Reijo Sund

# TILASTOLLISIA MENETELMIÄ DYNAAMISTEN POTILASPOPULAATIOIDEN MALLINTAMISEEN

## Tapahtumahistoria-analyysia hoitoilmoitusrekisterin skitsofreenikoille

Aiheita 26/2000
Stakes

ISBN 951-33-1046-9 (verkkojulkaisu)

ISBN 951-33-1034-5 (paperijulkaisu)
ISSN 1236-9845

# KUVAILULEHTI

**Julkaisun päivämäärä**
Syyskuu 2000

**Tekijät**
Reijo Sund

**Julkaisija**
Stakes

**Kustantaja**
Stakes

**Julkaisun nimi** TILASTOLLISIA MENETELMIÄ DYNAAMISTEN POTILASPOPULAATIOIDEN MALLINTAMISEEN
Tapahtumahistoria-analyysia hoitoilmoitusrekisterin skitsofreenikoille

**Julkaisun sarja ja numero**
Aiheita-monistesarja 26/2000, Stakes

**Tiivistelmä**

Tässä tutkimuksessa tarkastellaan yksilötason rekisteriaineistojen dynaamiseen mallintamiseen sopivia tilastollisia menetelmiä. Pääasiallisena tavoitteena on esittää nykyisestä valtavirrasta poikkeavia metodologisia näkökulmia, jotka mahdollistavat rekisteriaineistojen entistä monipuolisemman hyödyntämisen niin tutkimuksessa kuin tilastotuotannossakin.

Mallintamisen lähtökohtana on tapahtumahistoria-analyysin viitekehys. Toisin sanoen kiinnostus kohdistuu useiden ajassa ilmenevien tapahtumien muodostamiin tapahtumasarjoihin. Tutkimuksessa kuvataan, kuinka dynaaminen ilmiö voidaan operationalisoida varannoista ja virtaamista koostuvaksi järjestelmäksi. Peruskäsitteet ja niiden väliset yhteydet esitetään sekä demografian että elinaika-analyysin merkintöjä käyttäen. Muilta osin rajaudutaan tarkastelemaan järjestelmän kokonaisvaltaiseen mallintamiseen soveltuvan semi-Markov-mallin muodostamiseen ja estimointiin tarvittavia ei-parametrisiä menetelmiä. Tutkimuksessa perustellaan myös, miten semi-Markov-malli on mahdollista sisällyttää Markovin ketjujen teoriakehykseen ja miksi mallin käyttö ei vaadi juuri tapahtumataulujen laskemista vaativampia toimenpiteitä.

Tilastolliset menetelmät ja mallit toimivat oletusten ollessa voimassa teknisessä mielessä riippumatta sovelluskohteesta, joten käytettävä aineisto luo ne puitteet, jotka sitovat abstraktin mallin havainnoitaviin ilmiöihin. Tässä tutkimuksessa aineiston lähteenä on Stakesin ylläpitämä hoitoilmoitusrekisteri. Rekisterin ominaisuuksia, syntymekanismeja, luotettavuutta ja käyttötapoja tarkastellaan omassa luvussaan.

Tutkimuksen soveltavassa osassa osoitetaan esitettyjen menetelmien soveltuvuus käytäntöön. Aineistona käytetään hoitoilmoitusrekisterin mukaan vuosien 1994-1998 aikana skitsofreniadiagnoosilla hoitoa saaneita potilaita. Potilaspopulaation yleisten ominaisuuksien tarkastelun lisäksi muodostetaan populaation liikkeitä kuvaava semi-Markov-malli sekä tarkastellaan sen validiteettia ja käyttökelpoisuutta ennustamiseen.

Tärkeimmät lähteet:

Bartholomew, David J. & Forbes, Andrew F. & McClean, Sally I. (1991): Statistical Techniques for Manpower Planning. Second Edition. John Wiley & Sons. Chichester.

Hakkarainen, Anni (1989): Psykiatrisen potilasväestön muutokset: teoreettinen malli ja sen sovellutukset Suomessa. Lääkintöhallituksen tutkimuksia 48. Helsinki.

**Avainsanat**
tapahtumahistoria-analyysi, Markov-mallit, hoitoilmoitusrekisteri, skitsofrenia

**Muut tiedot (esim. elektroninen julkaisu tai verkkojulkaisun osoite)**
Pro gradu –tutkielma (Helsingin yliopisto, Tilastotieteen laitos)

**ISSN**
1236-9845

**ISBN**
951-33-1034-5

**Kokonaissivumäärä**
106

**Kieli**
suomi

**Hinta**
113 mk (sis.alv)

**Jakaja ja myyjä**
Stakes, PL 220, 00531 Helsinki, puh (09) 3967 2140 tai (09) 3967 2141 tai automaatti (09) 3967 2308, faksi (09) 3967 2450

# Sisällys

**1 Johdanto ..... 1**

1.1 Tutkimuksen tarkoitus ja rajaukset ..... 4
1.2 Aikaisempia tutkimuksia ..... 6

**2 Dynaamisten ilmiöiden mallintamisesta ..... 8**

2.1 Dynaaminen ilmiö mallinnettavana järjestelmänä ..... 9
2.2 Dynaamisen järjestelmän mallintamisen perusteita ..... 11
*2.2.1 Tapahtumataulu ..... 13*
*2.2.2 Eloonjäämisfunktiot ja niiden väliset yhteydet ..... 15*
*2.2.3 Ei-parametrisesta eloonjäämisfunktioiden estimoinnista ..... 16*
*2.2.4 Eloonjäämisfunktioiden vertailusta ..... 20*
2.3 Markov-malleista ..... 22
*2.3.1 Markovin ketjut ..... 23*
*2.3.2 Semi-Markov-malli ..... 32*
*2.3.3 Semi-Markov-mallin ei-parametrinen estimointi ..... 37*
*2.3.4 Mallin validiteetti ..... 39*

**3 Hoitoilmoitusjärjestelmä ..... 42**

3.1 Hoitoilmoitusrekisterin luotettavuus ja kattavuus ..... 44

**4 Skitsofreenikoiden potilaspopulaatio ..... 51**

4.1 Skitsofreenisten häiriöiden potilaspopulaation ominaisuuksia ..... 55
*4.1.1 Laitosvaranto ..... 55*
*4.1.2 Hoitokerrat ..... 57*
*4.1.3 Ryhmittely ja ryhmien ominaisuudet ..... 59*
*4.1.4 Volyymit ja ottamistaajuudet ..... 61*
*4.1.5 Hoito- ja uusiutumisaikajakaumat ..... 65*
4.2 Potilaspopulaatiota kuvaava dynaaminen malli ..... 67
*4.2.1 Mallin estimointi ja validiteetti ..... 69*

