

# Spatiaalista mallinnusta pistedatalla: kyselypohjainen analyysi koetusta terveydestä ja turvattomuudesta Helsingin metropolialueella

Tässä menetelmäartikkelissa esitellään spatiaalisten mallien soveltamista sosiaalitieteellisessä kyselytutkimuksessa. Havainnollistavina esimerkkeinä käytetään koettua terveyttä ja koettua turvattomuutta Helsingin metropolialueella. Spatiaalinen analyysi tarjoaa joustavia työkaluja paitsi analyysin eksploraatiiviseen vaiheeseen, myös mallinnukseen, jossa havaintojen spatiaalinen korrelaatio voidaan ottaa huomioon paikkatietopohjaisen naapuruusmäärittelyn avulla. Artikkelissa käydään läpi spatiaalisten mallien soveltamisen kannalta keskeistä käsitteistöä ei-tekniisellä ja käytännöllisellä otteella. Havainnollistavassa esimerkissä verrataan tavanomaisten regressiomallien tuloksia aluemallien vastaaviin. Artikkelissa todetaan, että koetulla turvattomuudella on selvästi enemmän alueellista rakennetta kuin terveydellä. Turvattomuus on siinä määrin spatiaalisesti korreloitunut ilmiö, että spatiaalisten mallien soveltaminen on perusteltua, minkä seurauksena selittävien kontekstimuuttujien piste-estimaatit pienenevät tuntuvasti tavanomaisten mallien tuottamiin estimaatteihin nähden.

## TEEMU KEMPPAINEN

### JOHDANTO

Sosiaalitieteissä on perinteisesti oltu kiinnostuneita paitsi yksilöistä myös heidän ympäristönsä tai kontekstinsa merkityksestä kulloisenkin tutkittavan ilmiön kannalta. Viime aikoina suosiota saanut monitasoanalyysi pyrkii tarttumaan juuri tämänkaltaisiin tutkimusasetelmiin, joissa havainnot ovat joidenkin luonnollisten ryväsraakenteiden suhteen korreloituneita eli riippuvia (Sampson ja Raudenbush 1997, Karvonen ym. 2001, Kauppinen 2004, Blomgren 2005; vrt. Chaix ym 2005). Ryvästyneissä asetelmissä havainnon sijainti määrättyssä kontekstirakenteessa on analyysien keskeinen lähtökohta. Tässä artikkelissa esitellään toinen, eräiltä osin joustavampi tapa hyödyntää sijaintitietoa lähestymällä aihetta spatiaalisten ekonometristen mallien (*spatial econometric models*) näkökulmasta (Ward ja Gleditsch 2008, Bivand ym. 2008, Anselin 2007 ja 2009).

Viimeaikainen menetelmien, ohjelmistojen ja laskentatehon kehitys on avannut kontekstuaalisesti suuntautuneelle sosiaalitutkimukselle uusia mahdollisuuksia. Tässä artikkelissa esiteltävä spatiaalisen ekonometrian lähestymistapa sijaintitietoon tarjoaa lupaavia sisällöllisiä ja teknisiä ratkaisuja tutkimuksen eri vaiheisiin. Sen avulla on mahdollista päästä monipuolisesti käsiksi tutkittavan ilmiön alueelliseen jakautumiseen niin visuaalisen kuvaamisen kuin konfirmatoriseman testausotteenkin mielessä. Alueellisen jakautumisen pohjalta voidaan määrittää tutkittavan ilmiön kannalta sopivimmalta vaikuttava mittakaava regressiopohjaisiin analyyseihin. Ero monitasoanalyysien kankeisiin, hallinnollisiin yksiköihin pohjautuviin aluejaotteluihin nähden on selkeä. Teknisessä mielessä sijainti on regressiomallien diagnostiikan ja kehittämisen kannalta keskeinen tekijä, sillä sen avulla voidaan tarkastella havaintojen spatiaalista korrelaatiota, joka rikkoisi oletusta havaintojen riippumattomuudes-

ta. Spatiaalisen korrelaation huomioiminen alue-  
malleissa on askel kohti parempia regressioker-  
toimia. (Chauvin 2005, Ward ja Gleditsch 2008,  
LeSage ja Pace 2010, Virrantaus 2011.)

Aluetematiikka on ollut viimeaikaisessa suo-  
malaisessa yhteiskuntatieteellisessä tutkimuksessa  
esillä mm. tutkittaessa alueiden erilaistumista  
(Vaattovaara 1998, Lankinen 2001, Rasinkangas  
2013), asumistoiveita (Kortteinen ym. 2005), päi-  
vittäisen matkustamisen ekologisuutta (Salonen  
ym. 2014), huono-osaisuutta (Blomgren 2005),  
elämäntyyliä (Ratvio 2012), fyysistä toimintaky-  
kyä ja psyykkistä hyvinvointia (Sipilä ja Marti-  
kainen 2005), koulutusta (Kauppinen 2004, Ber-  
nelius 2013), kuntien verokilpailua (Lyytikäinen  
2012), mielenterveyttä (Aaljoki ym 2000, Lehto-  
nen ja Kauronen 2013), muuttoliikettä (Lehto-  
nen ja Tykkyläinen 2009; Vilkama 2011), nuor-  
ten elinoloja (Paju 2004) ja tupakointia (Karvo-  
nen ym. 2001), tulotason kehitystä (Lehtonen ja  
Tykkyläinen 2010), turvattomuuden kokemusta  
(Kääriäinen 2002, Kemppainen ym. 2014) ja vä-  
kivaltaa (Kortteinen ym. 2001, Herttua ym.  
2008). Näiden lisäksi on erikseen syytä mainita  
Heikki Loikkasen ja Juho Saaren toimittama teos  
”Suomalaisen sosiaalipolitiikan alueellinen ra-  
kenne” (2000), joka sisältää kymmenkunta tutki-  
musta, joissa tarkastellaan mm. tuloeroja, työttö-  
myyttä, terveyspalvelujen käyttöä ja toimeentulo-  
tukea alueellisesta näkökulmasta. Kuitenkin vain  
harvoin spatiaalisuus on otettu huomioon tekni-  
sesti riittävällä tavalla, mikä perustelee käsillä  
olevan tekstin tarpeen.

Paikkatietoaineiston eri lajeista keskitytään  
tässä artikkelissa pistedataan eli aineistoon, jossa  
havainnon sijainti tiedetään koordinaatteina.  
Tässä tutkimuksessa paikkatietokoodattua kyse-  
lydataa käytetään spatiaalisten ekonometristen  
mallien havainnollistamiseen; sen sijaan nk. spa-  
tiaalisia pisteprosesseja ei käsitellä (esim. Heikki-  
nen 2007). Artikkelin tarkoituksena on esitellä  
aidon empiirisen aineiston avulla (vrt. Merlo ym.  
2005), kuinka pistedatan sijaintitietoa voidaan  
hyödyntää sosiaalitieteellisessä kyselytutkimuk-  
sessa. Havainnollistavina esimerkkeinä käytetään  
koettua terveyttä (*self-rated health*, SRH) ja koet-  
tua turvattomuutta. Esimerkit on valittu seuraavista  
systä. Niitä voidaan ensinnäkin pitää poh-  
joismaisen hyvinvointitutkimuksen ajattelun mu-  
kaisesti hyvinvoinnin tai elämänlaadun osatekijä-  
nä tai ulottuvuutena. Toiseksi, ilmiöillä on varsin  
erilainen rakenne spatiaalisuuden suhteen, mikä  
havainnollistaa artikkelissa esiteltävän menetel-

män käyttöaluetta ja etuja. Kolmanneksi, ilmiöt  
sijoittuvat tutkimusaiheina ekologisen sosiaalitut-  
kimuksen kummankin historiallisen pääjuonteen,  
modernin epidemiologian ja kriminologian  
alueille.

Esimerkkien osalta tutkimuskysymyksenä on  
selvittää, eroavatko Helsingin seudun lähiöt  
muusta alueesta koetun terveyden ja turvatto-  
muuden suhteen. Mahdollisia eroja pyritään se-  
littämään sosiaalisen disorganisaation teoriaper-  
innettä mukaillen erityisesti asuin ympäristön  
sosio-ekonomisen luonteen avulla. Aikaisemmin  
ei ole päästy käsiksi suomalaisten lähiöasukkai-  
den koettuun hyvinvointiin otteella, jossa yhdis-  
tyy lähiön rekisteripohjainen määrittely ja sovel-  
tuva kyselyaineisto. Koska kyseessä on ennen  
muuta menetelmäartikkeli, on tuloksiin syytä  
suhtautua alustavina luonnoksina, joiden päätehtävä  
on konkretisoida käytettyä menetelmällistä  
lähestymistapaa.

Kappaleessa 2 esitellään spatiaalisten mallien  
soveltamisen kannalta keskeistä käsitteistöä, min-  
kä jälkeen kuvaillaan käytetyn empiirisen aineis-  
ton rakenne ja tutkimusasetelma (kappale 3).  
Varsinaiseen analyysivaiheeseen päästään kapp-  
leessa 4, jossa keskitytään tavanomaisen ja spa-  
tiaalisen mallinnuksen tekniseen vertailuun. Lo-  
puksi kootaan yhteen artikkelin keskeinen sisältö  
ja avataan keskeisimpiä pohdintaa vaativia kysy-  
myksiä (kappale 5).

## SPATIAALISTEN MALLIEN KÄSITTEELLISTÄ TAUSTAA

Spatiaalista mallintamista voidaan pitää nykyai-  
kaisena työkaluna, joka auttaa tarttumaan eri  
aloilla jo pitkään askarruttaneisiin kysymyksiin  
ympäristön, sijainnin tai yleisemmin kontekstu-  
aalisuuden merkityksestä. Eri tieteenaloilla tema-  
tiikkaa on lähestytty keskenään varsin saman-  
tyyppisillä käsitteillä. Epidemiologisessa tutki-  
muksessa aihetta käsitellään kontekstuaalisen  
vaikutuksen (*contextual effect*) käsitteellä, kau-  
punktutkimuksessa puhutaan naapurustovaiku-  
tuksesta (*neighbourhood effect*) ja spatiaalisen  
ekonometrian teksteissä käytetään spatiaalisen  
vaikutuksen (*spatial effect*) käsitettä (Anselin  
1988, Morgenstern 2008, Sampson 2012). Jois-  
sakin tapauksissa käsitteet erotellaan selväsanai-  
sesti toisistaan, kuten tekevät esimerkiksi Dujar-  
din ym. (2009), jotka erottelevat naapurustovai-  
kutuksen spatiaalisesta vaikutuksesta siten, että  
edellisessä on kyse kuulumisesta maantieteellises-  
ti määräytyneeseen sosiaaliseen ryhmään, jälkim-  
mäisessä taas puhtaasti sijaintiin liittyvästä vaiku-

tuksesta. Tässä yhteydessä ei ole mahdollisuutta eritellä tarkemmin näiden kontekstuaalisuuteen liittyvien käsitteiden erilaisia määritelmiä ja käsitteiden keskinäisiä yhtäläisyyksiä ja eroja.

