,750	0,886
3	Länsi-Uusimaa	2017	0,864	0,763	0,607	0,792
		2018	0,860	0,758	0,615	0,790
		2019	0,858	0,753	0,615	0,787
4	Vantaa ja Kerava	2017	0,901	0,831	0,572	0,821
		2018	0,904	0,841	0,569	0,824
		2019	0,872	0,826	0,555	0,800
5	Varsinais-Suomi	2017	1,027	1,007	1,093	1,036
		2018	1,026	1,010	1,086	1,034
		2019	1,015	1,008	1,083	1,027
6	Satakunta	2017	1,046	1,093	1,178	1,082
		2018	1,039	1,090	1,167	1,075
		2019	1,044	1,103	1,162	1,080
7	Kanta-Häme	2017	1,059	1,011	1,105	1,058
		2018	1,051	1,009	1,087	1,049
		2019	1,052	1,020	1,091	1,053
8	Pirkanmaa	2017	1,005	0,960	1,005	0,995
		2018	1,003	0,953	1,003	0,992
		2019	0,998	0,944	0,999	0,986
9	Päijät-Häme	2017	1,053	1,023	1,120	1,060
		2018	1,069	1,050	1,168	1,085
		2019	1,082	1,055	1,170	1,093
10	Kymenlaakso	2017	1,042	1,076	1,297	1,100
		2018	1,088	1,103	1,344	1,142
		2019	1,096	1,127	1,378	1,158
11	Etelä-Karjala	2017	1,043	1,028	1,245	1,079
		2018	1,044	1,018	1,233	1,076
		2019	1,033	1,013	1,207	1,063
12	Etelä-Savo	2017	1,139	1,222	1,467	1,221
		2018	1,137	1,225	1,463	1,221
		2019	1,142	1,240	1,456	1,225
13	Pohjois-Savo	2017	1,132	1,212	1,253	1,173
		2018	1,141	1,213	1,257	1,180
		2019	1,135	1,211	1,260	1,176

	Hyvinvointialue	Vuosi	Terveystarvekerroin	Sosiaalihuollon tarvekerroin	Vanhustenhuollon tarvekerroin	Kokonaistarvekerroin
14	Pohjois-Karjala	2017	1,126	1,272	1,316	1,195
		2018	1,128	1,270	1,295	1,192
		2019	1,202	1,304	1,361	1,255
15	Keski-Suomi	2017	1,010	1,042	1,049	1,025
		2018	1,009	1,052	1,045	1,025
		2019	0,990	1,042	1,022	1,008
16	Etelä-Pohjanmaa	2017	1,094	1,090	1,294	1,133
		2018	1,085	1,095	1,289	1,127
		2019	1,095	1,103	1,280	1,134
17	Pohjanmaa	2017	0,944	0,844	0,985	0,931
		2018	0,950	0,849	0,979	0,934
		2019	0,957	0,859	0,987	0,942
18	Keski-Pohjanmaa	2017	1,086	1,083	1,112	1,090
		2018	1,131	1,123	1,177	1,138
		2019	1,142	1,149	1,233	1,161
19	Pohjois-Pohjanmaa	2017	0,971	1,134	0,929	0,997
		2018	0,966	1,138	0,938	0,997
		2019	0,976	1,146	0,965	1,010
20	Kainuu	2017	1,117	1,264	1,355	1,195
		2018	1,132	1,280	1,382	1,213
		2019	1,152	1,292	1,419	1,235
21	Lappi	2017	1,086	1,300	1,188	1,152
		2018	1,087	1,310	1,182	1,153
		2019	1,086	1,314	1,189	1,155
90	Helsinki	2017	0,905	0,843	0,772	0,865
		2018	0,893	0,822	0,759	0,851
		2019	0,898	0,816	0,752	0,852

Taulukoissa 2–4 esitettyjen mallien mukaiset tarvekertoimet. Tarvekerroin kertoo alueen asukkaiden keskimääräisen palvelutarpeen suhteessa maan keskiarvoiseen. Tarvekertoimien väkilukupainotettu keskiarvo on siis yksi. Vuosi viittaa siihen vuoteen, jonka väestö- ja tarvetekijätietoa on käytetty. Kokonaistarvekerroinlaskettu painotettuna keskiarvona terveyden-, vanhusten- ja sosiaalihuollon mallien perusteella lasketuista tarvekertoimista. Painot (terveydenhuolto 0,5886, vanhustenhuolto 0,1975 ja sosiaalihuolto 0,2139) perustuvat Kuntataloustilaston tietoihin sektorikohtaisista toteutuneista menoista vuonna 2020. Sosiaali- ja vanhustenhuollon tarvekertoimien laskennassa neutralisoitu taustamaa. Terveystarvekerroinlaskennassa neutralisoitu taustamaa ja etäisyys erikoissairaanhoidon.

Taulukossa 9 on esitettyä vuoden 2019 kokonaistarvekertoimet erilaisilla mallimäärittelyillä. Kolmannessa sarakkeessa kursivoituna esitetty neutralisointi2-tarvekerroin on rahoitukseen esitetty vuoden 2019 tarvekerron, joka löytyy myös taulukosta 8.