**5 Pohdintaa ..... 78**

5.1 Aineiston hyödyntämisen ennakkoehdoista ..... 81
5.2 Heterogeenisuuden hallinnasta ..... 85
5.3 Lopuksi ..... 86

**Kirjallisuus ..... 88**

# 1 Johdanto

Sosiaali- ja terveydenhuollon resurssit ovat rajallisia. Näiden resurssien mahdollisimman tehokas hyödyntäminen vaatii hallinnollista suunnittelua ja poliittista tahtoa. Päätöksenteon ja valvonnan tueksi tarvitaan tieteellisesti validia ja vertailukelpoista tietoa. Suomessa terveyttä koskevat tilastot ja rekisterit ovat kansainvälisestikin katsoen varsin monipuoliset ja niiden hyödyntämisen edellytykset periaatteessa hyvät (Aromaa 1980; Koskinen ym. 1998). Tietojärjestelmissä on kuitenkin aukkoja ja tietojen hyväksikäytössä puutteita. Aromaata (1980) ja Hakkaraista (2000) lainaten on suorastaan paradoksaalista, että maassamme on ollut varaa luoda hallinnollisia tarkoituksia varten koko väestöä koskevia ainutlaatuisia yksilötasoisia rekistereitä, mutta ettei tietojen monipuolisempaan hyödyntämiseen panosteta enempää voimavaroja edes nykyisen tietoyhteiskunnan aikakautena.

Vallitsevan tiedontarpeen tyydyttämiseksi on viime vuosina sosiaali- ja terveydenhuollon tilastojärjestelmien uudistuksen myötä tärkeänä tavoitteena ollut parantaa rekisteritietojen hyväksikäyttöä, mihin on pyritty nopeuttamalla tietojen käyttöön saamista ja kehittämällä tiedoille erilaisia tietopalvelukanavia (Hakkarainen 1998; Hämäläinen 1998). Tilastotuotannossa säästetään huomattavasti yhteiskunnan kustannuksia käyttämällä rekistereitä muun muassa siksi, ettei tietoja tarvitse erikseen kerätä (Hämäläinen 1998). Rekisteritietojen kerääminen ei siten saa olla itsetarkoitus, vaan kaikella kerätyllä tiedolla on oltava käyttöä niin tietoja tuottavalle yksikölle itselleen, sosiaali- ja terveydenhuollon päätöksentekijöille kuin myös alan tutkijoille (vrt. Miikki & Orre 1998; Orre 1998; Stakes 1999b). Rekisterien tiedoista tuotetaankin erilaisia kokoavia tilastojulkaisuja ja lasketaan alueittaisia tunnuslukuja, jolloin tiedot soveltuvat myös kuntien, palvelujen tuottajien, läänien, sairaanhoitopiirien sekä valtakunnallisen suunnittelun ja seurannan apuvälineiksi.

Rajoittumalla pelkkiin valmiiksi jalostettuihin tilastoihin jätetään valtaosa laajan rekisteriaineiston mahdollisuuksista kokonaan hyödyntämättä (Hakkarainen 1998). Yksilötason rekisteriaineiston hyödyntäminen vaatii kuitenkin tutkimuksellista otetta ja niin vahvaa metodista osaamista, ettei sellaista voida edellyttää jokaiselta tiedon

haluajalta. Onkin yleensä tapana puhua rekisteritutkimuksesta, kun varsinaista rekisteriaineistoa käytetään tutkimuksen tietolähteenä (ks. esim. Gissler 1999a; 1999b; Keskimäki ym. 1997; Koskinen ym. 1998; Mähönen 1993, 90-124; Sosiaalivakuutus 1980, 201-253; Teperi 1994; Valkonen ym. 1998a).

Periaatteessa nykyiset tietojärjestelmät mahdollistavat tietojen automaattisen työstämisen haluttuun muotoon monimutkaisissakin tapauksissa, jos vain tarvittavat säännöt pystytään muotoilemaan riittävän täsmällisesti. Ideaalitapauksessa rekisteriaineistoa voidaankin hyödyntää entistä monipuolisemmin yksinkertaisesti ”upottamalla” tietojärjestelmän sisälle sopivia malleja, jotka tuottavat automaattisesti käyttökelpoisia ja helposti tulkittavia tuloksia sekä tunnuslukuja niitä tarvitseville. Kuvatun kaltaisessa automatisoinnissa on kysymys enemmän yksilötason rekisteriaineiston ja parantuneiden tietojenkäsittelyresurssien luomista metodisista mahdollisuuksista kuin tiedontuotannon uudesta ”vallankumouksellisesta” viitekehyksestä.

Vaikka tutkimuskohde ja siihen liittyvät kysymyksenasettelut ovat olemassa teknisistä ratkaisuista riippumatta, ei ole itsestään selvää, millaisia tietoja tarvitaan (vrt. Dillon 1994; Knowles 1994; Millard 1994) ja millä ehdoilla tietoja on mahdollista tuottaa (ks. kohta 5.1). Rekisteriaineistoja kannattaakin lähestyä metodologisesti riittävän abstraktilla ja joustavalla tasolla, jotta menetelmiä voitaisiin hyödyntää useamman kuin yhden ilmiön mallintamisessa.

Analyyttiset eli matemaattisesti tai loogisesti täsmälliset menetelmät sopivat hyvin rekisteriaineistoja hyödyntävään mallintamiseen. Niiden piirissä voidaan erottaa toisistaan deterministiset ja ei-deterministiset sekä staattiset ja dynaamiset lähestymistavat seuraavan nelikentän mukaisesti (vrt. McClean 1994):

| | **Staattinen** | **Dynaaminen** |
|---|---|---|
| **Deterministinen** | Algebralliset yhteydet | Differentiaali-/ differenssiyhtälöt |
| **Ei-deterministinen** | Probabilistiset yhteydet | Stokastiset mallit |