### SPATIAALINEN KORRELAATIO

Spatiaaliset vaikutukset jaetaan tavallisesti kahden luokkaan, spatiaaliseen riippuvuuteen (*spatial dependence*) ja spatiaaliseen heterogeenisyyteen (*spatial heterogeneity*). Spatiaalisen riippuvuuden tapauksessa havainnoilla on sijainnin mukaan jäsentynyttä (auto)korrelaatiota (*spatial correlation*): jos esimerkiksi koetussa turvattuudessa esiintyy positiivista spatiaalista korrelaatiota, muistuttavat lähekkäin asuvien vastaajien turvattomuuskokemukset toisiaan. Spatiaalinen heterogeenisuus puolestaan viittaa havaittujen yhteyksien alueelliseen vaihteluun, joka muistuttaa ajatuksena interaktiotermejä tavanomaisessa regressiossa tai vaihtelevia kertoimia sekä eli monitasomalleissa. Tässä artikkelissa keskitytään spatiaaliseen riippuvuuteen, mikä on tavanomainen lähestymistapa myös soveltavassa kirjallisuudessa. (Manski 1993, Ward ja Gleditsch 2008, Anselin 2009, Elhorst 2010, Fornango 2012.) Spatiaalisesta heterogeenisuudesta ja sen yhteydessä käytettävästä maantieteellisesti painotetusta regressiosta (*geographically weighted regression, GWR*) kiinnostuneille on tarjolla paitsi kattavia esityksiä englanninkielisissä tutkimuskirjallisuudessa (Fotheringham 1998 ja 2009, Wheeler ja Paez 2010) myös joitakin suomalaisia esimerkkejä (Sund ja Nouko-Juvonen 2000, Lehtonen ja Kauronen 2013).

Spatiaalisen korrelaation ajatus liittyy läheisesti ns. ensimmäiseen maantieteen lakiin, jonka mukaan lähekkäiseen sijaintiin liittyy tavallisesti muutakin samankaltaisuutta (Tobler 1970). Kyseessä on ajatus, jonka juuret voidaan jäljittää tilastotieteen varhaisten pioneerien töihin (Getis 2010). Tilastotieteestä tutut sisäkorrelaatio ja aikasarjojen autokorrelaatio ovat käsitteellisesti sukua spatiaaliselle korrelaatiolle. Spatiaalisen korrelaation tapauksessa samanlaisuus perustuu lähekkäisyyteen, jonka ei kuitenkaan välttämättä tarvitse olla maantieteellistä, sillä ajatus voidaan yleistää koskemaan vaikkapa sosiaalista lähekkäisyyttä. Euklidisen etäisyyden sijaan voidaan myös tarkastella esimerkiksi ajallista etäisyyttä maantieteellisten sijaintien välillä.

Klassinen tapa määrittää alue-erojen suuruusluokkaa on laskea ns. globaali Moranin I, joka kertoo ilmiön spatiaalisesta korrelaatiosta

koko tarkastelualueella. Muitakin spatiaalisen korrelaation mittoja on (gamma, Gearyn c ja paikalliset mittarit: *LISA, local indicators of spatial association*), mutta tässä tutkimuksessa keskitytään globaaliin Moranin indeksiin, joka käytännössä on spatiaalisen korrelaation ”johtava testi ja mittari” (Getis 2010, 262). Periaatteessa ilmiö voi olla alueen suhteen satunnaisesti jakautunut, hajaantunut (negatiivinen autokorrelaatio) tai klusteroitunut (positiivinen autokorrelaatio).

Valitun muuttujan alueellista jakautumista kuvaava globaali Moranin I lasketaan koko alueen havainnoista ja se kuvaa alueella keskimäärin vallitsevaa havainnon arvon ja sen naapuruston arvojen painotetun keskiarvon välistä korrelaatiota. Painotus perustuu havaintojen naapurussuhteeseen, mitä esitellään tarkemmin seuraavassa kappaleessa. Kyse on siis muuttujan korreloimisesta itsensä kanssa spatiaalisesti eli auto-korrelaatiosta tilassa. I vaihtelee tietyn oletuksen välillä -1 (täydellinen negatiivinen autokorrelaatio) – 1 (täydellinen positiivinen autokorrelaatio). Tilastollista testaamista varten tarvitaan tieto, että I:n odotusarvo on  $-1 / (n-1)$ , missä n on havaintojen määrä. Lisäksi tarvitsemme oletuksen suureen varianssista, missä on käytetty kahta vaihtoehtoa, normaalisuusoletusta ja Monte Carlo -satunnaistamista. Normaalisuusoletusta on kritisoitu pitämättömyydestä, mutta se on käytännössä yleisimmin käytetty valinta. Pääsääntöisesti kumpikin tapa tuottaa samansuuntaiset tulokset. (Ward ja Gleditsch 2008, Getis 2010, Virrantaus 2011, Fornango 2012.)

### NAAPURUSTO

Tässä artikkelissa esiteltävät analyysityökalut edellyttävät, että kullekin aineiston havainnolle määritetään naapurusto. Aineistotyyppi asettaa rajoja sille, miten yksittäisen havainnon naapurit voidaan määrittellä. Aineiston havaintoyksiköt voivat olla esimerkiksi epäsäännöllisen muotoisia, rajojen erottamia alueita; esimerkiksi Suomen kunnat muodostavat tämäntyyppisen aineiston (esim. Sund ja Nouko-Juvonen 2000). Alueet voivat olla myös säännöllisen muotoisia, jolloin on käytännössä kyse abstrakteista alueista, joita ei sellaisenaan ole kokemusmaailmassa olemassa. Ruudukko tai kuusikulmaisten kennojen muodostama aluejako ovat esimerkkejä tällaisesta aluedatasta. Tilastokeskuksen ylläpitämä ruututietokanta (Tilastokeskus 2013) jakaa Suomen nimensä mukaisesti ruutuihin (tietokannasta lisää kappaleessa 3). Alueiden lisäksi aineisto voi olla

pistedataa, jossa jokaisella havainnolla on yksikäsitteinen sijainti tilassa, tässä tapauksessa kaksiolotteiseksi pelkistetyssä Helsingin seudun ta-  
sossa. (Dubin 2009.)

Rajojen erottamat alueet voidaan määrittellä naapureiksi esimerkiksi niin, että yhteinen raja tekee alueista naapureita. Tätä vaihtoehtoa ei ole tarjolla pistedatassa, koska erottavia rajoja ei ole. Sen sijaan voidaan valita tietty määrä lähimpiä naapureita etäisyydestä riippumatta, jolloin jokaisella havainnolla on sama määrä naapureita. Tällöin naapurustojen maantieteellinen koko vaihtelee havaintojen tiheyden mukaan. Naapuruussuhteen mahdollinen epäsymmetria käy tästä tapauksesta ilmi: a voi kuulua b:n k lähimmän naapurin joukkoon ilman, että b välttämättä kuuluisi a:n vastaavaan joukkoon.

Toinen tapa on määrittää tietty etäisyys ja lukea tietyn havainnot naapureiksi kaikki korkeintaan tällä etäisyydellä olevat pisteet, jolloin puolestaan naapureiden määrä vaihtelee pisteestä toiseen. Etäisyyteen ja lukumäärään perustuvat naapuriratkaisut sopivat myös rajojen erottamien alueiden aineistolle, missä tapauksessa on määriteltävä kullekin alueelle piste, jonka suhteen etäisyydet määritellään. Naapureita voidaan kohdella kategorisesti, jolloin havainto joko on tai ei ole määrätyn toisen havainnon naapuri. Toinen vaihtoehto on ottaa käyttöön jonkinlainen kasvavan etäisyyden mukana vähenevä funktio. (Bivand ym. 2008; Ward ja Gleditsch 2008; Dubin 2009: ks. myös Vauramo ym. 1992) Käytännössä naapuruus tuodaan analyysiin matriisimuodossa (ns. W-matriisi), jossa kategorisen määrittelyn tapauksessa ykkönen ilmaisee naapuriparin (kai-  
killa naapureilla on sama painoarvo) ja muualla, erityisesti diagonaalilla, on nollaa. Ei-kategorisessa tapauksessa naapuriparin välinen etäisyys heijastuu käänteisesti painomuuttujan suuruuteen, jolloin ainoastaan diagonaalilla on nollaa (yksi-  
kään havainto ei ole itsensä naapuri).

Tässä artikkelissa on yksinkertaisuuden ja vastaajatiheyden suuren aluevaihtelun vuoksi valittu tarjolla olevista lähestymistavoista kategori-  
nen kiinteän naapurimäärän ratkaisu. Lisäksi on pidetty kiinni ns. isotropiaoletuksesta, jonka mukaan ainoastaan havaintojen välinen etäisyys ratkaisee, eikä suunnalla ole väliä (Fortin ja Dale 2009). Tämä on usein todellisuuteen nähden liiallisen yksinkertaistava oletus, mutta se on tapana tehdä tässä esiteltävien mallien sovelluksissa. Eräs tapa tämän korjaamiseen olisi käyttää etäisyyksittain esimerkiksi joukkoliikenteen aika-

taulujen (Toivonen 2014) pohjalta laskettua aikaetäisyyttä, mutta tämä ratkaisu ei olisi tyydyttävä ympäryskuntien tapauksessa, sillä julkinen liikenne ei siellä samalla tavalla kerro asuin-  
ympäristöjen välimatkoista kuin vaikkapa Helsingissä.

Naapurien määrittely edellä mainituilla tavoilla sisältää tutkijan harkinnasta riippuvan subjektiivisen tekijän. Naapurirakenne voidaan myös yrittää estimoida aineistosta, mutta tässä artikkelissa pyritään havainnollistamaan spatiaalisia malleja yksinkertaisimmalla mahdollisella tavalla ja pitäytytään siksi eksogeenisessä määrittelyssä, mikä on käytännössä tyyppinen valinta. (Dubin 2009.) Naapurusto- tai kontekstirakenne voi vai-  
kuttaa saataviin tuloksiin huomattavastikin (vrt. MAUP, *modifiable areal unit problem*, esim. Fischer 2001, Montello 2001), minkä vuoksi rakenteen määrittämiseen ja eri vaihtoehtojen vahvuuksiin, heikkouksiin ja seurauksiin kiinnitetään artikkelin empiirisissä esimerkeissä huomiota. Kiinteän naapurimäärän tapauksissa tämä tarkoittaa erityisesti naapuruston optimaalisen koon etsimistä ja eri kokojen vertailua.

#### SPATIAALINEN VIRHE- JA VIIVEMALLI

Spatiaalisten mallien keskeinen anti soveltajalle on siinä, että ne tarjoavat keinon tarttua spatiaaliseen korrelaatioon liittyviin haasteisiin. Niiden teoreettista puolta esitellään seuraavassa tarkoituksellisen ei-teknisesti ja eräänlaisessa ”sovel-  
tajan soveltajalle” -hengessä Juan Merlon tutkimusryhmän julkaisemien monitasoartikkelien tyyliin (esim. Merlo ym. 2005). Esityksen matemaattisesta muodollisuudesta on siksi tingitty ja sen sijaan keskitytään mallien peruserätyyden esittämiseen käsitteellisellä ja käytännöllisellä tasolla. Esimerkiksi Anselin (2009), Elhorst (2010) ja Gibbons ja Overman (2012) tarjoavat lisätietoja tarkkaa matemaattista ilmaisua kaipaaville.

Spatiaalisten mallien logiikkaa voidaan havainnollistaa tarkastelemalla selitettävän muuttujan alueellista jakaumaa. Seuraan esityksessäni Fornangoa (2012; ks. myös Ward ja Gleditsch 2008 ja Sampson 2012). Jos selitettävällä muuttujalla on spatiaalista korrelaatiota, voidaan ensiksi tarkastella, johtuuko se selittävien tekijöiden alueellisesta rakenteesta. Mikäli selittävien tekijöiden tuominen malliin poistaa spatiaalisen korrelaation mallin jäännöstermistä, on alkuperäisellä selitettävän muuttujan spatiaalisella korrelaatiolla rakenteellinen selitys. Jos esimerkiksi väestön ikäjakaumassa on alueellista rakennetta, se saattaisi osaltaan selittää erityyppisten terveysin-

dikaattoreiden lähtökohtaista spatiaalista korrelaatiota.