Keskeisenä huomiona taulukosta 9 – ja tämän raportin tärkeimpänä tuloksena – voidaan todeta, että siirtyminen vanhasta sairausluokitukselta uuteen sairausluokitukselle kasvattaa olennaisesti Uudellemaalle arvioitua palvelutarvetta ja vähentää sitä pääosin muuta maata (vertaa saraketta 3 sarakkeeseen 9). Myös Pohjois-Karjalan ja Keski-Pohjanmaan arvioitu palvelutarve kasvaa huomattavasti. Muista mallinnoista johtuen sarakkeessa 9 esitetty vanhan sairausluokituksen mukainen tarvekerroin ei vastaa tarkalleen aiemman mallinnoituksen (Häkkinen ym. 2020) mukaista vuoden 2019 tarvekertoimista, mutta on kuitenkin varsin lähellä sitä (vertailua ei esitetty tässä). Muutokset tarvekertoimissa selittyvät siis pääosin sairausluokituksen muutoksilla, ts. päivitettyllä luokituksella havaitaan palvelujen tarvetta paremmin. Vertaamalla taulukon 8 vuosittaisia tarvekertoimia nähdään myös, että muutos pätee suhteellisen vakaasti yli ajan eikä ole vuoden 2019 aineistosta riippuvainen: siirtyminen uuteen sairausluokitukselle aiheuttaa tasomuutoksen tarvekertoimissa.

Taustamaan neutralisointi vaikuttaa selkeästi erityisesti Helsingin tarvekertoimeen (vertaa taulukon 9 sarakkeita 2 ja 1). Koska henkilöt, joiden taustamaa ei ole Suomi, käyttävät kaikilla sektoreilla muuhun

väestöön verrattuna vähemmän palveluita (muut tekijät vakioituna), vaikuttaa muuttuja negatiivisesti arvioituun tarpeeseen alueilla, joilla on suhteellisesti enemmän maahanmuuttajaväestöä. Siksi sen neutralisointi kasvattaa näiden alueiden, erityisesti Helsingin, arvioitua palvelutarvetta. Erikoissairaanhoidon päivystyksen matka-ajan neutralisoinnilla ei ole suhteessa kovin suurta vaikutusta. Jos malleissa käytettäisiin päivitettyä sairausluokitusta kokonaisuudessaan ilman mallittaista muuttujakarsintaa, saataisiin sarakkeessa 4 esitetyt tarvekertoimet. Vertaamalla niitä sarakkeen 3 tarvekertoimiin nähdään, että karsinnalla on joillain alueilla, kuten Pohjois-Karjalassa, jonkin verran vaikutusta. Tämä johtunee muutamista, huomattavan negatiivisen kertoimen saavista sairauksista (erityisesti älyllinen kehitysvammaisuus terveydenhuollossa). Vaikka hoitajakäyntien diagnoositiedon käyttämisellä ei ollut suurta vaikutusta taulukoissa 5–7 nähtyyn mallien ennustekykyyneen, nähdään taulukon 9 sarakkeesta 5, että erityisesti Etelä-Savossa niiden huomioinen vaikuttaisi kuitenkin positiivisesti palvelutarpeeseen. Kuten sairauksien karsimisellakaan, hoitajakäyntien huomioimisella ei kuitenkaan ole kokonaiskuvassa suurta merkitystä.

Lisäksi voidaan todeta, ettei ole kovin suurta merkitystä, minkä vuoden aineistoa käytetään mallien estimoinnissa (vertaa saraketta 3 sarakkeisiin 6 ja 7). Sarakkeen 8 malliin on sisällytetty tarjontatekijöitä kontrolloimaan hyvinvointialueittaiset indikaattorit, jotka on tarvekertoimen laskennassa neutralisoitu. Niiden suhteen voidaan jälleen todeta vaikutuksen olevan kokonaisuudessaan suhteellisen vähäinen, mutta suurimpien vaikutusten keskittyvän tietyille alueille (Lappi ja Pohjois-Karjala).

Englannin sairausluokituksen (Chaplin ym. 2016) käyttämisellä olisi huomattavasti vaikutusta erityisesti Helsingin, Päijät-Hämeen, Etelä-Karjalan ja Keski-Suomen tarvekertoimiin. Koska Englannin sairausluokituksen käyttäminen parantaisi terveydenhuollon mallia mutta heikentäisi vanhusten- ja sosiaalihuollon malleja (ja siis muuttaisi vanhusten- ja sosiaalihuollon malleja enemmän asukasperusteisiksi), se muuttaisi tarvekiointia enemmän terveydenhuollolle tärkeisiin sairauksiin ja vähemmän vanhusten- ja sosiaalihuollossa esiintyvään sairastavuuteen perustuvaksi.

Taulukko 9. Eri mallityyppien tuottamat vuoden 2019 kokonaistarvekertoimet.