Todellisuuden satunnaisuus tekee mallinnettavista ilmiöistä lähes poikkeuksetta ei-deterministisiä. Alkutilasta ei toisin sanoen voida varmuudella päätellä lopputilaa, joten epävarmuus on otettava malleissa huomioon esimerkiksi todennäköisyyslaskennan keinoin. Staattisessa lähestymistavassa aineistoa käytetään panosten ja tuotosten - resurssien ja suoritteiden - mittaamiseen. Halutaan esimerkiksi tietää, kuinka monta asiakasta on tuottanut havaitun määrän hoitopäiviä ja mitä kautta asiakasvirta kulkee järjestelmän läpi (vrt. Hakkarainen 2000). Dynaamisessa ajattelutavassa kiinnostuksen kohteena ovat puolestaan ajan myötä tapahtuvat muutokset. Esimerkiksi Härö (1980) on pukenut dynaamisen ajattelutavan sanoiksi ehdottaessaan henkilön sairaalakäytön seurantaa pidemmän ajanjakson kuluessa. Staattinen ja dynaaminen näkökulma eivät ole toisiaan poissulkevia tai erillisiä vaan pikemminkin toisiaan täydentäviä. Perustavanlaatuinen ero näkökulmissa ilmenee esimerkiksi hoitopäivän käsitteessä: staattisessa ajattelutavassa kysytään, kuinka monta hoitopäivää potilaalle kertyy, ja dynaamisessa puolestaan, kuinka kauan potilas viipyy sairaalassa. Staattisessa ajattelussa hoitopäivä on siis tuotosten mittari ja dynaamisessa puolestaan keston mittari (Hakkarainen 1989, 3).

Dynaaminen lähestymistapa on jäänyt suomalaisessa terveydenhuollon tutkimuksessa hyvin vähälle huomiolle, vaikka suuri osa käytettävissä olevista yksilötason rekisteriaineistoista soveltuu sellaisenaan myös dynaamisen mallintamisen tarpeisiin. Dynaamisten mallien puitteissa on mahdollista tehdä myös monipuolisia ”entä jos” - ennustuksia (what-if analysis); voidaan esimerkiksi kokeilla erilaisten hoitopolitiikkojen lopputuloksia (vrt. McClean 1994; Hakkarainen 2000).

## 1.1 Tutkimuksen tarkoitus ja rajaukset

Tässä tutkimuksessa tarkastellaan yksilötason rekisteriaineistojen dynaamiseen mallintamiseen sopivia tilastollisia menetelmiä. Pääasiallisena tavoitteena on esittää nykyisestä valtavirrasta poikkeavia metodologisia näkökulmia, jotka mahdollistavat rekisteriaineistojen entistä monipuolisemman hyödyntämisen niin tutkimuksessa kuin tilastotuotannossakin.

Tutkimus on jaettu viiteen lukuun, jotka kaikki täydentävät toisiaan. Ensimmäinen luku toimii kysymyksenasettelua pohjustavana yleisenä johdatteluna aihepiiriin sekä tiivistelmänomaisena katsauksena tutkimuksen sisältöön. Lisäksi tarkoituksena on kuvata lyhyesti kunkin luvun merkitystä kokonaisuuden kannalta.

Tämän jälkeen paneudutaan tarvittaviin käsitteisiin ja menetelmiin tilastoteoreettisella tasolla. Metodologisena lähtökohtana on tapahtumahistoria-analyysin viitekehys, jossa kiinnostus kohdistuu useiden ajassa ilmenevien tapahtumien muodostamiin tapahtumasarjoihin. Tarkoituksena on kuvata, kuinka dynaaminen ilmiö voidaan operationalisoida mallinnettavaksi järjestelmäksi ja edelleen, kuinka järjestelmä koostuu yksilöiden varannoista ja virtaamista. Dynaamisten ilmiöiden osoitetaan järjestelmän puitteissa olevan formalisoitavissa sekä demografian että elinaika-analyysin käsitteitä käyttäen. Järjestelmän osien mallintamiseen sopivien menetelmien kirjo on varsin suuri, joten johdonmukaisen kokonaisuuden muodostamiseksi rajaudutaan tarkastelemaan ei-parametrisia menetelmiä, joita tarvitaan koko järjestelmää kuvaavan stokastisen mallin muodostamisessa ja estimoinnissa. Stokastisten mallien osalta keskitytään Markovin ketjujen ja semi-Markov-mallien matemaattisiin toimintaperiaatteisiin. Lopuksi perustellaan, miten semi-Markov-malli on mahdollista sisällyttää Markovin ketjujen teoriakehykseen ja kuinka mallin estimointi sekä oletusten testaaminen voidaan palauttaa hyvin yksinkertaisten ei-parametristen mallien soveltamiseksi. Toisin sanoen näennäisestä monimutkaisuudestaan huolimatta varannoista ja virtaamista muodostuvaa dynaamista järjestelmää kuvaavan semi-Markov-mallin käyttö ei periaatteessa vaadi juurikaan väestötieteilijän tapahtumataulujen laskemista vaativampia toimenpiteitä. Lisäksi rajautumalla ei-parametrisiin menetelmiin mallin käyttökelpoisuuteen

oleellisesti vaikuttavat taustaoletukset saadaan minimoitua, joten malli on periaatteessa hyödynnettävissä vähäisilläkin matemaattisilla taustatiedoilla.

Tilastolliset menetelmät ja mallit toimivat oletusten ollessa voimassa teknisessä mielessä riippumatta sovelluskohteesta, joten käytettävä aineisto luo ne puitteet, jotka sitovat abstraktin mallin havainnoitaviin ilmiöihin. Tässä tutkimuksessa rajaudutaan käyttämään hoitoilmoitusrekisterin tietoja. Kolmannessa luvussa tarkastellaankin hoitoilmoitusjärjestelmää, sen ominaisuuksia, syntymekanismeja, luotettavuutta ja käyttötapoja. Luvussa kuvataan, millaisia tietoja hoitoilmoitusrekisteri sisältää ja millaisia ongelmia sen hyödyntämiseen liittyy. Päämääränä on antaa käsitys olemassa olevan aineiston yleisestä käyttökelpoisuudesta ja toisaalta tunnetuista rajoituksista.

Neljäs luku on tutkimuksen soveltava osa. Siinä aikaisemmat luvut punotaan yhteen käyttämällä hoitoilmoitusrekisterin tietoja aikaisemmin esiteltyjen menetelmien aineistona. Varsinaisena tutkimusaineistona käytettävien skitsofreniapotilaiden valinta perustellaan ensin rekisterin tietoja hyödyntäen, jonka jälkeen populaation ominaisuuksia tarkastellaan yksinkertaisten kuvailevien analyysien avulla. Tarkastelut on pyritty rajaamaan siten, että ne pystytään esittämään kumulatiivisessa järjestyksessä eli saatuja tuloksia käytetään uusien aikaansaamiseksi. Alustavien analyysien jälkeen muodostetaan ja estimoidaan valitun skitsofreenikoista koostuvan potilaspopulaation liikkeitä kuvaava semi-Markov-malli sekä tarkastellaan sen validiteettia ja käyttökelpoisuutta varantojen ennustamiseen. Luvun tarkoitus on osoittaa esitettyjen menetelmien käytännön soveltuvuus analyysiin ja mallintamiseen hoitoilmoitusrekisterin tapauksessa.