Jos sen sijaan lopullisen mallin jäännöstermiin jää spatiaalista korrelaatiota, on periaatteessa kolme vaihtoehtoa. Ensiksi, kyseessä voi olla selitettävän muuttujan kuvaaman ilmiön vuorovaikutteinen luonne itsensä kanssa. Ilmiössä on tällöin tarttumista (*contagion*) tai diffuusiota, ja siihen viitataan spatiaalisen viiveen (*spatial lag*) käsitteellä. Vastaava malli, spatiaalinen viivemalli, sisältää selitettävän muuttujan spatiaalisen viiveen – tässä tapauksessa naapuruston keskiarvon – selittävänä tekijänä. Toinen vaihtoehto on rakenteellinen selitys sellaisten selittävien muuttujien suhteen, joita ei ole mallissa tai joita ei ole lainkaan mitattu. Tähän viitataan termillä spatiaalinen häiriö (*spatial disturbance*) tai virhe (*spatial error*), ja sitä vastaa käytännössä spatiaalinen virhemalli, jossa jäännöstermiin on lisätty tavallisen satunnaisen komponentin rinnalle residuaalien spatiaalinen viive. Spatiaalisen viive- ja virhemallin erikoiskertoimet ( $\rho$  ja  $\lambda$ ) tulkitaan kuten muutkin regressiokertoimet: ne kertovat, missä määrin naapuruston keskiarvo (vastaavasti)  $y$ -muuttujassa tai virhetermissä kontribuoi kullekin havainnolle mallin perusteella laskettavaan ennustearvoon. Kolmantena vaihtoehtona on jo mainittu spatiaalinen heterogeenisuus, jonka vallitessa regressiokertoimet eli selittävien ja selitettävän muuttujan yhteydet vaihtelevat alueellisesti.

Spatiaalisen riippuvuuden huomiotta jättämisellä voi olla vakavia seurauksia tuloksien kannalta, sillä tavanomaisin menetelmin lasketut regressiokertoimien piste-estimaatit saattavat olla virheellisiä. Tarkemmin määritettynä kyse on ongelmista estimaattien tarkentuvuuden ja mahdollisesti myös harhaisuuden ja tehokkuuden suhteen, spatiaalisen prosessin luonteesta riippuen. Myös tavanomaiset keskivirhe-estimaatit voivat olla virheellisiä. (Ward ja Gleditsch 2008, LeSage ja Pace 2010.)

## AINEISTO JA ASETELMA

### AINEISTO

Tutkimuksen aineisto on yhdistelmä otos- ja rekisteriaineistoista. Yksilötason otosaineistona toimii Katumetro-hankkeen (Katumetro 2014) kyselyaineisto. Vastaajien sijainti on peräisin otoksen mukana pyydytyistä perustiedoista, joihin kuuluu asuinalueiden keskipisteen koordinaatit. Chauvin (2005) pitää yksilötasoisien, riit-

tävän tarkasti geokoodattujen aineistojen harvinaisuutta yhtenä keskeisistä esteistä kontekstuaalisen terveystutkimuksen kehittymiselle, joten tämän tutkimuksen aineisto ja lähestymistapa vastaavat näiltä osin tutkimuskirjallisuudessa esitettyyn tarpeeseen. Kontekstimuuttujat ovat rekisteripohjaisia ja peräisin ruututietokannasta. Kontekstimuuttujia varten määritettiin kunkin vastaajan ruutu koordinaattien perusteella.

Katumetro-aineisto kerättiin vuonna 2012 Helsingin metropolialueelta (14 kuntaa) sekä Lohjalta ja Lahdesta. Helsingin metropolialueeksi luettiin Espoo, Helsinki, Hyvinkää, Järvenpää, Kauniainen, Kerava, Kirkkonummi, Mäntsälä, Nurmijärvi, Pornainen, Sipoo, Tuusula, Vantaa ja Vihti. Otannan näkökulmasta aineisto koostuu kahdesta ositetusta satunnaisotoksesta, joista toinen ositettiin Tilastokeskuksen ruututietokannan tilastoruutujen (250 x 250 metriä) mediaanitulotason pohjalta ja toinen asuinkunnan mukaisesti. Kummankin otoksen tavoiteperusjoukon tilastoyksiköitä ovat 25–74-vuotiaat suomen- ja ruotsinkieliset henkilöt. Ikä ja asuinpaikka määritettiin vuoden 2012 alussa väestörekisteritietojen perusteella. Ruutuotos poimittiin ylimpään ja alimpaan tuloviidennekseen kuuluvien ruutujen asukkaista. Kuntaotos on ehdollinen ruutuotokseen nähden siten, että mikäli henkilö tuli poimituksi ruutuotokseen, suljettiin hänen kotitaloutensa ulos toisesta otoksesta: näin varmistettiin, ettei sama henkilö voisi tulla kahteen kertaan poimituksi.

Brutto-otoksen koko on yhteensä 27 000 henkilöä, joista 9609 vastasi (vastausosuus 35,6 %). Korkean yksikkövastauskadon tilanteessa on perinpohjainen katoanalyysi keskeisessä asemassa. Kadon analyysissä käytimme paitsi erikseen pyytämämme yksilötason rekisteritietoja tuloista ja työttömyydestä, myös ruututietokannan tarjoamia, asuinympäristöä kuvaavia sosio-ekonomisia rekisterimuuttujia kadon mallintamisessa. Mallintamisen tuloksena saatavat vastaamistodennäköisyydet tuotiin lopulliseen painomuuttujaan, mikä parantaa aineiston käyttöedellytyksiä tavanomaisessa tilastollisen päättelyn kehikossa, jossa oletuksena on edustava satunnaisotos tai superpopulaatioon viittaava otostulkinta kokonaisaineistosta. Aineiston otanta-asetelmasta ja painotuksesta löytyy yksityiskohtainen selonteko lähteestä Laaksonen ym. (2014). Aineistosta jätettiin Lahden vastaajat pois, jotta saadaan mahdollisimman yhtenäinen, jatkuvan kokonaisuuden muodostava alue. Näin ollen lopullinen aineisto on kooltaan 9254 havaintoa.

Ruututietokanta on Tilastokeskuksen ylläpitämä aineisto, joka sisältää tilastoruuduittain koottua tietoa koko maan alueelta mm. ruudun väestöstä, asukkaiden talous- ja työmarkkina- asemasta sekä koulutuksesta (Tilastokeskus 2013). Tässä analyysissä on käytetty vuoden 2009 tietokantaa 250 x 250 metrin ruutukoolla, koska uudempaa versiota ei ollut tätä tutkimusta tehtäessä käytettävissä. Otosaineiston ja ruututiedon yhdistäminen onnistui koordinaattien avulla. Poikkeuksena olivat muutamat (54 kpl) otoshenkilöt, jotka asuivat kyselyn aikaan alueilla, joita ei vuonna 2009 ollut vielä asutettu tai rakennettu (Stjernberg 2013a). Tämä havainnollistaa eri aikaa edustavien aineistojen yhdistämiseen liittyviä haasteita ja erityisesti sitä, että jollakin tavalla muuttuneiden alueiden tapauksessa kontekstietto on vanhentunutta. Toisaalta käytetyt ruutupohjaiset muuttujat ovat varsin karkeita, joten voidaan olettaa, että suurin osa ruuduista tuli luokitelluksi oikealla tavalla.

Kuvio 1 havainnollistaa aineiston rakennetta. Mukana olevista 4134 ruudusta yli puolet on yhden vastaajan ruutuja ja yli kymmenen vastaajan ruutuja on 46 kappaletta.

Artikkelin havainnollistavina esimerkkeinä käytetään koettua terveyttä ja turvattomuutta. Kumpaakin vastetta on mitattu aineistossa vakiintuneilla kysymysmuotoiluilla. Terveyden osalta kysymys on: ”Kuinka hyvä terveydentilanne on nykyisin?” (koetusta terveydestä ks. esim. Manderbacka 1998, Heistaro ym 2001, Jylhä 2009). Vastausvaihtoehdot ovat: erittäin hyvä – melko hyvä – keskitasoa – melko huono – erittäin huono. Turvattomuutta tiedusteltiin kysymyksellä ”Kuinka turvalliseksi tunnettu olonne kävelles-

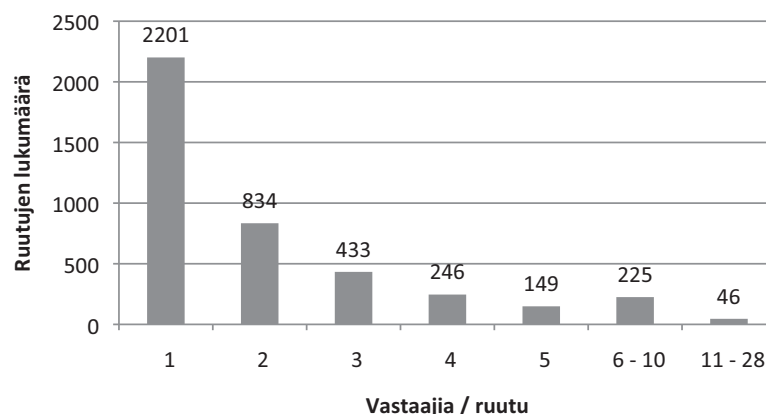
sänne omalla asuinalueellanne yksin myöhään viikonloppuiltana?” (ks. Ferraro ja LaGrange 1978, Farrall ym. 1997, Kempainen ym. 2014), ja vastausvaihtoehdot ovat: turvalliseksi – melko turvalliseksi – melko turvattomaksi – turvattomaksi – en uskalla käydä tuolloin ulkona – en liiku iltaisin ulkona. Turvattomuuskysymyksen viimeinen vastausvaihtoehto määritettiin analyysissä puuttuvaksi, koska se eroaa laadullisesti muusta skaalasta, mikä ratkaisuna jonkin verran lisää yksikkövastaukset. Muuttuja ei ole tarkalleen ottaen välimatka-asteikollinen, mutta tässä sitä käytetään likimääräisesti sellaisena.

#### HAVAINNOLLISTAVAN TUTKIMUSESIMERKIN ASETELMASTA

Spatiaalisten mallien havainnollistamiseksi tutkimuksessa käytetään kolmenlaisia regressiomalleja: spatiaalisten mallien lisäksi käytetään asetelmaperusteista regressiomallia (jatkossa AP) sekä mahdollisimman yksinkertaista lineaarista regressiomallia, johon viitataan tavallisen pienimmän neliösumman estimointitekniikan nimellä OLS (ordinary least squares). Otanta-asetelman monimutkaisuuden vuoksi AP-malli on luonteva valinta pääasialliseksi lähestymistavaksi, sillä sen avulla voidaan huomioida otannan erityispiirteitä ja kadon vaikutus osite-muuttujan ja painomuuttujan määrittelyllä. AP-mallit estimoitettiin SPSS:n Complex Samples -moduulilla (esim. Siller ja Tompkins 2006) ja aluemallit R:n *spdep*-paketilla Anselinin (2007) tapaan (ks. myös. Bivand 2014); R on nopeasti päivittyvä avoimen lähdekoodin analyysi- ja ohjelmointiympäristö, jonka voi ladata maksutta osoitteesta <http://cran.rstudio.com/>.

OLS-malli ei ota huomioon otanta-asetelmaa eikä painoja, mutta sen käyttö on tässä yhtey-

**Kuvio 1.**  
Vastaajien lukumäärä ruutua kohti.



dessä kuitenkin perusteltua, koska se tarjoaa hyvän vertailukohdan spatiaalisille malleille, koska niissäkin otanta-asetelma ja painot jäävät tämänhetkisillä tekniikoilla huomiotta. Sekä spatiaaliset mallit että AP-malli ovat OLS-mallia kehittyneempiä ratkaisuja, mutta ne paikkaavat eri asioita, jolloin ei ole itsestään selvää, kumpi olisi absoluuttisessa mielessä parempi. Kumpikin tarjoaa pikemminkin joitakin etuja, mutta tuo mukanaan myös omia rajoituksiaan.