Hyvinvointialue	Perusmalli	Neutralisointi1	Neutralisointi2	Täysi sairastavuus	Hoitajakäynnit	Vuoden 2017 aineisto	Vuoden 2018 aineisto	Hv-alueindikaattorit neutralisoituna	Vanha sairausluokitus	Englannin sairausluokitus
Itä-Uusimaa	0,944	0,941	0,945	0,945	0,952	0,949	0,948	0,950	0,915	0,953
Keski-Uusimaa	0,889	0,884	0,886	0,887	0,883	0,891	0,891	0,891	0,864	0,895
Länsi-Uusimaa	0,775	0,787	0,787	0,790	0,788	0,792	0,793	0,794	0,775	0,775
Vantaa ja Kerava	0,779	0,804	0,800	0,803	0,799	0,801	0,803	0,805	0,774	0,799
Varsinais-Suomi	1,028	1,028	1,027	1,030	1,026	1,028	1,028	1,027	1,027	1,010
Satakunta	1,088	1,079	1,080	1,082	1,079	1,081	1,080	1,078	1,103	1,082
Kanta-Häme	1,059	1,052	1,053	1,053	1,053	1,057	1,056	1,054	1,066	1,031
Pirkanmaa	0,993	0,987	0,986	0,986	0,984	0,988	0,989	0,988	0,992	0,992
Päijät-Häme	1,100	1,095	1,093	1,093	1,095	1,092	1,091	1,093	1,108	1,139
Kymenlaakso	1,159	1,156	1,158	1,161	1,158	1,155	1,154	1,156	1,179	1,150
Etelä-Karjala	1,065	1,063	1,063	1,063	1,073	1,061	1,060	1,064	1,092	1,117
Etelä-Savo	1,234	1,225	1,225	1,223	1,235	1,221	1,219	1,219	1,248	1,230
Pohjois-Savo	1,183	1,173	1,176	1,170	1,177	1,173	1,173	1,170	1,172	1,161
Pohjois-Karjala	1,261	1,253	1,255	1,246	1,259	1,249	1,247	1,246	1,204	1,242
Keski-Suomi	1,015	1,007	1,008	1,007	1,009	1,006	1,007	1,006	1,029	1,030
Etelä-Pohjanmaa	1,142	1,131	1,134	1,130	1,131	1,134	1,131	1,131	1,150	1,128
Pohjanmaa	0,942	0,942	0,942	0,945	0,938	0,944	0,944	0,947	0,951	0,940
Keski-Pohjanmaa	1,172	1,163	1,161	1,154	1,157	1,163	1,158	1,158	1,145	1,180
Pohjois-Pohjanmaa	1,017	1,007	1,010	1,007	1,012	1,007	1,009	1,006	1,029	0,987
Kainuu	1,243	1,232	1,235	1,228	1,239	1,229	1,227	1,227	1,261	1,227
Lappi	1,162	1,152	1,155	1,153	1,152	1,151	1,150	1,146	1,164	1,137
Helsinki	0,836	0,856	0,852	0,855	0,849	0,850	0,851	0,854	0,830	0,867

Esitetyt kokonaistarvekertoimet käyttävät vuoden 2019 sairaustietoja. Tarvekertoimet on laskettu painotettuna keskiarvoina terveyden-, vanhusten- ja sosiaalihuollon mallien perusteella lasketuista tarvekertoimista. Painot (terveydenhuolto 0,5886, vanhustenhuolto 0,1975 ja sosiaalihuolto 0,2139) perustuvat Kuntataloustilaston tietoihin sektorikohtaisista menoista (v. 2020). Kolmen ensimmäisen sarakkeen tarvekertoimet ovat rahoitukseen ehdotettujen mallien (taulukot 2–4) mukaisia. Neutralisointi1 tarkoittaa, että taustamaa on neutralisoitu ja Neutralisointi2, että sekä taustamaa että etäisyys erikoissairaanhoidon päivystykseen on neutralisoitu; nämä tekijät on neutralisoitu myös kaikissa sarakkeiden 4–10 tarvekertoimissa. Sarakkeiden 4–10 tarvekertoimet on laskettu malleista, joissa rahoitukseen ehdotettuihin perusmalleihin on tehty seuraavia muutoksia. Täysi sairastavuus: kaikissa kolmessa mallissa on käytetty koko päivitettyä sairausluokitusta. Tarvekerroin, v. 2017: laskettu malleista, joissa on käytetty (muidenkin kuin sosiaalihuollon osalta) vuoden 2017 palvelukäyttöaineistoa. Tarvekerroin, v. 2018: käytetty 2018 aineistoa (sosiaalihuollossa 2017). Hv-alueindikaattorit neutralisoituna: malleihin on sisällytetty hyvinvointialueille indikaattorit (sosiaalihuollon mallissa indikaattorit Eksotelle, Päijät-Hämeen hyvinvointikuntayhtymälle, Oululle ja Helsingille). Vanha sairausluokitus: käytetty raportissa Häkkinen ym. (2020) esitettyä sairausluokitusta; Englannin sairausluokitus: Chaplin ym. (2016) käyttämä sairausluokitus. Muilta osin mallit vastaavat taulukoiden 2–4 malleja. Rahoitukseen ehdotetut tarvekertoimet esitetty kursivoituna.

5 Johtopäätökset

Tutkimuksen tulokset osoittavat, että sairausluokituksen kehittämisellä on vaikutusta palvelutarpeen arvioinnin malleihin ja niiden perusteella arvioituun suhteelliseen palvelutarpeeseen. Päivitetyn sairausluokituksen mukaisten tulosten vertailu vanhan sairausluokituksen (Häkkinen ym. 2020) mukaisiin malleihin osoittaa, että mallien selityskyky paranee. Tämä näkyy myös mallien perusteella lasketuissa tarvekertoimissa: arvioitu palvelutarve kasvaa selkeästi Uudellamaalla, Pohjois-Karjalassa ja Keski-Pohjanmaalla suhteessa muuhun maahan. Lisäksi tuloksista nähdään, ettei pienehköillä mallimuutoksilla ole huomattavaa vaikutusta mallien hyvytyteen tai niiden perusteella laskettuihin tarvekertoimiin.

Henkilön taustamaalla on huomattava vaikutus palvelukäyttöön kaikissa malleissa. Siksi myös sen neutralisointi vaikuttaa selkeästi tarvekertoimiin. Neutralisointi on kiistatta normatiivinen valinta, mutta se tuntuisi hyvin perustellulta: maahanmuuttajien vähäisempi palvelukäyttö kertoo todennäköisemmin tyydyttämättömästä palvelutarpeesta kuin vähäisemmästä tarpeesta palveluille. Laajempi kansallinen keskustelu tekijöistä, joiden halutaan vaikuttavan arvioituun palvelutarpeeseen ja siten rahoitukseen, olisi kuitenkin tärkeää.