Lopuksi esitetään pohdintaa, joka täydentää tutkimusta muita osioita kriittisemmillä ja vapaamuotoisemmilla näkökulmilla. Tarkoituksena on myös yhteenvetävästi arvioida tämän tutkimuksen tarkoituksenmukaisuutta.

## 1.2 Aikaisempia tutkimuksia

Tämän tutkimuksen lähtökohtana on Anni Hakkaraisen väitöskirjassa kulminoitunut perustavanlaatuinen työ dynaamisen näkökulman puolesta, jossa hän istuttaa hoitoilmoitusrekisterin tietoja käyttävän psykiatrista potilasväestöä kuvaavan dynaamisen mallin demografiseen viitekehykseen sekä esittää siihen kuuluvaa käsitteistöä ja sovelluksissa tarvittavia menetelmiä (Hakkarainen 1980; 1982; 1985; 1989). Hoitoilmoitusrekisterin aineistoja hyödyntävissä tutkimuksissa dynaamista lähestymistapaa on sovellettu lisäksi ainakin päiväkirurgian potilaiden hoitoaikajakaumien kuvaamiseen (Hakkarainen ym. 1999), selkäleikkausten uusintariskien tarkasteluun (Keskimäki ym. 2000) ja hoitopäivätuotosten (Ruotsalainen 1991, 95-98) sekä erikoissairaanhoidon kustannusten (Häkkinen ym. 1994) simulointiin. Viime aikoina on alettu puhua myös hoitoketjujen seurannasta ja joustavan hoidon ideologiasta (ks. esim. Kauppinen ym. 1999; Nylander & Nenonen 2000; Nylander ym. 1998; Rissanen & Noro 1999). Nämä käsitteet sisältävät myös dynaamisen lähestymistavan edellyttämiä piirteitä, mutta toistaiseksi aikaulottuvuus on jäänyt tarkasteluissa varsin vähälle huomiolle.

Millardin ja McCleanin (1994; 1996) toimittama kirjapari toimii erinomaisena johdatteluna terveydenhuollon suunnitteluun ja mallintamiseen liittyviin kysymyksiin. Kirjoissa käsitellystä vanhustenhuollon mallintamisen teemasta löytyy myös teknisempiä esityksiä (ks. esim. McClean & Millard 1993; Taylor ym. 1998a; 1998b). Delasien ym. (1998) toimittamassa European Journal of Operational Research -lehden erikoisnumerossa tarkastellaan yleisemmin terveydenhuollon suunnittelun viimeaikaisia suuntauksia operaatioanalyysin näkökulmasta. Erityisesti Delasien (1998), Lagergrenin (1998) sekä Roystonin (1998) katsaukset johdattavat aihetta käsittelevään kirjallisuuteen. Kiinnostava on myös Lin ja Bentonin (1996) terveydenhuollon organisaatioiden tehokkuuden mittaamista koskava katsaus. Suomalaisesta kirjallisuudesta tutustumisen arvoisia ovat esimerkiksi Gisslerin (1999b), Keskimäen (1997), Linnan (1999), Rissasen (1996), Ruotsalaisen (1991) ja Teperin (1994) tutkimukset, Valkosen ym. (1998a) toimittama rekisteriaineistoja käsittelevä teos sekä Kanniston (2000) väestötutkimuksen ja sosiaalilääketieteen yhteyksiä kartoittava puheenvuoro.

Menetelmäkirjallisuuden tasolla dynaamista mallintamista käsittelevää materiaalia on olemassa varsin runsaasti. Tämän tutkimuksen kannalta käyttökelpoisimmiksi lähteiksi osoittautuivat loppujen lopuksi työvoiman suunnittelun menetelmiä kuvaavat esitykset (Bartholomew 1971; Bartholomew ym. 1991), sosiaalisten prosessien mallintamiseen liittyvä kirjallisuus (Bartholomew 1981; 1982) ja kokoava aktuaarimatemaattinen katsaus (Elandt-Johnson & Johnson 1980). Tärkeitä lähteitä olivat myös elinaika-analyysia (Cox & Oakes 1984; Kalbfleisch & Prentice 1980; Le 1997) ja sovellettuja todennäköisyysmalleja (Feller 1968; Ross 1970) käsittelevät teokset. Dynaamista lähestymistapaa soveltavasta skitsofreniaa ja psykiatrista potilasväestöä koskevasta kirjallisuudesta hyödyllisiä ovat esimerkiksi Brownin (1960), Eatonin ja Whitmoren (1977), Hakkaraisen (1989) ja Von Korffin (1978) tutkimukset.

# 2 Dynaamisten ilmiöiden mallintamisesta

Dynaamisia ilmiöitä tarkasteltaessa mielenkiinto kohdistuu useiden ajassa ilmenevien tapahtumien muodostamiin tapahtumasarjoihin eli tapahtumahistorioihin. Yleensä seurataan sopivien yksiköiden liikkumista eri ”tilojen” välillä; yksinkertaisimmassa tapauksessa vain yhdenlainen muutostapahtuma on mahdollinen. Monissa tilanteissa yhden tapahtuman tarkastelu ei kuitenkaan kuvaa ilmiön dynamiikkaa realistisella tavalla. Esimerkiksi hitaasti kehittyvää sairautta, joka etenee kuolemaan monien toisiaan seuraavien tilojen kautta ei pystytä kattavasti mallintamaan havainnoimalla ainoastaan ensi diagnoosista kuolemaan kuluvaa aikaa.