Sisällöllisesti mallinnus lähtee liikkeelle asuin- ympäristöstä. Käsillä oleva tutkimus on osa Suomen Akatemian rahoittamaa hanketta (PREFA- RE, 2012–2015), jossa tarkastellaan hyvinvointia suomalaisissa 1960- ja 1970-luvuilla rakennetuissa lähiöissä. Asuinkontekstia kuvaava lähiömuuttuja on määritetty Ruututietokannan 250 x 250 -metrin tilastoruutujen avulla ja sen määrittämisen on kehittänyt kaupunkimaantieteilijä Mats Stjernberg (2013b): tässä yhteydessä lähiöllä viitataan alueeseen, joka on rakennettu 1960- tai 1970-luvulla, sijoittuu keskusta-alueiden ulkopuolelle (SYKE:n keskustarajaustyökalu) ja on kerrostalovaltainen (vähintään puolet ruudun asukkaista asuu lähiökerrostaloissa). Mallinnuksen lähtökohtana on näissä lähiöissä asumisen yhteys koettuun terveyteen ja turvallisuuteen, mikä on sikäli luontevaa, että lähiökeskustelulle on tyypillistä huoli sosio-ekonomisen huonosuuden alueellisesta keskittymisestä. Tätä yhteyttä lähdetään tarkastelemaan tuomalla malliin kontekstuaalinen, alueen sosio-ekonomista luonnetta kuvaava muuttuja sekä yksilötason demografisia ja sosio-ekonomisia muuttujia. Eri-tyistä huomiota kiinnitetään residuaalien spatiaaliseen korrelaatioon.

Sekä lähiöiden että muiden alueiden joukossa on sisäistä vaihtelua monen tekijän suhteen, joista kirjallisuuden perusteella merkittävimpiä lienee tutkittavien ilmiöiden kannalta alueen sosio-ekonominen rakenne. Tuoreessa suomalaisessa tutkimuksessa (Kemppainen ym. 2014) havaittiin, että Helsingissä asuinalueen sosio-ekonomisen tilan (SES) ja koetun turvallisuuden välillä on merkitsevä yhteys kohtuullisten vakiointienkin jälkeen siten, että mitä matalampi on alueen SES, sitä enemmän se altistaa asukkaitansa turvallisuuden kokemukselle (ks. myös Brunton-Smith ja Sturgis 2011). Näyttöä on myös alueen sosio-ekonomisen deprivaaation kontekstuaalisesta roolista terveysongelmien riskitekijänä (Chauvin 2005, Fone ja Dunstan 2006). Niinpä myös käsillä olevassa tutkimuksessa on analyysiin sisäl-

lytetty kontekstimuuttuja, joka kuvaa asuin- ympäristön – eli käytännössä ruudun – moniulotteista sosio-ekonomista haavoittuvuutta. Kolmeen yhtä suureen luokkaan katkaistu muuttuja perustuu valikoitujen ruututietokannan muuttujien (ruudun koulutus-, tulo- ja työmarkkinarakenne) faktorointiin.

Näiden kontekstimuuttujien jälkeen malleihin tuodaan demografisia ja sosio-ekonomisia yksilömuuttujia; subjektiivinen tulomuuttuja lisätään erikseen erotuksena objektiivisempiin sosio-ekonomisiin muuttujiin nähden. Näin päästään käsiksi kontekstimuuttujien ja selitettävien yksilötason muuttujien mahdollisiin ekologisiin yhteyksiin: mikäli yhteys häviää yksilötason muuttujien lisäämisen myötä, on kyse pohjimmiltaan ruututyypin väestökoostumuksen eroista ilman aitoa kontekstuaalista vaikutusta. Jos taas ekologinen yhteys säilyy, on mahdollista että mukana on kontekstivaikutusta. Näissä tapauksissa voidaan ajatella, että asuinkonteksti vaikuttaa tällöin kaikkiin, yksilöllisestä sosio-ekonomisesta asemasta riippumatta.

## **SPATIAALISTEN MALLIEN SOVELTAMISTA: KOETTU TERVEYS JA TURVATTOMUUS HELSINGIN METROPOLIALUEELLA**

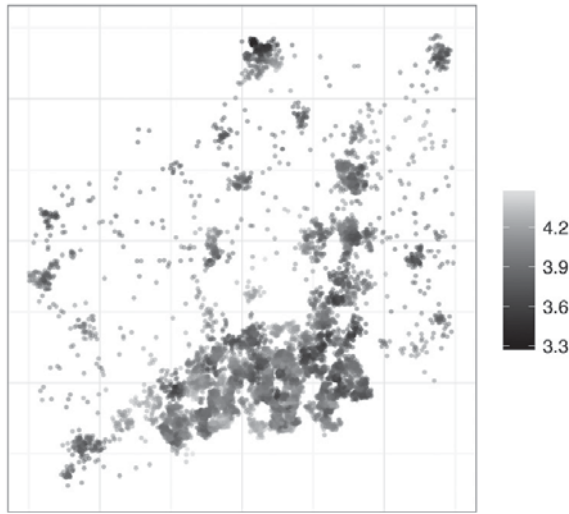
### **KOETUN TERVEYDEN JA TURVATTOMUUDEN ALUEELLINEN JAKAUTUMINEN**

Luvussa 2.2 esiteltyyn naapurustomatriisiin avulla voidaan muodostaa spatiaalinen viivemuuttuja, joka laskee jokaiselle havainnolle tämän naapuruston keskiarvon koetussa terveydessä ja turvallisuudessa. Yksittäisten vastausarvojen suora visualisointi olisi paitsi informaatiivisuuden näkökulmasta tehottomampi (vrt. Vauramo ym. 1992), myös tietoturvan mielessä vähemmän suotuisa vaihtoehto. Naapuruston kokoa vaihtelemalla voidaan joustavasti tarkastella koetun terveyden ja turvallisuuden alueellista rakennetta.

Kuviot 2 ja 3 kuvaavat koetun terveyden ja turvallisuuden alueellista jakautumista. Kuvi- oissa esitetään 50 lähimmän naapurin määrittämisen mukaisesti laskettu naapuruston keskiarvo kullekin havaintopisteelle. Naapuruston keskiarvon avulla päästään käsiksi ilmiöiden alueellisiin trendeihin, sillä menetelmä tasoittaa (smoothing) ilmiöstä saatavaa kuvausta poistamalla siitä yksilöllistä, ei-spatiaalista vaihtelua: koska vierekäisten pisteiden naapurustot ovat lähes samat, ovat keskiarvot hyvin lähellä toisiaan, jolloin ilmiön alueellinen rakenne piiryy näkyviin. Ha-

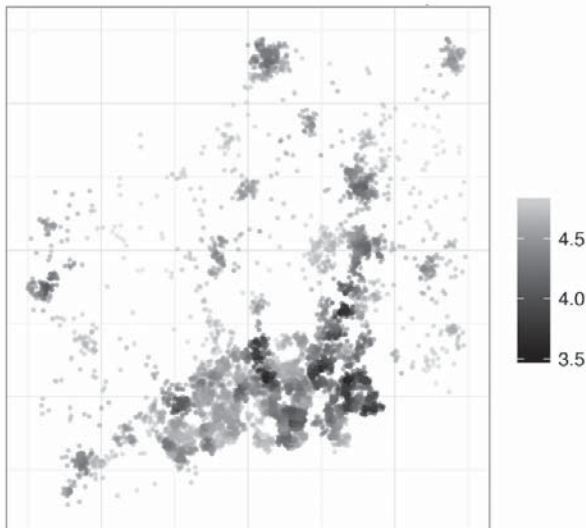
### Kuvio 2.

Koettu terveys. Naapuruston keskiarvo kullekin havaintopisteelle. 50 lähintä naapuria. Mitä tummempi piste, sitä heikompi koettu terveys. Katumetro-aineisto (ei Lahti), n=9133, vastauskaala 1–5.



### Kuvio 3.

Koettu turvattomuus. Naapuruston keskiarvo kullekin havaintopisteelle. 50 lähintä naapuria. Mitä tummempi piste, sitä turvattomampi. Katumetro-aineisto (ei Lahti), n=8630, vastauskaala 1–5.



vaintojen sijaintiin on lisätty pieni määrä satunaiskohinaa päällekkäin menevien pisteiden erottamiseksi ja tietoturvan parantamiseksi. Kuviot on piirretty verrattain suurilla ja sopivasti läpikuultavilla pisteillä mahdollisimman informatiivisen trendipinnan luomiseksi. Piirtämisessä on käytetty R:n ggplot2 -pakettia.

Koska karttatyylinen visualisointi helposti ohjaa ajatuksia kohti alueita koskevaa aineistoa,

on aiheellista korostaa, että kuviot 2 ja 3 perustuvat otospohjaiseen kyselyaineistoon, jossa on sekä mittausvirhettä että katoa. Tällaisen aineiston visuaalinen havainnollistaminen tasoittavalla viive-menetelmällä saa aineistosta esiin kiinnostavaa alueellista rakennetta. Voidaan myös ajatella, että keskiarvoistaminen tasoittaa mittausvirheestä johtuvaa satunnaisuutta. Erisuuruiset naapurustomääritykset tasoittavat kuvaa eri tavoin. Valittuun naapurustomäärityksiin päädyttiin kokeilemalla lukuisia erilaisia vaihtoehtoja. Pienen naapuruston tapauksessa (esim. 10 lähintä) on vaikea erottaa selkeitä trendejä ja suurilla naapurustoilla (esim. 400 lähintä) taas keskiarvojen vaihteluväli kutistuu epäedullisen pieneksi. 50 lähimmän naapurin ratkaisu muodosti lupaavalta vaikuttavan kompromissin näiden kahden piirteen suhteen, sillä siinä yhdistyy varsin selväpiirteinen aluetrendi kohtuullisiin keskiarvoeroihin.

Kuvien perusteella näyttää siltä, että kummassakin ilmiössä on tarkemman tutkimuksen arvoisia alue-eroja. Turvattomuuden alue-erot vaikuttavat suuremmilta kuin terveyden vastaavat. Turvattomimmat alueet eli *hot spotit* keskittyvät pitkälti juna- ja metroliikenteen varrelle, kun taas terveyden suhteen kuva on sirpaleisempi. Espoon kaakkoisrannikko on kummankin mittarin valossa vertailumielessä varsin hyvinvoivaa aluetta.

Visuaaliseen vaikutelmaan ei pidä kuitenkaan luottaa liiaksi, ja seuraavaksi onkin luontevaa ottaa askel kohti konfirmatorisempaa analyysiotetta, mihin tarjoaa mahdollisuuden edellä kuvattu Moranin I. Taulukkoon 1 on koottu kummankin muuttujan kohdalta kyseinen suure ja sen tilastollinen merkitsevyys normaalisuusoletuksella. Laskennassa on käytetty neljää erisuuruista naapurustomääritystä.

Moranin I:n nojalla arvioiden sekä turvattomuudessa että terveydessä on spatiaalista korrelaatiota. Spatiaalinen korrelaatio heikkenee odotetusti kontekstin kokoa kasvatettaessa, mutta sitä kuitenkin esiintyy kaikilla käytetyillä naapurustomäärityksillä. Spatiaalinen korrelaatio on turvattomuuden suhteen selvästi voimakkaampaa kuin terveydessä, mikä tukee karttakuvien pohjalta tehtyä eksploratiivista havaintoa.

Mikäli I mielletään korrelaatioosuureeksi, on arvioitava, missä määrin terveyden kohdalla esiintyvät hyvin pienet kertoimet ovat enää sisällöllisesti kiinnostavia. Tilastolliset merkitsevyydet sinänsä ovat voimakkaita kautta linjan aineiston melko suuren koon vuoksi, ja niiden laskennassa



## Taulukko 1.

Moranin I:t koetulle terveydelle ja turvallisuudelle. Neljä naapurustomäärittystä: 10, 50, 100 ja 400 lähintä naapurua.