Mallien määrittelyssä on pyritty sikäli yksinkertaisuuteen, että malleihin ei ole sisällytetty suurta määrää muuttujia, jotka parantaisivat niitä vain lievästi. Ehdotettuihin malleihin ei ole esimerkiksi sisällytetty sairausindikaattoreita useammalle aikaisemmalle vuodelle, vaikka niiden nähdään hieman parantavan ennustekykyyä. Palvelutarvetta voisi mallintaa myös koneoppimismenetelmillä, mutta tällöin jouduttaisiin tinkimään mallien läpinäkyvyydestä. Mallien selkeyden ja ennustekyvyn välillä on vaihtosuhde, eikä mitään tiettyä monimutkaisuuden astetta voi nostaa selkeästi parhaaksi valinnaksi. Siksi valinta enemmän ja vähemmän monimutkaisten mallien välillä on lopulta myös jossain määrin poliittista harkintaa vaativa arvovalinta.

Englannin NHS:n G&A-mallin sairausluokituksen (Chaplin ym. 2016) käyttämisellä olisi huomattavasti vaikutusta arvioihin palvelutarpeesta, mikä osaltaan korostaa tarvetta kehittää sairausluokitusta edelleen. Englannin sairausluokitusta ei ole perusteltua ottaa suoraan käyttöön, sillä vaikka se toimii hyvin terveydenhuollon mallissa, ei se kuvaa hyvin vanhusten- ja sosiaalihuollon palveluiden käyttöä. Toteutetuista tarkasteluista on kuitenkin nähtävissä, että erityisesti terveydenhuollon palvelutarpeen mallintamista on mahdollista yhä kehittää paremmalla sairausluokittelulla. Sairausluokituksen jatkokehittämisen ohella jatkotutkimuksessa tulee tarkastella mahdollisuutta toteuttaa akuuttien sairauksien aiheuttaman palvelutarpeen huomiointia esimerkiksi jonkinlaisen *concurrent*-mallinnuksella toteutetun jälkikäteiskompensaation kautta, esimerkiksi mukaillen Dudleyn ym. (2003) mallia.

Yksi keskeinen kysymys jatkotyölle on aineistojen hitaasta valmistumisesta rahoitukseen aiheutuvan viiveen korjaaminen. Lisäksi tulevan palvelutarpeen mallinnukseen liittyvän epävarmuuden kvantifiointi mielekkäällä tavoin olisi tärkeää lisäämään rahoitusjärjestelmän uskottavuutta ja parantamaan sen ennustettavuutta. Hankkeessa kerätty aineisto tarjoaa mahdollisuuden kehittää tarveperusteista rahoitusta vastaamaan myös tuleviin muutoksiin rahoitusjärjestelmässä, esimerkiksi monikanavarahoituksen purkua koskien.

Tutkimuksen heikkoutena on edelleen se, ettei tarjontatekijöiden mahdollista vaikutusta ole kontrolloitu kovin uskottavalla joukolla muuttujia. Kuitenkin raportoitujen alustavien kokeilujen, joissa malleihin sisällytettiin hyvinvointialueindikaattorit, perusteella tarjontatekijöiden huomioimisen vaikutus voisi lopulta jäädä suhteellisen vähäiseksi.

Erytisesti terveydenhuollon mallista huomataan jälleen, että selitysaste jää vaatimattomaksi. Tämä on kuitenkin melko väistämätön seuraus palvelukäytön vahvasti vinoutuneesta jakaumasta: jotkut henkilöt käyttävät vuosittain palveluja jopa miljoonien eurojen edestä. Tällaisilla havainnoilla on suuri vaikutus selityssasteeseen, ja samaan aikaan niitä on vaikea mallintaa. Edistyneemmällä mallinnusmenetelmillä ja/tai tarkemmalla sairausluokituksella pystytään parempiin selityssasteisiin, mutta hintana on monimutkaisuuden lisääntyminen. Joka tapauksessa palvelukäytön hankala jakauma on terveydenhuollon todellisuutta, minkä mikä vain rahoitusjärjestelmä joutuu kohtaamaan. Siksi sitä ei voida pitää tarveperusteisen rahoituksen ominaisongelmana.

Lopuksi on todettava, että kuten muullekin palvelujärjestelmän tutkimukselle, on tarveperusteisen rahoituksen kehittämiseksi ensiarvoisen tärkeää taata laadukas hoitoilmoitusrekisteritieto ja kustannustietojen keräämisen kehittäminen.

Liitteet

Liite 1. Terveysthuollon mallin painokertoimet.

	Painokerroin
Nainen, ikäryhmä	
Ikä 0	1,415
Ikä 1–6	0,904
Ikä 7–12	1,003
Ikä 13–18	1,220
Ikä 19–25	1,021
Ikä 26–30	0,986
Ikä 31–35	0,961
Ikä 36–40	0,937
Ikä 41–45	0,901
Ikä 46–50	0,947
Ikä 51–55	1,000
Ikä 55–60	0,966
Ikä 61–65	0,980
Ikä 66–70	1,070
Ikä 71–75	1,231
Ikä 76–80	1,250
Ikä 81–85	1,374
Ikä 86–90	1,118
Ikä 91–95	0,674
Yli 95 vuotta	0,424
Mies, ikäryhmä	
Ikä 0	1,570
Ikä 1–6	0,996
Ikä 7–12	1,017
Ikä 13–18	0,934
Ikä 19–25	0,928
Ikä 26–30	0,936
Ikä 31–35	0,883
Ikä 36–40	0,904
Ikä 41–45	0,886
Ikä 46–50	0,899
Ikä 51–55	0,957
Ikä 55–60	0,975
Ikä 61–65	1,012
Ikä 66–70	1,132
Ikä 71–75	1,320
Ikä 76–80	1,451
Ikä 81–85	1,635
Ikä 86–90	1,662