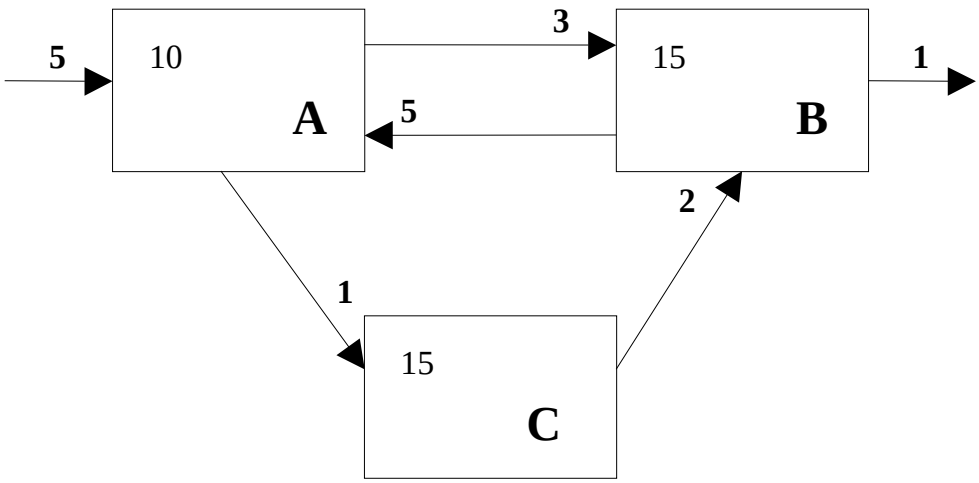
Tarkasteltaessa ajassa ilmeneviä tapahtumasarjoja tilastollisin menetelmin on tapana puhua tapahtumahistoria-analyysista. Tapahtumahistoria-analyysin viitekehys sisältää käyttökelpoisia menetelmiä monista eri tutkimusperinteistä ja tarjoaa siten mahdollisuudet erittäin monipuoliseen analyysiin (ks. esim. Allison 1984; Andersen ym. 1993; Bartholomew 1982; Blossfeld & Rohwer 1995; Cox & Oakes 1984; Elandt-Johnson & Johnson 1980; Janssen & Limnios 1999; Kalbfleisch & Prentice 1980; Yamaguchi 1991). Sovellusesimerkkejä löytyy niin yhteiskuntatieteistä (ks. esim. Barton & Pillai 1995; Box-Steffensmeier & Jones 1997; Prais 1955; Tuma ym. 1979), demografiasta (ks. esim. Courgeau & Lelièvre 1997; Trussel ym. 1992), taloustieteistä (ks. esim. Heckman & Singer 1984; Kiefer 1988; Lancaster 1992; Shorrocks 1978), epidemiologiasta (ks. esim. Moolgavkar & Prentice 1986), työvoiman suunnittelusta (ks. esim. Bartholomew ym. 1991), biometriasta (ks. esim. Clayton 1988), terveydenhuollon suunnittelusta (ks. esim. Davies ym. 1975; Kao 1972; 1974) kuin teollisuuden laadunvalvonnastakin (ks. esim. Ansell & Phillips 1989; Ouhbi & Limnios 1996).

## 2.1 Dynaaminen ilmiö mallinnettavana järjestelmänä

Järjestelmä määritellään joukoksi toisiinsa liittyviä asioita tai osia, jotka toimivat yhdessä tai ovat jonkinlaisessa yhteydessä siten, että niiden voidaan ajatella muodostavan eriteltävissä olevan kokonaisuuden. Tilastollinen malli määritellään tässä yhteydessä matemaattiseksi kuvaukseksi siitä, kuinka järjestelmässä tapahtuu muutoksia. Nämä määritelmät tarvitsevat tuekseen kuvauksen järjestelmään liittyvistä olioista, ilmiöistä ja toisaalta myös rajoituksista. Periaatteessa mallin muodostaminen voidaan tulkita prosessiksi, jossa hahmotetaan ongelmanasetteluun liittyvät ilmiöt mahdollisimman yksinkertaiseksi järjestelmäksi ja esitetään, kuinka järjestelmään liittyviä tekijöitä voidaan kuvata matemaattisesti. Puhtaasti tilastollista työtä vaativaa on siis vain jälkimmäinen osuus.

Varsin yleispätevä tapa kuvata yksilötason dynaamisiin ilmiöihin liittyvää järjestelmää on esittää se joukkona varantoja ja virtaamia. Tällöin järjestelmässä olevat yksilöt voidaan jokaisella ajanhetkellä luokitella ongelmanasetteluun sopivien tekijöiden perusteella ryhmiin. Yksilöiden määrää kussakin ryhmässä nimitetään varannoksi. Tiettyä ajanhetkeä kuvaavat varannot eivät kerro suoraan mitään ajan myötä tapahtuvista muutoksista, joten muutosten käsittelyyn tarvitaan virtaaman käsite. Virtaamat ilmaisevat niiden yksilöiden määrät, jotka siirtyvät jollain aikavälillä ryhmästä toiseen. On huomionarvoista, että varannot liittyvät aikapisteeseen ja virtaamat puolestaan aikaväliin (Bartholomew ym. 1991, 3-6). Demografiassa vastaavanlaista järjestelmää on ajateltu kenttänä, jossa on tiloja ja niiden välillä tapahtuvaa liikettä. Jokainen tilojen välinen siirtyminen eli transitio on muutostapahtuma järjestelmää kuvaavassa muutostapahtumakentässä (Hakkarainen 1989, 8-13; Hoem 1976).

Formaalisti järjestelmä voidaan esittää suunnattuna verkkona, jonka solmuilla ja kaarilla on tiloja ja transitioita kuvailevia ominaisuuksia (vrt. Butler & Huzurbazar 1997; Huzurbazar 1999). Kuviossa 1 on esimerkki varantoja ja virtaamia kuvaavan järjestelmän verkkoesityksestä. Laatikot kuvaavat solmuja eli tiloja tai ryhmiä ja nuolet puolestaan kaaria eli mahdollisia transitioita tai siirtymisiä. Vain nuolten suuntaiset siirtymät ovat mahdollisia. Numerot laatikoiden sisällä ilmaisevat, kuinka monta yksilöä kyseisessä




Kuvio 1. Kolmetilainen avoin järjestelmä

tilassa on tarkasteluajanjakson alussa ja nuolien yläpuoliset numerot puolestaan kuinka monta yksilöä tekee nuolen suuntaisen transition tarkasteluajanjakson aikana. Kaaret, jotka eivät lähde solmuista kuvaavat järjestelmän ulkopuolelta siihen kohdistuvaa liikettä ja vastaavasti solmuihin päättymättömät kaaret esittävät järjestelmän piiristä poistuvaa liikettä. Järjestelmä on suljettu, jos sen piiriin ei tule eikä siitä poistu yksilöitä; muutoin se on avoin.