Naapuruston koko	Terveys		Turvattuus	
	I	<i>p</i>	I	<i>p</i>
10	0.028	$\leq 0.0001$	0.165	$\leq 0.0001$
50	0.019	$\leq 0.0001$	0.131	$\leq 0.0001$
100	0.014	$\leq 0.0001$	0.103	$\leq 0.0001$
400	0.008	$\leq 0.0001$	0.060	$\leq 0.0001$

käytetyt varianssit ovat hyvin samansuuruiset niin normaalisuusoletuksella kuin Monte Carlo-satunnaistamisessakin. Tässä tapauksessa onkin pystyttävä arvioimaan tuloksen sisällöllistä merkittävyyttä muilla keinoin. Yksi tapa olisi ottaa pienempi otos, jonka avulla arvioida tilastollista merkitsevyyttä, mutta tähän ratkaisuun sisältyy liian mielivaltaisen tekijä siinä, kuinka suuren otoksen ottaa päättelyn pohjaksi. Arvioin tilastollisen merkitsevyyden käyttökelpoisuutta katsomalla, miten Moranin testisuure reagoi havaintojen sijainnin satunnaistamiseen, jolloin myös naapurustomatriisit määritellään kullekin naapurustokoolle uusiksi. Tällaisessa tapauksessa muuttujan alueellinen rakenne on satunnainen, jolloin testin pitäisi pystyä erottelemaan tämä tapaus todellisista jakaumista, joissa sen mukaan on spatiaalista korrelaatiota. Tässä se myös onnistuu, sillä testi ei aidon satunnaisessa tapauksessa anna kummankaan muuttujan osalta aiheita nollahypoteesin hylkäämiseen millään käytetyistä naapurustomäärittäyksistä.

### MALLINNUKSEN TEKNISIÄ HUOMIOITA JA SISÄLLÖLLISTÄ TULKINTAA

Tarkastellaan seuraavaksi regressiomallien avulla koetun terveyden ja turvallisuuden maantieteellisen puolen lisäksi niiden demografista ja sosioekonomista luonnetta. Mallinnuksen myötä pääsemme käsiksi mallin jäännöstermeihin eli residuaaleihin. Mikäli jäännöstermeissä on spatiaalista korrelaatiota, on ilmiöllä alueellista rakennetta, jota malli ei pysty selittämään. Tämän toteamista varten on kehitetty erityinen residuaaleille soveltuva versio Moranin indeksistä (Anselin 2007 ja 2009). Residuaalien spatiaalisen korrelaation tapauksessa on syytä vakavasti harkita, voisiko tilanne korjaantua esimerkiksi spatiaalisten mallien avulla. Vaihtoehtoisen tavan residuaaliongelman ratkaisuun tarjoaa ns. suodatusmenetelmä, jonka esittely ja soveltaminen on oman erillisen artikke-

linsa arvoinen aihe (spatial filtering, ks. esim. Getis ja Griffith 2002 sekä Griffith 2003).

Kuten edellä todettiin, mallinnuksen avulla haetaan vastausta kysymykseen, eroavatko Helsingin metropolialueen 1960- ja 1970-lukujen lähiöt muusta kaupunkirakenteesta asukkaiden kokeman terveyden tai turvallisuuden suhteen. Spatiaalisen mallinnuksen naapurirakenteen lähtökohdaksi valittiin kuvioissa käytetty 50 lähimmän naapurin määrittely, koska se on visuaalisesti arvioiden ja Moranin I:n perusteella hyvä kompromissi, kun halutaan säilyttää luokkien välisten keskiarvoerojen suuruus ilman että hukataan alueellisten trendien selkeyttä; myös mallinnusvertailu erisuuruuksilla naapurustoilla tukee tätä ratkaisua Akaiken informaatiomitalalla arvioituna.

Kuten edellä todettiin, merkitsee kiinteän naapurimäärän ratkaisu käytännössä sitä, että laskennallisen ”naapuruston” maantieteellinen koko vaihtelee havaintojen tiheyden mukaan. 50 lähimmän naapurin ratkaisussa valtaosalla (75 %) vastaajista naapurit löytyvät keskimäärin noin 900 metrin etäisyydeltä. Vain viidellä prosentilla keskimääräinen etäisyys on 2900 metriä tai enemmän. Harvimman havaintotiheyden alueilla tämä etäisyys on 15 kilometrin luokkaa, mutta näiden havaintojen osuus on varsin pieni, eikä naapuruston maantieteellistä laajuutta ole syytä pitää ongelmana.

Käsitellään aluksi artikkelin luonteen mukaisesti mallinnuksen teknisempää puolta, minkä jälkeen luodaan tiivis, mallinnuksiin pohjautuva sisällöllinen kuvaus koetun terveyden ja turvallisuuden alueellisesta luonteesta Helsingin seudulla. Sisällöllisen puolen osalta tulokset ovat ainoastaan suuntaa-antavia ja alustavia, koska täysi sisältöartikkelin vaatima kirjallisuuskehystys ja siihen pohjautuva mallin määrittely jouduttiin jättämään käsillä olevan tekstin tavoitteenasettelun ulkopuolelle.

Täysien OLS-mallien residuaalien tarkastelu osoittaa, että terveyden osalta tavanomainen mallinnus riittää: residuaalien Moranin I on käytännössä nolla. Turvattomuuden tapauksessa residuaalien autokorrelaatio sen sijaan antaa aihetta spatiaalisten mallien soveltamiseen. Havainnollisuuden vuoksi esitän taulukoissa myös koetun terveyden osalta vastaavat aluemallit (Taulukko 2). Mallien taulukoissa on esitetty vain tekstissä käsiteltävät muuttujat. Muut vakioidut muuttujat on mainittu taulukoiden alaviitteessä.

## Taulukko 2.

Koettu terveys: asetelmaperusteiset (AP) mallit, OLS-malli ja alueelliset viive- ja virhemallit.

	Yhden selittäjän AP-mallit			AP2			AP3	
	B	S.E.	<i>p</i>	B	S.E.	<i>p</i>	B	S.E.
Vakiotermi	.	.	.	3,809	.	.	4,442	.
Lähiöruutu								
Ei lähiö	ref.	.	.	ref.	.	.	ref.	.
Lähiö	-0,164	0,032	< 0.0005	-0,072	0,034	0,036	-0,041	0,031
Ruudun SES								
Matala SES	ref.	.	.	ref.	.	.	ref.	.
Keskitaso SES	0,196	0,027	< 0.0005	0,181	0,029	< 0.0005	0,056	0,026
Korkea SES	0,296	0,025	< 0.0005	0,276	0,027	< 0.0005	0,071	0,026
Kotitalouden toimeentulo (subj.)								
Erinomaisesti	ref.	.	.	.	.	.	ref.	.
Hyvin	-0,181	0,029	< 0.0005	.	.	.	-0,202	0,029
Kohtalaisesti	-0,528	0,030	< 0.0005	.	.	.	-0,482	0,032
Melko huonosti	-0,773	0,056	< 0.0005	.	.	.	-0,706	0,055
Huonosti	-1,160	0,101	< 0.0005	.	.	.	-1,034	0,104
n	8279			8279			8279	

1) Malleissa AP3, OLS3, VIIVE(50) ja VIRHE(50) vakioitu myös sukupuoli, ikä, kieli, työmarkkina-asema ja kotitalouden muoto.

	OLS3			VIIVE(50)			VIRHE(50)		
<i>p</i>	B	S.E.	<i>p</i>	B	S.E.	<i>p</i>	B	S.E.	<i>p</i>
.	4,475	.	.	4,161	.	.	4,473	.	.
.	ref.	.	.	ref.	.	.	ref.	.	.
0,177	-0,040	0,032	0,058	-0,046	0,026	0,073	-0,052	0,027	0,052
.	ref.	.	.	ref.	.	.	ref.	.	.
0,029	0,058	0,022	0,008	0,051	0,022	0,022	0,056	0,023	0,013
0,006	0,070	0,023	0,002	0,059	0,023	0,013	0,068	0,023	0,004
.	ref.	.	.	ref.	.	.	ref.	.	.
< 0.0005	-0,223	0,025	< 0.0005	-0,223	0,025	< 0.0005	-0,224	0,025	< 0.0005
< 0.0005	-0,487	0,026	< 0.0005	-0,486	0,026	< 0.0005	-0,487	0,026	< 0.0005
< 0.0005	-0,761	0,042	< 0.0005	-0,759	0,042	< 0.0005	-0,762	0,042	< 0.0005
< 0.0005	-0,987	0,070	< 0.0005	-0,987	0,070	< 0.0005	-0,989	0,070	< 0.0005
	8279			8279			8279		
				Rhoo	S.E.	<i>p</i>	Lambda	S.E.	<i>p</i>
				0,080	0,045	0,075	0,089	0,052	0,092

### Taulukko 3.

Koettu asuinalueen turvallisuus: asetelmaperusteiset (AP) mallit, OLS-malli ja alueelliset viive- ja virhemallit.

	Yhden selittäjän AP-mallit			AP2			AP3	
	B	S.E.	<i>p</i>	B	S.E.	<i>p</i>	B	S.E.
Vakiotermi	.	.	.	4,132	.	.	4,520	.
Lähiöruutu								
Ei lähiö	ref.	.	.	ref.	.	.	ref.	.
Lähiö	-0,240	0,033	< 0.0005	-0,090	0,035	0,010	-0,081	0,033
Ruudun SES								
Matala SES	ref.	.	.	ref.	.	.	ref.	.
Keskitaso SES	0,299	0,026	< 0.0005	0,279	0,027	< 0.0005	0,218	0,025
Korkea SES	0,480	0,024	< 0.0005	0,456	0,026	< 0.0005	0,355	0,025
Kotitalouden toimeentulo (subj.)								
Erinomaisesti	ref.	.	.	.	.	.	ref.	.
Hyvin	-0,152	0,026	< 0.0005	.	.	.	-0,086	0,025
Kohtalaisesti	-0,320	0,028	< 0.0005	.	.	.	-0,189	0,027
Melko huonosti	-0,486	0,052	< 0.0005	.	.	.	-0,308	0,051
Huonosti	-0,737	0,134	< 0.0005	.	.	.	-0,528	0,128
n	7817			7817			7817	

1) Malleissa AP3, OLS3, VIIVE(50) ja VIRHE(50) vakioitu myös sukupuoli, ikä, kieli, työmarkkina-asema ja kotitalouden muoto.

Regressiokertoimet ja niiden keskivirheet voivat kärsiä, mikäli selitettävän ilmiön spatiaalinen korrelaatio jätetään huomiotta (Ward ja Gleditsch 2008). Kuten taulukosta 3 huomataan, turvattomuuden tapauksessa asuin ympäristön lähiöluonnetta ja sosio-ekonomista tilaa luonnehtivien kontekstimuuttujien piste-estimaatit pienevät tuntuvasti, kun jäännöstermien spatiaalinen korrelaatio otetaan malleissa huomioon joko y-muuttujan viiveenä tai mallivirheen korrelaatiopakenteessa. Lähiömuuttujan kohdalla kumpikin spatiaalinen malli tuottaa varsin samansuuruisen kertoimen, mutta kontekstuaalisen SES-muuttujan tapauksessa viivemallin kertoimet ovat selvästi pienemmät kuin virhemallissa, mikä vastaa Wardin ja Gleditschin (2008, 69) havaintoa.

Lähiö-muuttujan suhteen piste-estimaatin pieneminen johtaa siihen, että perusteita nollahypoteesin hylkäämiselle ei lopullisessa mallissa enää ole. Vaikka spatiaalisten mallien kertoimet rho ja lambda ovatkin tilastollisesti melkein merkitseviä myös koetun terveyden tapauksessa, ovat ne sen verran pieniä, että niiden sisällöllinen merkittävyys on kyseenalainen.