	Painokerroin
Ikä 91–95	1,454
Yli 95 vuotta	1,447
Sairastavuus	
Tuberkuloosi	0,759
HIV, C-hepatiitti	1,688
B-hepatiitti	0,490
Ei-tuberkuloottiset mykobakteerit	1,370
Ruusut	0,957
Veneeriset syylät	0,957
Huulen, suun ja nielun pahanlaatuiset kasvaimet	0,198
Ruuansulatuselinten pahanlaatuiset kasvaimet	1,442
Hengityselinten ja rintaontelon elinten pahanlaatuiset kasvaimet	2,377
Luun ja nivelruston pahanlaatuiset kasvaimet	2,151
Ihon melanooma ja muut pahanlaatuiset ihokasvaimet	2,899
Muut pehmytkudoksen pahanlaatuiset kasvaimet	0,354
Rintasyöpä	2,298
Naisen sukupuolielinten pahanlaatuiset kasvaimet	0,945
Miehen sukupuolielinten pahanlaatuiset kasvaimet	1,669
Virtsaelinten pahanlaatuiset kasvaimet	0,761
Silmän, keskushermoston ja aivohermojen pahanlaatuiset kasvaimet	1,090
Kilpirauhasen ja muiden umpirauhasten pahanlaatuiset kasvaimet	1,979
Pahanlaatuiset kasvaimet, joiden sijaintipaikka on epäselvä, sekundaarinen tai määrittämätön	0,874
Imukudoksen, verta muodostavien kudosten ja lähisukuisten kudosten pahanlaatuiset kasvaimet	3,241
Pintasyövät	0,162
Keskushermostokalvojen ja aivohermojen hyvänlaatuiset kasvaimet	0,666
Kasvu- ja leviämistaipumukseltaan epäselvät tai tuntemattomat kasvaimet	1,610
Krooniset hankinnaiset ja perinnölliset anemiat, hyytymyshäiriöt, neutropenia	1,486
Immuunipuutokset / immunol. Häiriöt	1,800
Amyloidoosit	2,395
Kilpirauhasen vajaatoiminta	0,046
Diabetes	0,367
Hypertyreoosi	0,555
Struuma	0,399
Hyperparatyreoosi	0,129
Lihavuus	0,438
Päihde- ja riippuvuushäiriöt (pl. Opioidiriippuvuus)	1,427
Opioidiriippuvuus	4,162
Tupakoinnin aiheuttamat haitat	0,071
Psykoosisairaudet ja kaksisuuntainen mielialahäiriö	3,164
Masennus- ja ahdistuneisuushäiriöt	0,474
Dissosiaatio- ja somatisaatiohäiriöt	1,030
Laihuushäiriö	4,045
Syömishäiriöt (pl. Laihuushäiriö)	1,589
Unihäiriöt	0,171
Persoonallisuushäiriöt	0,488

	Painokerroin
Oppimiskyvyn vaikeudet yhdistettynä muihin kuin laaja-alaisiin kehityshäiriöihin	0,493
Laaja-alaiset kehityshäiriöt ("autismispektri")	0,259
Tarkkaavaisuus- ja käyttöhäiriöt	0,921
Sekavuustilat ja elimelliset aivo-oireyhtymät	0,473
Transsukupuolisuus tai muu syntymässä määriteltyä (juridista) sukupuolta vastaamaton sukupuoli-identiteetti	1,327
Määrittämätön mielenterveyden häiriö	1,503
Hengityshalvaus	2,511
Parkinson ja muut rappeuttavat liikehäiriösairaudet	1,137
Epilepsia	0,332
Uniapnea	0,184
Neuroimmunologiset sairaudet	0,653
Keskushermoston ja ääreishermoston tulehdus/tulehdukselliset sairaudet	1,203
Migreeni ja muut päänsärkysairaudet, muu krooninen kipu	0,300
Hydrokefalus	1,567
Neuromuskulaarisairaudet	1,011
Allerginen silmätulehdus + allerginen nuha	0,032
Silmien rappeumataudit	0,632
Glaukooma	0,164
Näkövammaisuus (ne, joilla ei tietoa WHO-luokituksesta)	0,147
Silmien sarveiskalvosairaudet	0,212
Silmien verkkokalvoirtaumat ja verisuonitukokset	0,388
Silmien taittovirheet	0,295
Silmien lasiaissairaudet	0,263
Korvakirurgia	0,386
Huimaus/Korvan tasapainoelimen häiriöt	0,166
Johtumistyyppinen ja sensorineuraalinen kuulonalenema	0,144
Verenpainetauti	0,182
Sepelvaltimotauti	0,347
Eteisvärinä	0,555
Sydämen vajaatoiminta	1,159
Aivoverenkiertohäiriöt	0,660
Ateroskleroosi	2,297
Sydämen läppäsairaudet	0,881
Kardiomyopatiat	0,503
Sydämen johtumishäiriöt	0,411
Keuhkoveritulppa	0,819
Rinta-aortan aneurysmat	0,759
Keuhkokuume	1,081
COPD	0,801
Astma	0,222
Keuhkokudoksen sairaudet	1,199
Hengityselinten krooninen toimintavajaus	2,393
Hammaskaries ja hammasytimen ja hampaanjuuren kärkeä ympäröivien kudosten sairaudet	1,171
Purentaelimen poikkeavuudet	0,802