Mallin muodostamiseksi on myös tunnettava siirtymiä aiheuttavat mekanismit, sillä ne on otettava huomioon mallin virtaamiin liittyvinä oletuksina. Tilastollisessa mallissa on periaatteessa tehtävä ero virtaamien deterministisyyden ja stokastisuuden sekä diskreetin ja jatkuvan ajan välillä. Deterministinen ja stokastinen malli antavat yleensä kuitenkin keskiarvon mielessä varsin yhteneviä tuloksia (Bartholomew ym. 1991, 7) ja diskreettejä malleja voidaan käyttää jatkuvien approksimaatioina sekä päinvastoin, joten järkevien tulosten tavoittelun puolesta erot ovat hiuksenhienoja. Haluttaessa liittää varantoihin ja virtaamiin ehtoja, esimerkiksi rajoituksia tilojen koolle, joudutaan menetelmällisesti monimutkaisempiin ratkaisuihin, jotka on tarkoituksellisesti rajattu tämän tutkimuksen ulkopuolelle (ks. esim. Bartholomew 1982, 149-185 & 212-247). On myös selvää, että järjestelmän piirissä olevien yksilöiden ominaisuudet vaikuttavat oleellisesti järjestelmässä ilmeneviin tapahtumiin. Periaatteessa heterogeenisuuden hallintaan olisi mahdollista soveltaa erilaisia vakiointi- ja regressiomenetelmiä (ks. esim. Ugwuowo

& McClean 2000), mutta sellaisten käyttö monimutkaistaisi tarkasteluja oleellisesti, joten tässä tutkimuksessa on säännönmukaisesti oletettu yksilöiden heterogeenisuudestä aiheutuvien ongelmien olevan kierrettävissä käyttämällä analyyseissa mahdollisimman homogeenisten yksilöiden ryhmiä (vrt. Bartholomew ym. 1991, 14-15).

## 2.2 Dynaamisen järjestelmän mallintamisen perusteita

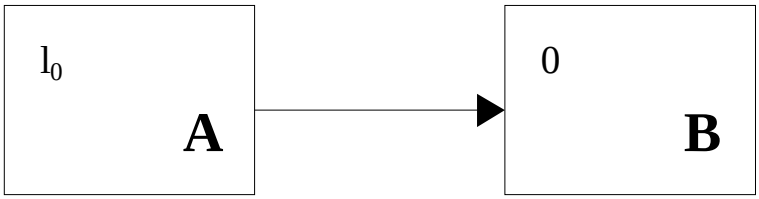
Järjestelmää mallinnettaessa on luontevinta tarkastella ensin siihen liittyviä käsitteitä ja menetelmiä perinteisiä - erityisesti demografiassa ja aktuaarimatematiikassa käytettyjä - lähestymistapoja soveltaen. Demografia eli väestötiede on tieteenala, jonka tehtävänä on tutkia ihmisten muodostamia populaatioita eli väestöjä – erityisesti niiden suuruutta, rakennetta (koostumusta) ja kehitystä sekä niiden kvantitatiivisesti mitattavissa olevia ominaisuuksia (Haimi 1987, 3).

Väestöntutkimuksen kohteena ovat väestöllisten tapahtumien ilmaantuvuus eli insidenssi (incidence) ja tapahtumien seurauksena syntyvä väestörakenne, vallitsevuus eli prevalenssi (prevalence). Ilmaantuvuuden mittaus tapahtuu pitkittäistutkimuksella (longitudinal research) ja vallitsevuuden puolestaan poikkileikkaustutkimuksella (cross-sectional research). Aineiston ominaisuuksista riippuen puhutaan myös periodi- (period/census analysis) ja kohorttitarkasteluista (cohort analysis). Kohortilla tarkoitetaan väestöryhmää, joka on kokenut jonkun väestöllisen ilmiön, esimerkiksi syntymän, suunnilleen samaan aikaan. Kohorttitarkastelun voi ymmärtää laajemmassa mielessä jonkun sopivasti rajatun ryhmän ominaisuuksien pitkittäistutkimukseksi. Perioditarkastelussa aineistossa on periaatteessa havaintoja lyhyemmältä ajanjaksolta kuin kohorttitarkastelusta, mutta koska jotkut ilmiöt saattavat käyttäytyä kohorttimaisesti jo lyhyelläkin ajanjaksolla, ei käsitteiden erottelu ole aina yksiselitteistä. Nykyisin rekisteripohjaiset aineistot ovat varsin kattavia ja voivat sisältää poikkileikkaustietoja pitkältäkin ajanjaksolta (vrt. luku 3), jolloin on kyse niin sanotuista paneeliaineistoista. Paneeliaineistoja käyttäen myös useiden rinnakkaisten kohorttien yhtäaikainen seuranta on mahdollista. Esitetyt käsitteet onkin ymmärrettävä joustavassa mielessä, sillä

usein päädytään menetelmällisiltä perusteiltaan yhteneviin ratkaisuihin, vaikka asioista puhuttaisiinkin eri nimillä. (Vrt. Haimi 1987, 31-36.)

Perusaineisto on yleensä joukko absoluuttisia lukuja, jotka liittyvät tiettyyn ajanhetkeen tai ajanjaksoon. Varannon määriteltiin olevan tietyssä tilassa tietyllä ajanhetkellä havaittavien yksilöiden lukumäärä, joten varantojen mittaamiseksi riittää laskea järjestelmän piirissä olevien yksilöiden lukumäärät kussakin tilassa halutulla ajanhetkellä suoraan perusaineiston absoluuttisista luvuista. Demografiassa ollaan kuitenkin useimmiten kiinnostuneita mittaamaan pelkkien absoluuttisten lukujen sijaan niiden suhdetta muihin tekijöihin. Esimerkiksi tietyssä tilassa olevien yksilöiden varannosta saadaan laskettua tilan vallitsevuus suhteuttamalla tilan varanto toiseen isompaan populaatioon, joka joko jo on kyseisessä tilassa tai jolla on mahdollisuus siirtyä siihen.

Toiseksi väestötieteen keskeiseksi mittauskohteeksi mainittiin ilmaantuvuus. Ilmaantuvuus ei ole vallitsevuuden tavoin osuus (proportion), vaan se kuvaa tapahtumien esiintymistaajuutta (rate) eli tapahtumien määrää suhteutettuna niin sanottuun väestöaikaan (vrt. Elandt-Johnson & Johnson 1980, 31-35). Väestö- eli riskiajalla tarkoitetaan väestön yhteenlaskettua elinaikaa tarkasteluvälillä eli sitä aikamäärää, jonka verran tapahtumalle ollaan tarkasteluvälillä yhteensä alttiina. Epidemiologiassa ilmaantuvuudella tarkoitetaan yleensä uusien tautitapausten esiintymistaajuutta, joten edellä kuvatun järjestelmän yhteydessä on havainnollisempaa puhua ilmaantuvuuden sijaan yleisemmin virtaamia mittaavista taajuusluvuista. Täsmällisemmin ilmaistuna niillä tarkoitetaan tietyllä aikavälillä tietystä tilasta toiseen tapahtuvien siirtymien lukumäärän suhdetta lähtötilassa oleville yksilöille kyseisessä aikavälissä kertyvään riski- eli viipymisaikaan. Lisäksi tietyllä aikavälillä tapahtuvien siirtymien absoluuttisista luvuista laskettua määrää kutsutaan volyymiksi. Useissa tapauksissa on luonnollista olettaa siirtymisen todennäköisyyden riippuvan lähtötilassa vietetystä ajasta, joten yksittäisen taajuusluvun tai volyymin sijaan on tarkoituksenmukaisempaa kuvata virtaaman ominaisuuksia lähtötilassa vietetystä ajasta riippuvan funktion avulla. Tällöin hetkittäinen taajuusluku (instantaneous rate) on luonnollista tulkita tapahtuman hetkittäiseksi intensiteetiksi (vrt. Keiding 1991).