Siinä missä residuaaleille kehitetty Moranin I -testin versio auttaa vastaamaan kysymykseen, onko tavanomainen mallinnus teknisesti soveltuva ratkaisu spatiaalisen korrelaation näkökulmasta, se ei kuitenkaan auta mallityypin valinnassa. Spatiaalisia malleja on lukuisia, lähtien ns. Manskin (1993) yleistä naapurustovaikutusmallia vastaavasta täydestä empiirisestä määrittämis-

	OLS3			VIIVE(50)			VIRHE(50)		
<i>p</i>	B	S.E.	<i>p</i>	B	S.E.	<i>p</i>	B	S.E.	<i>p</i>
.	4,465	.	.	2,168	.	.	4,604	.	.
.	ref.	.	.	ref.	.	.	ref.	.	.
0,015	-0,082	0,024	0,001	-0,043	0,023	0,064	-0,037	0,027	0,173
.	ref.	.	.	ref.	.	.	ref.	.	.
< 0.0005	0,234	0,020	< 0.0005	0,154	0,020	< 0.0005	0,170	0,022	< 0.0005
< 0.0005	0,363	0,021	< 0.0005	0,200	0,022	< 0.0005	0,240	0,025	< 0.0005
.	ref.	.	.	ref.	.	.	ref.	.	.
< 0.0005	-0,078	0,022	< 0.0005	-0,078	0,022	< 0.0005	-0,081	0,022	< 0.0005
< 0.0005	-0,183	0,024	< 0.0005	-0,180	0,023	< 0.0005	-0,183	0,023	< 0.0005
< 0.0005	-0,315	0,039	< 0.0005	-0,314	0,038	< 0.0005	-0,321	0,038	< 0.0005
< 0.0005	-0,429	0,065	< 0.0005	-0,434	0,063	< 0.0005	-0,433	0,063	< 0.0005
	7817			7817			7817		
				Rhoo	S.E.	<i>p</i>	Lambda	S.E.	<i>p</i>
				0,553	0,027	< 0.0005	0,612	0,028	< 0.0005

tä, jossa alueellinen komponentti on samaan aikaan sekä y-muuttujassa, selittäjissä että virhetermissä (Elhorst 2010). Tässä artikkelissa olen hahmottanut pitäytyä yksinkertaisimmista ja sovellusten tasolla vakiintuneimmista malleista, joita ovat edellä esiteltyt viive- ja virhemalli (esim. Fornango 2011, Sampson 2012, Troy ym. 2012). Näiden välisessä valinnassa on tapana käyttää sekä tilastollisia suureita että sisältöön pohjautuvaa harkintaa. Sisällöllisen harkinnan suhteen on hyvä muistaa, että virhe- ja viivemallit olettavat hyvin erityyppisen alueellisen luonteen ilmiölle: virhemallin kohdallaan ajatellaan, että sisällyttämällä mallin virhetermiin alueellinen komponentti saadaan huomioitua ainakin osa alueellisesti korreloituneesta, mittaamattomasta ja siten

mallin ulkopuolelle jäävästä selittäjästruktuurista. Viivemalli taas sisältää ajatuksen, että selitettävän muuttujan kuvaama ilmiö on itseensä nähden vuorovaikutteinen (esim. diffuusio, tarttuminen, kilpailu). Mallivalintaa ohjaavina tilastollisina suureina on käytetty Lagrange-kerrointen testisarjaa (Anselin 2007 ja Fornango 2011) ja erilaisia informaatiomittoja, esim. Akaiken informaatiokriteeriä (Troy ym. 2012).

Tässä tapauksessa Anselinin (2007) käyttämä Lagrange-kertoimien testisarja viittaa turvattuuden tapauksessa viivemallin suuntaan. On kuitenkin aihetta kriittisesti pohtia, voidaanko valinta jättää kokonaisuudessaan teknisen mittarin ratkaistavaksi (vrt. Gibbons ja Overmans 2012). Jotta viivemalli voitaisiin olettaa oikeaksi

valinnaksi, tulisi selitettävässä ilmiössä olla ”tarttuvaa”, itseään ruokkivaa tai muuten itsensä kanssa vuorovaikutteista luonnetta. Koetun turvattomuuden tapauksessa leviäminen on tiettyyn rajaan saakka teoreettisesti ajateltavissa: havainnot väkivallasta tai levottomuudesta – ja siten myös pelko – leviävät osaltaan huhuina ja kertomuksina kasvokkaisen vuorovaikutuksen välityksellä (vrt. Valkonen 1970). Turvattomuus voi toisaalta ruokkia itseään myös muiden tekijöiden kautta. Turvattomaksi koetun alueen käyttäjäkunta valikoituu, jolloin sosiaalista järjestystä tukevan epävirallisen sosiaalisen kontrollin edellytykset heikkenevät. Tällöin on kuitenkin kyse jo virhemallin mukaisesta tulkinnasta, koska oletetaan turvattomuuden kannalta relevantteja, mittaamattomia ja spatiaalisesti järjestyneitä tekijöitä, jotka erottavat muilta osin samantyyppisiä alueita toisistaan. Sosiaalisen disorganisaation teoriaperinteestä kumpuava epävirallinen sosiaalinen kontrolli (Sampson ym. 1997) on tämänhetkisen tiedon perusteella lupaavimpia tarjokkaita tähän rooliin. Teoreettisesti arvioituna tiukka viivemallitulkinta on lopulta epäuskottava, ja on turvallisempaa pitää kiinni virhemallin ajatuksista.

Mitä tässä esitettyjen mallien perusteella voidaan sanoa tutkittavista ilmiöistä sisällöllisen tulkinnan mielessä? Aloitetaan koetusta terveydestä, jonka osalta AP-mallit riittävät eikä aiheutta spatiaalisten mallien käyttöön ole. Lähtöasetelma on, että koettu terveys on lähiöissä muuhun seutuun verrattuna keskimäärin 0,16 pistettä heikompi (vastausskaala 1–5). Tästä valtaosa selittyy näiden asuin ympäristöjen sosio-ekonomisella luonteella (aluetta kuvaava kontekstimuuttuja) ja vastaajien demografisella ja sosiaalisella taustalla (vastaajia kuvaavat muuttujat). Matalan sosio-ekonomisen tason alueella asuminen on lopulta negatiivisesti, joskin varsin heikosti, yhteydessä koettuun terveyteen (ks. taulukko 3: AP3). Yksilöllisistä kovariaateista erityisesti koettu yksilöllinen taloustilanne näyttää olevan keskeinen koetun terveyden selittäjä: talousahdinkoa eniten kokevat raportoivat terveytensä keskimäärin noin yhden pisteen verran heikommaksi kuin parhaimmassa asemassa olevat.

Koetun turvattomuuden suhteen on sisällöllisiä johtopäätöksiä tehtäessä ammennettava eri malleista. Ensinnäkin tavanomainen mallinnustaktiikka kärsii pahoin jäännöstermien spatiaalisesta rakenteesta, joten spatiaalisten mallien tulokset ovat todennäköisesti oikeammassa suun-

nassa kuin AP3- ja OLS3-malleissa. Myös asetelman ja kadon huomiointi AP-malleissa tekee yhden varsin tuntuvaan korjauksen, jota ei sitäkään voi jättää huomiotta. Kootaan näitä seuraavaksi yhteen. Lähiöissä on varsin selvästi korkeampi turvattomuustaso kuin muilla alueilla: ero on lähtöasetelmassa 0,24 pistettä. Tämä selittyy lähes kokonaisuudessaan alueiden sosio-ekonomisella luonteella. Yksilötekijöiden vakioinnin ja alueellisen mallinnuksen jälkeenkin sosio-ekonomisen alueskaalan ääripäiden ero on 0,24 pistettä (virhemalli). Sisällöllinen tulkinta ei tässä tapauksessa muutu spatiaalisen korrelaation huomioimisen myötä. Koska kertoimen muutos on kuitenkin varsin tuntuva, on residuaalien korrelaatiokerenne syytä ottaa vakavasti: kertoimen pienemmällä lähtötasolla tai vastaavasti pienemmällä aineistolla tulos olisi saattanut laskea tavanomaisen merkitsevyydestason heikommalle puolelle.

Oma koettu toimeentulokin on kiinteästi yhteydessä koettuun turvattomuuteen: mitä hankalampi taloudellinen tilanne on, sitä enemmän turvattomuutta koetaan. Asetelmaperusteisesta mallista käy ilmi, että talousahdingon yhteys turvattomuuteen on jyrkempi kuin muiden mallien perusteella ajattelisi.

Eroavatko lähiöt siis muusta alueesta koetun terveyden ja turvattomuuden osalta? Kyllä, koettu hyvinvointi on kummankin muuttujan suhteen heikommalla tasolla lähiöissä kuin muualla. Mistä on kyse? Lähiöt ovat muuta aluetta tyyppillisemmin luonteeltaan matalan sosio-ekonomisen tason asuin ympäristöjä, mitkä puolestaan altistavat asukkaansa koetulle turvattomuudelle, pitkälti riippumatta asukkaan omasta sosio-ekonomisesta tilanteesta. Toisaalta lähiö- ja muiden alueiden välillä on eroja väestörakenteessa, mikä näkyy siinä, että vastaajien demografisen, sosiaalisen ja taloudellisen tilanteen huomioon ottaminen selittää käytännössä täysin terveyden tapauksessa alkuvertailussa havaitun eron. Tiivistäen voidaan todeta, että terveyden alue-eroissa on kyse enemmänkin kompositio- eli koostumus selityksestä, kun taas turvattomuuden tapauksessa näyttää jäävän tilaa myös varsin huomattavalle asuin ympäristön sosio-ekonomisen luonteen kontekstivaikutukselle (ks. Valkonen 1970 ja Chauvin 2005). Tuloksiin on syytä suhtautua varauksella, koska ne ovat sisällöllisessä mielessä vasta alustavia ja niiden tarkoitus on tässä yhteydessä toimia lähinnä artikkelissa esitellyn analyysimenetelmän havainnollistamisen tukena.

## YHTEENVETO JA POHDINTAA

Tässä artikkelissa esiteltä spatiaalisen ekonometrian lähestymistapa tarjoaa joustavan välineistön sosiaalitieteellisen hyvinvointitutkimuksen eri vaiheisiin ottamalla maantieteen kiinteästi mukaan analyysiin. Tutkimuksen eksploratiivisessa vaiheessa ilmiöiden alueellista rakennetta voidaan tarkastella monipuolisesti ja valita siltä pohjalta soveltuvimmalta vaikuttava naapurustorakenne tarkempaan analyysiin. Ilmiöstä riippuen spatiaalinen korrelaatio voi olla ongelma regressiomallinnuksessa. Kyse on havaintojen riippuvuudesta eli samasta ilmiöstä, johon esimerkiksi monitasoanalyysikin pyrkii tarttumaan ryvästyneiden ilmiöiden tapauksissa. Koettu terveys on ilmiönä siinä määrin ei-alueellista tekoa, että sen osalta tavanomaisen mallinusratkaisut riittävät. Toisin on koetun turvattomuuden laita, joka ymmärrettävästi on terveyttä alueellisempi ilmiö. Turvattomuuden tapauksessa tavanomainen mallinnus kärsii residuaalien spatiaalisesta korrelaatiosta ja sijainnin huomiointi vaikuttaa varsin tuntuvasti kontekstimuuttujien regressiokertoimiin. Spatiaalisen korrelaation ottaminen vakavasti mallien tasolla johtaa parempiin estimaatteihin, oikeampaan tilastolliseen päättelyyn ja näin ollen myös laadukkaampiin empiirisiin tuloksiin.