	Painokerroin
Hampaan kiinnityskudosten sairaudet	1,199
Suun protetiikka	0,428
Refluksi ja ulcus	0,352
Divertikkelit ja ärtyvä suoli	0,353
Sappirakon ja sappiteiden sairaudet	0,630
Maksan tulehdussairaudet ja vajaatoiminta	1,181
Crohnin tauti ja haavainen koliitti	0,630
Keliakia	0,273
Atooppinen ekseema	0,184
Psoriaasi	0,310
Krooniset haavat	3,087
Allerginen kosketusihottuma	0,272
Aktiininen keratoosi	0,101
Nivelreuma	0,593
Artroosisairaudet	0,479
Olkapään vaivat	0,100
Muulla luokittelemattomat muut nivelsairaudet	0,249
Selkärangan sairaudet	0,321
Luukato	0,617
Polven sisäiset viat	0,108
Varpaiden hankinnaiset epämuotoisuudet	0,471
Raajakipu	0,349
Munuaisten vajaatoiminta	4,336
Munuaissairaudet	0,988
Virtsakivet ja virtsaushäiriöt	0,440
Endometrioosi	0,674
Kohdunkaulan dysplasia	0,139
Kuukautisvuotohäiriöt	0,378
Hedelmättömyys	0,815
Miehen sukupuolielinten sairaudet	0,287
Vammat ja myrkytykset	0,297
Lonkkamurtuma	0,779
WHO:n näkövamma luokitus	
1. aste	0,252
2. aste	0,273
3. aste	0,047
4. aste	0,236
5. aste	0,959
9. aste (määrittelemätön)	0,720
Raskauden ennenaikaisuus päivinä	-0,170
Raskauden ennenaikaisuus päivinä²	0,010
Syntytyt	2,840
Työkyvytön, ikäkvantiili	
1. ikäkvantiili	1,612
2. ikäkvantiili	0,276

	Painokerroin
3. ikäkvantiili	-0,192
4. ikäkvantiili	0,146
5. ikäkvantiili	0,309
Pääasiallinen toiminta	
Työllinen	-0,125
Opiskelija	-0,059
Varusmies	-0,215
Koulutus	
Toinen aste	-0,035
Korkeakoulu	-0,052
Asuntokunnan käyttötulo per kulutusyksiköt, luonnollinen logaritmi	-0,058
Siviilisääty	
Naimaton	0,022
Eronnut	0,046
Leski	-0,056
Perhemuoto	
Yhden aikuisen perhe	0,048
Yksinasuja, alle 75-vuotias	-0,064
Yksinasuja, 75–84-vuotias	0,389
Yksinasuja, 85–89-vuotias	0,962
Yksinasuja, yli 89-vuotias	1,797
Taustamaa muu kuin Suomi	-0,107
Matka-aika minuutteina	-0,001
Matka-aika minuutteina²	0,000

Liite 2. Sosiaalihuollon mallin painokertoimet.

	Painokerroin
Nainen	0,008
Ikäryhmä	
Ikä 0	6,965
Ikä 1–6	6,99
Ikä 7–12	7,026
Ikä 13–18	8,802
Ikä 19–25	6,718
Ikä 26–30	6,727
Ikä 31–35	6,882
Ikä 36–40	6,957
Ikä 41–45	6,932
Ikä 46–50	7,003
Ikä 51–55	7,070
Ikä 56–60	7,129
Ikä 61–65	7,093
Ikä 66–70	6,983
Ikä 71–75	7,004
Ikä 76–80	6,811
Ikä 81–85	6,636
Ikä 86–90	6,259
Ikä 91–95	5,740
Yli 95 vuotta	5,256
Sairastavuus	
HIV, C-hepatiitti	1,727
Ruusut	1,027
Päihde- ja riippuvuushäiriöt (pl. Opioidiriippuvuus)	3,806
Opioidiriippuvuus	6,112
Psykoosisairaudet ja kaksisuuntainen mielialahäiriö	2,644
Masennus- ja ahdistuneisuushäiriöt	0,085
Älyllinen kehitysvammaisuus	43,019
Laaja-alaiset kehityshäiriöt ("autismispektri")	7,786
Tarkkaavaisuus- ja käyttöhäiriöt	11,34
Muistisairaudet ja Alzheimerin tauti	0,373
Hengityshalvaus	20,463
Parkinson ja muut rappeuttavat liikehäiriösairaudet	1,485
Epilepsia	3,122
CP-oireyhtymä	18,437
Neuroimmunologiset sairaudet	5,450
Keskushermoston ja ääreishermoston tulehdus/tulehdukselliset sairaudet	1,113
Hydrokefalus	3,039
Neuromuskulaarisairaudet	3,944
Aivoverenkiertohäiriöt	1,083
Keuhkokuume	0,976
COPD	0,285

	Painokerroin
Hengityselinten krooninen toimintavajaus	6,277
Hammaskaries ja hammasytimen ja hampaanjuuren kärkeä ympäröivien kudosten sairaudet	1,027
Hampaan kiinnityskudosten sairaudet	0,967
Krooniset haavat	1,779
Nivelreuma	0,202
Raajakipu	0,247
Munuaissairaudet	1,032
Hedelmättömyys	0,293
Vammat ja myrkytykset	0,652
WHO:n näkövammaluokitus	
1. aste	2,836
2. aste	5,642
3. aste	7,000
4. aste	9,281
5. aste	22,782
9. aste (määrittelemätön)	2,396
Pääasiallinen toiminta	
Työllinen	-0,018
Opiskelija	-0,511
Varusmies	0,045
Koulutus	
Toinen aste	-0,460
Korkeakoulu	-0,208
Työkyvytön, ikäkvantiili	
1. ikäkvantiili	4,748
2. ikäkvantiili	3,26
3. ikäkvantiili	1,492
4. ikäkvantiili	0,987
5. ikäkvantiili	1,154
Asuntokunnan käyttötulo per kulutusyksiköt, luonnollinen logaritmi	-0,667
Siviilisäätty	
Naimaton	0,236
Eronnut	-0,114
Leski	-0,178
Yhden aikuisen perhe	0,429
Taustamaa muu kuin Suomi	-0,755

Liite 3. Vanhustenhuollon mallin painokertoimet.