Kuvio 2. Kaksitilainen suljettu järjestelmä

Periaatteessa järjestelmää pystytään kuvaamaan kattavasti, kun tunnetaan järjestelmän alkutilan varannot, kaikkiin mahdollisiin siirtymiin liittyvät taajuusluvut sekä järjestelmään ilmaantuvien uusien yksilöiden määrät. Väestötieteestä periytyvä menetelmä järjestelmän (tai väestön muutostapahtumakentän) tilojen mielivaltaisten yhdistelmien tarkasteluun on tapahtumataulun (life table) käyttö. Tapahtumataulun käsitteitä ja funktioita hyödyntäen pystytään kuvaamaan järjestelmään liittyvät ilmiöt ja niiden yhteydet toisiinsa formaalilla tavalla.

### 2.2.1 Tapahtumataulu

Oletetaan yksinkertaisuuden vuoksi, että kuvion 2 kaltaisessa järjestelmässä on tarkastelun alkamishetkellä $l_0$ yksilöä tilassa A ja nolla yksilöä tilassa B. Oletetaan lisäksi, että järjestelmään ei tule uusia yksilöitä ja että kaikki tilan A yksilöt siirtyvät äärellisessä ajassa tilaan B. Olkoon $l_x$ siirtymisen välttäneiden yksilöiden lukumäärä ajanhetkellä $x$ $(x \geq 0)$. Luku $l_x$ on tarkoituksenmukainen vain, jos myös luku $l_0$ tunnetaan. Merkitään $P_x$:llä muutostapahtuman välttäneiden osuutta ajanhetkellä $x$ eli $P_x = l_x/l_0$. Muutosten odotettu lukumäärä aikavälillä $(x, x+1)$ on

$$d_x = l_x - l_{x+1}. \tag{1}$$

Tällöin

$$q_x = \frac{d_x}{l_x} = \frac{l_x - l_{x+1}}{l_x} \tag{2}$$

on muutostapahtuman todennäköisyys aikavälillä $(x, x+1)$ sillä ehdolla, että yksilö on tilassa A hetkellä $x$. Ehdollinen todennäköisyys sille, että hetkellä $x$ tilassa A oleva yksilö

ei koe muutostapahtumaa aikavälillä $(x, x+1)$ on

$$p_x = 1 - q_x = \frac{l_{x+1}}{l_x}. \tag{3}$$

Edelleen

$$l_x = l_{x-1}p_{x-1} = l_0 p_0 p_1 ... p_{x-1} \tag{4}$$

eli

$$P_x = \frac{l_x}{l_0} = p_0 p_1 p_2 ... p_{x-1}. \tag{5}$$

Lisäksi

$$L_x = \int_x^{x+1} l_y dy = \int_0^1 l_{x+t} dt \tag{6}$$

ilmaisee, kuinka kauan tilassa A ajanhetkellä $x$ olevat yksilöt yhteensä odottavat muutostapahtumaa aikavälillä $(x, x+1)$ ja

$$T_x = \int_0^\infty l_{x+t} dt = L_x + T_{x+1} \tag{7}$$

on ajankohdasta $x$ alkaen tilassa A kertyvä odotusaikojen yhteismäärä. Näin ollen

$$e_x = \frac{T_x}{l_x} = \frac{1}{l_x} \sum_{i=0}^{\infty} L_{x+i} \tag{8}$$

on keskimääräinen aika, jonka yksilö vielä viipyy tilassa A olleessaan siellä ajanhetkellä $x$. (Vrt. Elandt-Johnson & Johnson 1980, 83-93; Hakkarainen 1989, 11-12; Pollard 1973, 3-6.)

On merkillepantavaa, että $l_x$ ja $P_x$ liittyvät täsmälleen ajanhetkeen $x$; $q_x$, $p_x$, $d_x$ ja $L_x$ liittyvät aikaväliin $(x, x+1)$, kun taas $T_x$ ja $e_x$ liittyvät väliin $[x, \infty)$. Vaikka tapahtumataulut on perinteisesti konstruoitu käyttämättä todennäköisyysteoriaa, voidaan $P_x$ teoreettisessa mielessä tulkita jatkuvaksi funktioksi, jolloin se on analoginen elinaika-analyysin eloonjäämisfunktion kanssa. Seuraavissa tarkasteluissa $X$ oletetaan jatkuvaksi satunnaismuuttujaksi.

### 2.2.2 Eloonjäämisfunktiot ja niiden väliset yhteydet

Eloonjäämisfuntio kuvastaa muutostapahtuman välttämisen todennäköisyyttä eli on todennäköisyys sille, että tapahtuma-aikaa edustava satunnaismuuttuja $X$ on suurempi kuin $x$ eli

$$P_x = S(x) = P(X > x). \tag{9}$$

Yleensä $X$ ei voi olla negatiivinen, joten $S(0) = 1$ ja $S$ on vähenevä funktio sekä $\lim_{x\to\infty} S(x) = 0$. Edelleen

$$S(x) = 1 - F(x) = 1 - P(X \leq x) = 1 - \int_{-\infty}^{x} f(u)du = \int_{x}^{\infty} f(u)du, \tag{10}$$

jossa

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x} = \frac{dF(x)}{dx} = -\frac{dS(x)}{dx} \tag{11}$$

on tiheysfunktio.