Esitellyllä lähestymistavalla on myös heikkoutensa. Viive- ja virhemallien erityiset spatiaaliseen rakenteeseen liittyvät kertoimet eivät ole sinällään erityisen informatiivisia, sillä ne kuvaavat lähinnä yleisellä tasolla spatiaalisuuden astetta. Mutkikkaan otanta-asetelman ja vastauskadon oloissa on erityisen ongelmallista, että asetelmuuttujia ei saada nykyisillä tekniikoilla mukaan kuvattuihin malleihin. Myös varsinaisten kontekstimuuttujien käyttöön liittyy tiettyä mutkikkautta. Ei ole täysin selvää, miten kontekstimuuttujien piste-estimaattien pieneneminen turvattomuuden spatiaalisissa malleissa tulisi tulkita, koska spatiaalisia ekonometrisia malleja käytetään tyyppillisesti selkeästi yksitasoisten aineistojen kanssa.

Spatiaalisten ekonometrinen mallien tapa käsittelee konteksti onkin tietyllä tapaa erilainen kuin tavanomaisessa monitasoisessa ajattelussa. Spatiaalisen ekonometrian lähestymistapa koostaa havainnon naapuruston piirteistä aggregoitua kontekstuaalista informaatiota. Näissä kontekstimuuttujissa on monitasoiselle asetelmalle tavanomaista redundanssia eli päällekkäisyyttä: vierekkäisten havaintojen naapurustot ovat lähes sa-

mat, jolloin muuttuja ei sisällä informaatiota samassa määrin kuin yksilömuuttujat. Verrattuna hallinnolliseen aluerakenteeseen tai tilastoruu- tuun tilanne on siinä mielessä erilainen, että nyt naapurusto liukuu vastaajan asuinkiinteistön sijainnin mukaan, jolloin päällekkäisyys on vain osittaista. Hallinnollisessa rakenteessa päällekkäisyys on täydellistä, koska jokaisella tietyllä alueen vastaajalla on sama ”naapurusto”, jolloin myös naapurustotason muuttujat ovat täsmälleen samat. Päällekkäisyyden olettaisi näkyvän keskivirheissä, mutta nyt kuitenkin nimenomaan piste-estimaatit pienenevät. On mahdollista, että spatiaalinen korrelaatio ei ole nyt niin voimakasta, että se vaikuttaisi varianssiestimaatteihin. Jatkossa on syytä selvittää, miten toinen tavallinen spatiaalisen autokorrelaation korjauskeino, spatiaalinen suodatus (Griffith 2003), toimii tämänkaltaisessa asetelmassa: se mahdollistaa yksinkertaisten estimointimenetelmien käytön ja siten otanta-asetelman ja vastauskadon huomioon. Lisäksi ryvästyneiden aineistojen tapauksessa suodatus taipuu myös monitasoanalyysiin (esim. Park ja Kim 2014).

Yleisellä tasolla tässä esiteltä lähestymistapa on eräs keino päästä kohti kokemusmaailmassa esiintyviä asuinympäristöjä, jotka eivät useinkaan ole yksiselitteisesti hallinnollisten rajojen määrittämiä. Naapurustoja ei ole suomalaisessa rekisterijärjestelmässä sellaisenaan olemassa (vrt. Préteceille 2003), joten niiden tilastollinen tarkastelu vaatii kehitystyötä, josta tässä artikkelissa esitetty lähestymistapa on yksi esimerkki. Naapuruston tai naapuriyhteisön kokoa ei kuitenkaan voida mekaanisesti määrittää tässä käytetyn aineiston kaltaisissa tapauksissa, joissa käsiteltävä alue on verrattain suuri ja maantieteellisesti jatkuva. Lisäksi on huomioitava, että vastaajan käsitys omasta asuinalueesta tai naapurustosta on subjektiivinen. Spatiaalinen analyysi antaa eräällä tapaa aineiston kertoa skaalan sopivuudesta kunkin tarkasteltavan ilmiön kohdalla erikseen. Sopivan tarkastelutason valinnassa voidaan käyttää erilaisten määritysten vertailua, teknisiä tunnuslukuja ja aiheen sisältötuntemukseen pohjautuvaa subjektiivista harkintaa. Tässä päädyttiin viidenkymmenen lähimmän vastaajan muodostamaan ”naapurustoon”.

Alueellisen analyysin yhteydessä on syytä kriittisesti arvioida vastaajien tietoturva. Tässä esiteltä analyysi ei vaaranna lainkaan vastaajien tietoturva, vaikka siinä käsitelläänkin havainnon sijaintia. Kuvioista on jätetty taustakartta

tarkoituksella pois, jotta pisteen paikantaminen ei olisi suoraviivaisesti mahdollista. Havaintojen sijaintiin on myös lisätty sopiva määrä satunnaiskohinaa niin, että yksikään piste ei olisi täsmälleen oikealla paikallaan. Lisäksi vastaajan asuin-kiinteistön keskipisteessä, jossa kukin kuvan piste siis sijaitsee, on edustettuna vastaajan naapuruston keskiarvo, jossa vastaajan oma arvo ei ole lainkaan mukana. Regressiomalleissa puolestaan naapuruus tai sijainti ei näyntyä tutkijalle saati lukijalle sellaisessa muodossa, joka herättäisi huolta tietoturvasta.

Jatkotutkimuksia ajatellen olisi mielenkiintoista selvittää, mitkä tekijät selittävät suomalaisessa kontekstissa asuinalueen sosio-ekonomisen luonteen ja koetun turvallisuuden yhteyttä. Eri-tyisesti olisi kiinnostavaa tarkastella, missä määrin kaupunkitutkimuksen keskeisiin käsitteisiin kuuluva epämuodollinen sosiaalinen kontrolli selittää tätä yhteyttä. Lähiötutkimuksen alueella olisi hyödyllistä päästä vertailemaan erityyppisiä lähiöitä turvallisuuden kokemuksen näkökulmasta. Näiden teemojen ymmärtäminen auttaa hahmottamaan sosio-ekonomisen eriytymiskehityksen tulevaisuuden suuntia, sillä asuinalueen turvallisuus, rauhallisuus ja näihin liittyvät mielikuvat ovat tunnetusti keskeisiä tekijöitä muutto-aikeiden muotoutumisessa.

## KIITOKSET

Tutkimus on osa Suomen Akatemian rahoittamaa hanketta ”Uusi kaupunkiköyhyys ja lähiöiden peruskorjaus” (PREFARE, 2011–2015). Aineiston keruuta on rahoittanut Kaupunkitutkimus ja metropolipolitiikka -ohjelma. Haluan kiittää kah- ta anonyymiä arvioijaa sekä Mikko Aaltosta (OP-TULA), Timo Kauppista (THL), Matti Kortteista (HY), Laura Lyytikäistä (TY), Henrik Lönnqvistiä (Helsingin kaupungin tietokeskus), Pekka Martikaista (HY), Mats Stjernbergiä (HY), Reijo Sundia (HY), Lasse Tarkiaista (HY), Martti Tuomista (Helsingin kaupungin tietokeskus), Teemu Turusta (HY) ja Mikko Weckrothia (HY) arvok- kaista ideoista ja kommentteista. Tutkimuksen alustavista tuloksista on keskusteltu professori Serge Paugamin kaupunkitutkimuksen kurssilla (EHESS, Pariisi), WIS 2014 -konferenssissa Tu- russa, ESHSM 2014 -konferenssissa Helsingissä, Sosiaalilääketieteen päivillä ja VTE-seminaarissa: kiitokset kaikille keskustelijoille. Rakennetun ympäristön tutkijakoulu ja sosiaalitieteiden tohtori- ohjelma ovat osallistuneet mainittujen konferens- simatkojen rahoitukseen.

Kemppainen T. Spatial modeling of point data: a survey-based analysis of self-rated health and feeling of unsafety in the Helsinki metropolitan area.

Sosiaalilääketieteellinen aikakauslehti – Journal of Social Medicine 2014:51:253–271

This article demonstrates the use of spatial regression models in social scientific survey research. Self-rated health and feeling of unsafety were chosen as topics for the illustrative practical example. The rationale for this choice is threefold. First, both are central areas of well-being. Second, they possess very different spatial structures, which serves well the purposes of illustration. Third, they stem from the two historical roots of the modern spatial analysis, namely, the fields of 19th century epidemiology and criminology.

Spatial analysis offers flexible and underused tools for social scientist. The spatial distribution of the phenomenon of interest can be visualised using different scales or resolutions. From the

point of view of modelling, one is able to account for the possible spatial autocorrelation of the phenomena, which results in more correct estimates. It is possible that the substantial interpretations depend on whether or not one accounts for spatial autocorrelation.

In our illustrative example, feeling of unsafety has a stronger spatial structure than self-rated health. As a consequence, accounting for spatial correlation makes a difference in the case of unsafety, contrary to that of self-rated health. The point estimates of the contextual variables diminish rather markedly in the spatial models of unsafety when compared to the standard models.



## KIRJALLISUUS

- Aaljoki T, Hiltunen A, Maaniitty M, Torvinen-Toroi P, Vänskä A, Hintikka J. Onko väestön mielenterveydessä alueellisia eroja? *Sosiaalilääk Aikak* 2000;37:6–13.
- Anselin L. *Spatial econometrics: methods and models*. Kluwer, Dordrecht 1988.
- Anselin L. *Spatial regression analysis in R*. A workbook. Revised version, May 10, 2007. [http://www.drs.wisc.edu/documents/articles/curtis/cesoc977-11/W15\\_Anselin2007.pdf](http://www.drs.wisc.edu/documents/articles/curtis/cesoc977-11/W15_Anselin2007.pdf) [luettu 24.4.2014].
- Anselin L. *Spatial regression*. Teoksessa Fotheringham A, Rogerson P. (toim.) *The Sage handbook of spatial analysis*. Sage, Los Angeles 2009, 255–275.
- Bernelius V. Eriytyvät kaupunkikoulut. Helsingin peruskoulujen oppilaspohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevaikutukset osana kaupungin eriytymiskehitystä. *Helsingin kaupungin tietokeskuksen tutkimuksia* 2013:1, Helsinki 2013.
- Bivand R. Package 'spdep'. Versio 0.5-71. The Comprehensive R-Archive Network, 31.1.2014. [cran.r-project.org/web/packages/spdep/spdep.pdf](http://cran.r-project.org/web/packages/spdep/spdep.pdf) [luettu 22.5.2014].
- Bivand R, Pebesma E, Gómez-Rubio V. *Applied spatial data analysis with R*. Springer, New York 2008.
- Blomgren J. Huono-osaisuus Suomen kaupunkiseutukunnissa: alue-erot ja sosiaalisen ympäristön vaikutukset 1990-luvulla. *Helsingin yliopiston sosiologian laitoksen tutkimuksia* nro 246, Helsinki 2005.
- Brunton-Smith I, Sturgis P. Do neighbourhoods generate fear of crime? An empirical test using the British Crime Survey. *Criminology* 2011;49:331–369.
- Chaix B, Merlo J, Chauvin P. Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilization in France. *J Epidemiol Community Health* 2005;59:517–526.
- Chauvin P. Environnement social et santé: avancées et perspectives dans l'étude des effets du contexte sur la santé. Teoksessa Chauvin P, Parizot I. (toim.) *Santé et expériences de soins: de l'individu à l'environnement social*. Inserm, Paris 2005, 187–200.
- Dubin R. *Spatial weights*. Teoksessa Fotheringham A, Rogerson P. (toim.) *The Sage handbook of spatial analysis*. Sage, Los Angeles 2009, 125–157.
- Dujardin C, Peeters D, Thomas I. Neighbourhood effects and endogeneity issues. Core discussion paper 2009/56. [http://www.ecore.be/DPs/dp\\_1254903898.pdf](http://www.ecore.be/DPs/dp_1254903898.pdf) [luettu 29.4.2014].
- Elhorst J. *Applied spatial econometrics: raising the bar*. *Spatial Economic Analysis* 2010;5:9–28.
- Farrall S, Bannister J, Ditton J, Gilchrist E. Questioning the measurement of the 'fear of crime'. *The British Journal of Criminology* 1997;37:658–679.
- Ferraro K, LaGrange R. The measurement of fear of crime. *Sociological Inquiry* 1987;57:70–101.
- Fischer M. *Spatial analysis in geography*. Teoksessa Smelser N, Wright J. (toim.) *International Encyclopedia of the Social and Behavioral Sciences*. Elsevier, 2001, 14752–14758.
- Fone D, Dunstan F. Mental health, places and people: a multilevel analysis of economic inactivity and social deprivation. *Health and Place* 2006;12:332–344.
- Fornango R. When space matters: spatial dependence, diagnostics and regression models. Teoksessa Pratt T. (toim.) *Advancing quantitative methods in criminology and criminal justice*. Routledge, London 2011, 15–33.
- Fortin M-J, Dale M. *Spatial autocorrelation*. Teoksessa Fotheringham A, Rogerson P. (toim.) *The Sage handbook of spatial analysis*. Sage, Los Angeles 2009, 89–103.
- Fotheringham A, Charlton M. Geographically weighted regression: a natural evolution of the expansion method for spatial analysis. *Environment and Planning A* 1998;30:1905–1927.
- Fotheringham A. *Geographically weighted regression*. Teoksessa Fotheringham A, Rogerson P. (toim.) *The Sage handbook of spatial analysis*. Sage, Los Angeles 2009, 243–253.
- Getis A. *Spatial autocorrelation*. Teoksessa Fischer M, Getis A. (toim.) *Handbook of applied spatial analysis. Software tools, methods and applications*. Heidelberg, Springer 2010, 255–278.
- Getis A, Griffith D. Comparative spatial filtering in regression analysis. *Geographical Analysis* 2002;34:130–140.
- Gibbons S, Overmans H. Mostly pointless spatial econometrics? *Journal of Regional Science* 2012;52:172–191.
- Griffith D. *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*. Springer, Berlin 2003.
- Heikkinen J. *Spatiaaliset pisteprosessit*. Kurssimoniste, Helsingin yliopisto, 2007. <http://www.rni.helsinki.fi/~jmh/spp07/spp07.pdf> [luettu 5.9.2014].
- Heistaro S, Jousilahti P, Lahelma E, Vartiainen E, Puska P. Self-rated health and mortality. A long term prospective study in eastern Finland. *J Epidemiol Community Health* 2001;55:227–232.
- Herttua K, Mäkelä P, Martikainen P, Sirén R. The impact of a large reduction in the price of alcohol on area differences in interpersonal violence: a natural experiment based on aggregate data. *J Epidemiol Community Health* 2008;62:995–1001.
- Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Soc Sci Med* 2009;69: 307–316.
- Karvonen S, Vikat A, Rimpelä M. *Monitasoanalyysi nuorten tupakoinnin seutukuntavaihtelun selittämisestä*. *Sosiaalilääk aikak* 2001;38:336–345.