	Kerroin
Sukupuoli, ikäryhmä	
Nainen, ikä 65–70	2,311
Nainen, ikä 71–75	2,359
Nainen, ikä 76–80	2,525
Nainen, ikä 81–85	3,056
Nainen, ikä 86–90	4,025
Nainen, ikä 91–95	5,062
Nainen, ikä yli 96 vuotta	5,737
Ikäryhmä, mies	
Mies, ikä 65–70	2,328
Mies, ikä 71–75	2,367
Mies, ikä 76–80	2,407
Mies, ikä 81–85	2,672
Mies, ikä 86–90	3,327
Mies, ikä 91–95	4,151
Mies, ikä yli 96 vuotta	4,488
Sairastavuus	
Tuberkuloosi	0,678
Ruusut	0,289
Krooniset hankinnaiset ja perinnölliset anemiat, hyytymyshäiriöt, neutropenia	0,115
Diabetes	0,201
Lihavuus	0,440
Päihde- ja riippuvuushäiriöt (pl. Opioidiriippuvuus)	0,846
Psykoosisairaudet ja kaksisuuntainen mielialahäiriö	1,520
Masennus- ja ahdistuneisuushäiriöt	0,725
Dissosiaatio- ja somatisaatiohäiriöt	0,130
Sekavuustilat ja elimelliset aivo-oireyhtymät	0,754
Muistisairaudet ja Alzheimerin tauti	3,941
Hengityshalvaus	2,365
Parkinson ja muut rappeuttavat liikehäiriösairaudet	1,034
Epilepsia	0,636
CP-oireyhtymä	0,855
Neuroimmunologiset sairaudet	1,889
Keskushermoston ja ääreishermoston tulehdus/tulehdukselliset sairaudet	0,214
Hydrokefalus	1,598
Neuromuskulaarisairaudet	1,116
Eteisvärinä	0,033
Sydämen vajaatoiminta	0,542
Aivoverenkiertohäiriöt	1,433
Ateroskleroosi	0,189
Keuhkoveritulppa	0,308
Keuhkokuume	0,448
Hengityselinten krooninen toimintavajaus	0,350
Hammaskaries ja hammasytimen ja hampaanjuuren kärkeä ympäröivien kudosten sairaudet	0,119

	Kerroin
Refluksi ja ulcus	0,117
Krooniset haavat	0,977
Nivelreuma	0,136
Artroosisairaudet	0,124
Luukato	0,734
Munuaissairaudet	1,011
Virtsakivet ja virtsaushäiriöt	0,136
Vammat ja myrkytykset	0,357
Lonkkamurtuma	1,158
WHO:n näkövammaluokitus	
1. aste	0,55
2. aste	1,058
3. aste	1,003
4. aste	1,663
5. aste	2,277
9. aste (määrittelemätön)	0,209
Koulutus	
Toinen aste	-0,057
Korkeakoulu	-0,03
Asuntokunnan käyttötulo per kulutusyksiköt, luonnollinen logaritmi	-0,257
Siviilisäät	
Naimaton	0,462
Eronnut	0,21
Leski	0,359
Taustamaa muu kuin Suomi	-0,266