- Kauppinen T. Asuinalueen ja perhetaustan vaikutukset helsinkiläisnuorten keskiasteen tutkintojen suorittamiseen. Helsingin kaupungin tietokeskuksen tutkimuksia 2004:6, Helsinki 2004.
- Katometro. Kaupunkitutkimus ja metropolipolitiikka -ohjelma. <http://www.helsinki.fi/kaupunkitutkimus/> [luettu 20.5.2014].
- Kempainen T, Lönnqvist H, Tuominen M. Turvattomuus ei jakaudu tasan: mitkä asuinalueen piirteet selittävät helsinkiläisten kokemaan turvattomuutta? *Yhteiskuntapolitiikka* 2014;79:5–20.
- Kortteinen M, Tuominen M, Vaattovaara M. Helsingin sosiaalimaantieteellinen rakenne ja pahoinpitelyrikollisuus. *Yhteiskuntapolitiikka* 2001;66:318–328.
- Kortteinen M, Tuominen M, Vaattovaara M. Asumistoiveet, sosiaalinen epäjärjestys ja kaupunkisuunnittelu pääkaupunkiseudulla. *Yhteiskuntapolitiikka* 2005;79:121–131.
- Kääriäinen J. Rikollisuuden pelko kasvaa Itä-Helsingissä. Asuinalueiden erilaistuminen ja turvattomuuden kokeminen Helsingissä ja Espoossa vuosina 1997 ja 2001. *Yhteiskuntapolitiikka* 2002;67:214–222.
- Laaksonen S, Kempainen T, Stjernberg M, Kortteinen M, Vaattovaara M, Lönnqvist H. Tackling city-regional dynamics in a survey using grid sampling. *Tulossa*, 2014.
- Lankinen M. Alueellisen eriytymisen suunta Helsingissä ja pääkaupunkiseudulla 1990-luvulla. Helsingin kaupungin tietokeskuksen tutkimuksia 2001:6, Helsinki 2001.
- Lehtonen O, Kauronen M-L. Aikuisväestön masennusta aiheuttavia paikallisia riskitekijöitä Suomessa. *Sosiaalilääk Aikak* 2013;50:114–126.
- Lehtonen O, Tykkyläinen M. Muuttoliikkeen alueelliset muodostumat ja pulssi Suomessa 1980–2006. *Terra* 2009;121:119–137.
- Lehtonen O, Tykkyläinen M. Tulotason spatiaalinen kaksoispolarisaatio Pohjois-Karjalassa 1996–2003. *Terra* 2010;122:63–87.
- LeSage J, Pace R. Spatial econometric models. Teoksessa Fischer M, Getis A (toim.) *Handbook of applied spatial analysis. Software tools, methods and applications*. Heidelberg, Springer 2010, 355–376.
- Loikkanen H, Saari J. (toim.) *Suomalaisen sosiaalipolitiikan alueellinen rakenne. Sosiaali- ja terveysturvan keskusliitto*, Helsinki 2000.
- Lyytikäinen T. Tax competition among local governments: evidence from a property tax reform in Finland. *Journal of Public Economics* 2012;96:584–595.
- Manderbacka K. Questions on survey questions on health. *Swedish Institute for Social Research* 30. Helsinki 1998.
- Manski C. Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies* 1993;60:531–542.
- Merlo J, Chaix B, Yang M, Lynch J, Råstam L. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health* 2005;59:443–449.
- Montello D. Scale in geography. Teoksessa Smelser N, Wright J. (toim.) *International Encyclopedia of the Social and Behavioral Sciences*. Elsevier, 2001, 13501–13504.
- Morgenstern H. Ecologic studies. Teoksessa Rothman K, Greenland S, Lash T. (toim.) *Modern epidemiology*. Third edition. Lippincott, Philadelphia 2008, 511–531.
- Paju P (toim.) *Samaan aikaan toisaalla... Nuoret, alueellisuus ja hyvinvointi. Nuorten elinolot vuosikirja IV. Nuorisotutkimusverkosto*, 2004.
- Park YM, Kim Y. A spatially filtered multilevel model to account for spatial dependency: application to self-rated health status in South Korea. *International Journal of Health Geographics* 2014;13.
- Préteceille E. La division sociale de l'espace francilien. *Typologie socioprofessionnelle 1999 et transformations de l'espace résidentiel 1990–99*. Observatoire sociologique du changement, Paris 2003.
- Rasinkangas J. Sosiaalinen eriytyminen Turun kaupunkiseudulla. Tutkimus asumisen alueellisista muutoksista ja asumispreferensseistä. *Siirtolaisuusinstituutti, tutkimuksia A 43*, Turku 2013.
- Ratvio R. Elämää keskustassa ja kaupunkiseudun reunoilla. *Urbaani ja jälkiesikaupungillinen elämäntytyli asumisen valinnoissa ja arkiliikkumisessa Helsingin seudulla*. Department of Geosciences and Geography A15, Helsinki 2012.
- Salonen M, Broberg A, Kyttä M, Toivonen T. Do suburban residents prefer the fastest or low-carbon travel modes? Combining public participation GIS and multimodal travel time analysis for daily mobility research. *Applied Geography* 2014;53:438–448.
- Sampson R. *Great American city: Chicago and the enduring neighborhood effect*. University of Chicago Press, Chicago 2012.
- Sampson R, Raudenbush S, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science* 1997;277:9–18.
- Siller A, Tompkins L. The big four: analyzing complex sample survey data using SAS, SPSS, STAT, and SUDAAN. *SUGI 31 Proceedings, Paper 172-31*. San Francisco, 2006. <http://www.bettycjung.net/Pdfs/Big4.pdf> [luettu 22.5.2014].
- Sipilä P, Martikainen P. Fyysisen ja psyykkisen terveyden väliset alue-erot ja niiden taustat pääkaupunkiseudulla. *Sosiaalilääk Aikak* 2005;42:202–218.
- Stjernberg M. *Henkilökohtainen tiedonanto*, 17.1.2013(a).
- Stjernberg M. *Concrete suburbia: a socio-economic analysis of suburban housing estates in Finland*. Konferenssipaperi, Nordic urban and housing research conference (NSBB), Roskilde 2013b.

- Sund R, Nouko-Juvonen S. Kehitysvammaapalvelujen käyttö ja kustannukset alueellisina ilmiöinä. Teoksessa Loikkanen H, Saari J. (toim.) Suomalaisen sosiaalipolitiikan alueellinen rakenne. Sosiaali- ja terveysturvan keskusliitto, Helsinki 2000.
- Tilastokeskus. Ruututietokanta 2013. Käyttäjän opas. [https://www.tilastokeskus.fi/tup/ruututietokanta/rttk2013\\_opas\\_fi.pdf](https://www.tilastokeskus.fi/tup/ruututietokanta/rttk2013_opas_fi.pdf) [luettu 2.5.2014].
- Tobler W. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 1970;46:234–240.
- Toivonen T. MetropAccess-reititin. Teoksessa Toivonen T ym. (toim.) MetropAccess: Helsingin metropolialueen moniulotteista saavutettavuutta tutkimassa. Helsingin yliopiston Geotieteiden ja maantieteen laitoksen julkaisuja C8 (painossa), Helsinki 2014.
- Troy A, Grove J, O’Neil-Dunne J. The relationship between tree canopy and crime rates across an urban–rural gradient in the greater Baltimore region. *Landscape and Planning* 2012;106:262–270.
- Vaattovaara M. Pääkaupunkiseudun sosiaalinen erilaistuminen. Helsingin kaupungin tietokeskuksen tutkimuksia 1998:7, Helsinki 1998.
- Valkonen T. Yhteisön vaikutukset yksilöön ja niiden kvantitatiivinen tutkiminen. Helsingin yliopisto, Helsinki 1970.
- Vauramo E, Mikkola P, Sippo-Tujunen I, Aro S, Alanko A, Pelanteri S, Hokkanen E. Coordinate-based mapping – a new method in health services research. *Med Inform* 1992;17:1–9.
- Vilkama K. Yhteinen kaupunki, eriytyvät kaupunginosat? Kantaväestön ja maahanmuuttajataustaisten asukkaiden eriytyminen ja muuttoliike pääkaupunkiseudulla. Helsingin kaupungin tietokeskuksen tutkimuksia 2011:2, Helsinki 2011.
- Virrantaus K. Paikkatiedoista hyötyä suunnitteluun ja päätöksentekoon – spatio-tilastollisen analyysin menetelmin. *Maankäyttö* 2011:1:6–11
- Ward M, Gleditsch K. Spatial regression models. Quantitative applications in the social sciences 155. Sage, Los Angeles 2008.
- Wheeler D, Páez A. Geographically weighted regression. Teoksessa Fischer M, Getis A. (toim.) Handbook of applied spatial analysis. Software tools, methods and applications. Heidelberg, Springer 2010, 461–486.

TEEMU KEMPPAINEN

*VTM, tohtorikoulutettava*

*Helsingin yliopisto*

*Sosiaalitieteiden laitos*