Lähteet

- ACRA (2015). Cost of unavoidable smallness due to remoteness. Advisory Committee on Resource Allocation. 2015:36.
- Anselmi, L., Everton, A., Shaw, R., Suzuki, W., Burrows, J., Weir, R., Tatarek-Gintowt, R., Sutton, M. & Lorrimer S. Estimating local need for mental healthcare to inform fair resource allocation in the NHS in England: cross-sectional analysis of national administrative data linked at person level. *The British Journal of Psychiatry* 2019:216; 338-344.
- Barrow, M. M. (1986). Central grants to local governments: a game theoretic approach. *Environment and planning C: government and policy*, 4(2), 155-164.
- Buchner, F., Wasem, J., & Schillo, S. (2017). Regression trees identify relevant interactions: can this improve the predictive performance of risk adjustment? *Health Economics*, 26(1), 74-85.
- Chaplin, M., Beatson, S., Yiu-Shing, L., Davies, C., Smyth, C., Burrows, J., Weir, R. & Tatarek-Gintowt, R. 2016. Refreshing the Formulae for CCG Allocations. For allocations to Clinical Commissioning Groups from 2016-2017. Report on the methods and modelling. ANHS England, Analytical Services (Finance).
- Dudley, R. A., Medlin, C. A., Hammann, L. B., Cisternas, M. G., Brand, R., Rennie, D. J., & Luft, H. S. (2003). The best of both worlds? Potential of hybrid prospective/concurrent risk adjustment. *Medical Care*, 56-69.
- Ellis, R. P., Martins, B. & Rose, S. (2017). Risk Adjustment for Health Plan Payment. In McGuire, T.G. & van Kleef, R. (eds.) *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*. 1-45. Elsevier Press.
- Haula, T., Peltola, M. (2020). Monialainen tai kallias palvelujen käyttö Suomessa. Teoksessa: Koivisto, J., Tiirinki, H. (toim.) Monialaisen palvelutarpeen tunnistaminen sosiaali-, terveys- ja työvoimapalveluissa. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2020:38.
- Holster, T., Nguyen, L. & Häkkinen, U. (2022). The role of occupational healthcare in ambulatory healthcare in Finland. *Nordic Journal of Health Economics*, tulossa.
- Holster, T., Ji, S. & Marttinen, P. (2022). Risk Adjustment for Regional Health Care Funding Allocations with Ensemble Methods: A Empirical Study and Interpretation. Julkaisematon työpaperi.
- Häkkinen U, Holster T, Haula T, Kapiainen S, Kokko P, Korajoki M, Mäklin S, Nguyen L, Puroharju T, Peltola M. (2020). Sote-rahoituksen tarvevakiointi. Terveiden ja hyvinvoinnin laitoksen raportti 6/2020.
- Häkkinen, U., Laukkonen, M-L., Nguyen, L., Holster, T. & Kapiainen, S. (2022). Olosuhdetekijät sosiaali- ja terveydenhuollon rahoituksessa. Terveiden ja hyvinvoinnin laitos (THL). Työpaperi 25/2022.
- Juhnke, C., Bethge, S. & Muhlbacher, A.C. (2016). A Review on Methods of Risk Adjustment and their Use in Integrated Healthcare Systems. *International Journal of Integrated Care* 16(4):4; 1-18.
- Kestilä, L., Kapiainen, S., Mesiäislehto, M. & Rissanen, P. (toim.) (2020). Covid-19-epidemian vaikutukset hyvinvointiin, palvelujärjestelmään ja kansantalouteen - Asiantuntija-arvio, kevät 2022. Terveiden ja hyvinvoinnin laitoksen raportti 4/2022.
- Kestilä, L., Härmä, V. & Rissanen, P. (toim.) (2020). Covid-19-epidemian vaikutukset hyvinvointiin, palvelujärjestelmään ja kansantalouteen - Asiantuntija-arvio, kevät 2021. Terveiden ja hyvinvoinnin laitoksen raportti 3/2021.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. (Vol. 26, p. 13). New York: Springer.
- Kuhn, M., & Johnson, K. (2020). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.
- Kuusio H, Seppänen, A, Jokela S, Somersalo L & Lilja E (2020). Ulkomaalaistaustaisten terveys ja hyvinvointi Suomessa. FinMonik-tutkimus. Raportti 1. Helsinki: Terveiden ja hyvinvoinnin laitos
- Laatikainen, T., Koponen, P., Reinikainen, J., Tolonen, H., Jousilahti, P., Suvisaari, J., Mattila, T., Niiranen, T. & Koskinen, S. (2020). Kansanterveyden seuranta, arviointi ja ennakointi: Mitä tietoa Suomessa saadaan hoitoilmoitusrekisteristä ja mitä väestötutkimuksista? *Suomen Lääkärilehti*, 2020:37(75):1853-1858a.
- Lyytikäinen, T., Saxell, T., Siikanen, M. & Toikka, M. Alueelliset palkkaerot sosiaali- ja terveydenhuollossa. VATT Tutkimukset 192/2022.
- Malmi, T., Riipinen, T., Roos, J. & Tukiainen, J. (2021). Arvio sote-uudistuksesta ja sote-rahoituksen tarvevakioinnista. Tempo Leon.
- Mäklin S & Kokko P. Terveiden- ja sosiaalihuollon yksikkökustannukset Suomessa vuonna 2017. Terveiden ja hyvinvoinnin laitoksen Työpaperi 21/2020.
- Pope, G. C., Kautter, J., Ellis, R. P., Ash, A. S., Ayanian, J. Z., Iezzoni, L. I., J., I. M., Levy, J. M. and Robst, J. (2004). Risk adjustment of medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, 25(4):119-141.
- Rissanen, P., Parhiala, K., Kestilä, L., Härmä, V., Honkatukia, J. & Jormanainen, V. (2020). Covid-19 epidemian vaikutukset väestön palvelutarpeisiin, palvelujärjestelmään ja kansantalouteen -nopea vaikutusarvio. Terveiden ja hyvinvoinnin laitoksen raportti 8/2020.
- Rose, S. (2016). A machine learning framework for plan payment risk adjustment. *Health Services Research*, 51(6):2358-2374.
- Shokkaert, E., & Van de Voorde, C. (2004). Risk selection and the specification of the conventional risk adjustment formula. *Journal of Health Economics*, 23(6), 1237-1259.

- Shmueli G. To Explain or to Predict? *Statistical Science* 2010;25(3):289-310.
- Shrestha, A., Bergquist, S., Montz, E., and Rose, S. (2018). Mental health risk adjustment with clinical categories and machine learning. *Health Services Research*, 53(4):3189–3206.
- Smith, P. C. (2003). Formula funding of public services: an economic analysis. *Oxford Review of Economic Policy*, 19(2), 301-322.
- Smith, P. C. (2007). *Formula funding of public services*. Taylor & Francis.
- Smith, P. C. (2008). Resource allocation and purchasing in the health sector: the English experience. *Bulletin of the World Health Organization*, 86, 884-888.
- Sund, R. Quality of the Finnish Hospital Discharge Register: a systematic review. *Scand J Public Health*. 2012; 40:505–15.
- Van Kleef, R.C., Eijkenaar, F., van Vliet, R.C.J.A. & de Ven, W.P.M.M. (2018). Health plan payment in the Netherlands. Teoksessa: McGuire, T.G. & van Kleef, R. (toim.). *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets*. Academic Press, 2018:397-429.
- Van Kleef, R.C. & Reuser, M. (2021). How the COVID-19 pandemic can distort risk adjustment of health plan payment. *The European Journal of Health Economics*. 22:1005-1016.
- Van Veen, S. H. C. M., Van Kleef, R. C., Van De Ven, W. P. M. M., & Van Vliet, R. C. J. A. (2018). Exploring the predictive power of interaction terms in a sophisticated risk equalization model using regression trees. *Health Economics*, 27(2), e1-e12.