Muutostapahtumaan kuluvan ajan jakaumaa voidaan kuvata eloonjäämisfunktion sijaan yksikäsitteisesti myös hasardifunktion avulla. Hasardifunktio kuvaa hetkittäistä muutostapahtuman taajuutta pisteessä $x$ eli

$$h(x) = \lim_{\Delta x \to 0} \frac{P(x \leq X < x + \Delta x \mid X \geq x)}{\Delta x}. \tag{12}$$

Näin ollen suure $h(x)\Delta x$ on approksimatiivinen todennäköisyys sille, että iän $x$ saavuttanut yksilö kokee muutostapahtuman välillä $(x, x + \Delta x)$. Hasardifunktio voidaan esittää myös seuraavissa yhtäpitävissä muodoissa

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{f(x)}{S(x)} = -\frac{d \log S(x)}{dx}. \tag{13}$$

Lisäksi integroimalla saadaan seuraava yhteys eloonjäämis- ja hasardifunktioiden välille:

$$S(x) = e^{-H(x)}, \tag{14}$$

jossa

$$H(x) = \int_{0}^{x} h(u)du = -\log S(x) \tag{15}$$

on kumulatiivinen hasardifunktio.

Tapahtumataulun ja elinaika-analyysin funktioiden merkintöjä yhdistellen hasardifunktio eli kuolevuuden voima (force of mortality) saadaan muotoon:

$$h(x) = \mu_x = \frac{-\frac{dl_x/l_0}{dx}}{l_x/l_0} = -\frac{1}{l_x}\frac{dl_x}{dx} = -\frac{d\log l_x}{dx}. \tag{16}$$

Tiheysfunktio on puolestaan muotoa

$$f(x) = -\frac{dl_x}{dx} = \mu_x l_x \tag{17}$$

ja edelleen

$$l_x = l_0 S(x) = \int_0^\infty \mu_{x+t} l_{x+t} dt, \tag{18}$$

$$d_x = l_0 \int_x^{x+1} f(y)dy = \int_0^1 \mu_{x+t} l_{x+t} dt, \tag{19}$$

$$\begin{aligned} q_x &= \frac{d_x}{l_x} = P(x \leq X < x+1 \mid X \geq x) \\ &= \frac{\int_x^{x+1} f(y)dy}{\int_x^\infty f(y)dy} = \frac{\int_x^{x+1} h(y)S(y)dy}{\int_x^\infty h(y)S(y)dy} = \frac{S(x)-S(x+1)}{S(x)} \end{aligned} \tag{20}$$

sekä

$$\begin{aligned} e_x &= r(x) = E(X-x \mid X \geq x) = \int_0^\infty t\frac{f(x+t)}{S(x)}dt \\ &= \frac{\int_x^\infty (t-x)f(t)dt}{S(x)} = \frac{\int_x^\infty tf(t)dt}{S(x)} - x = \frac{\int_x^\infty S(t)dt}{S(x)} \\ &= \frac{1}{l_x}\int_0^\infty t\mu_{x+t} l_{x+t} dt = -\frac{1}{l_x}\int_0^\infty t\frac{dl_{x+t}}{dt}dt. \end{aligned} \tag{21}$$

(Vrt. Elandt-Johnson & Johnson 1980, 50-57 & 93-95; Kalbleisch & Prentice 1980, 6-7; Klein 1998.)

### 2.2.3 Ei-parametrisesta eloonjäämisfunktioiden estimoinnista

Tapahtumataulujen funktioiden arvot on perinteisesti laskettu empiirisesti helposti ymmärrettävillä ja sovellettavilla tavoilla ilman oletuksia taustalla olevista parametrisista jakaumista (ks. esim. Elandt-Johnson & Johnson 1980, 150-180; Singpurwalla & Wong 1983). Analyyseissa joudutaan kuitenkin hyvin usein kohtaamaan aineiston puutteista

johtuvia erityisongelmia, sillä muutostapahtuma ei monista eri syistä johtuen välttämättä ehdi tapahtua tarkasteluajanjakson aikana kaikille aineiston yksilöille tai toisaalta muutostapahtuman odotusajan tarkkaa alkamisajankohtaa ei ehkä tunneta (Cox & Oakes 1984, 2-5). Puutteellisia havaintoja kutsutaan vastaavasti sensuroiduiksi (censored) ja katkaistuiksi (truncated) havainnoiksi. Puutteet heijastuvat suoraan tarvittaviin estimointimenetelmiin, joten aineiston ominaisuudet määräävät pitkälti käytettävät menetelmät (vrt. Little 1992).

Käyttökelpoisimpia eloonjäämisfunktion ei-parametrisia estimaattoreita on tuloraja- (product-limit) eli Kaplan-Meier-estimaattori (Kaplan & Meier 1958). Kaplan-Meier-estimaattori on empiirisen kertymäfunktion yleistys, jossa oikeanpuoleinen sensurointi otetaan huomioon. Voidaan lisäksi osoittaa, että samaa muotoa oleva eloonjäämisfunktion estimaattori on mahdollista johtaa myös silloin, kun aineistossa on mukana vasemmalta katkaistuja havaintoja (ks. esim. McClean & Gribbin 1987). Kaplan-Meier-estimaattori saadaan johdettua esimerkiksi seuraavalla tavalla:

Olkoot $t_1 < t_2 < ... < t_k$ erilliset muutostapahtumien havaitut tapahtuma-ajat $n$:n yksilön kokoisessa homogeenisessa aineistossa ($k \leq n$). Olkoon $n_i$ niiden yksilöiden määrä, joille muutostapahtuma on mahdollinen juuri ennen hetkeä $t_i$ ($1 \leq i \leq k$) ja $c_i$ tapahtumien määrä hetkellä $t_i$. Ajatellaan aineistoon liittyvää eloonjäämisfunktiota diskreettinä funktiona, jolla on todennäköisyysmassaa jokaisessa pisteessä $t_i$. Tällöin pistetodennäköisyysfunktio on muotoa

$$f(t_i) = h(t_i) \prod_{j=1}^{i-1} (1 - h(t_j)) \tag{22}$$

ja eloonjäämisfunktio vastaavasti muotoa

$$S(t_i) = \prod_{j=1}^{i} (1 - h(t_j)), \tag{23}$$

jossa $h(\cdot)$ on ehdollinen pistetodennäköisyysfunktio (diskreetti hasardifunktio), joka ilmaisee, millä todennäköisyydellä muutostapahtuma koetaan pisteessä $t_i$, jos se on sitä ennen vältetty. Ehdollinen todennäköisyys sille, että $c_i$ kappaletta $n_i$:stä muutostapahtumalle alttiista olevasta yksilöstä kokee muutostapahtuman hetkellä $t_i$,

voidaan esittää muodossa

$$L_i = \frac{[f(t_i)]^{c_i}[S(t_i)]^{n_i-c_i}}{[S(t_i^-)]^{n_i}} = h(t_i)^{c_i}(1-h(t_i))^{n_i-c_i}, \tag{24}$$

jossa $t_i^-$ on ajanhetki juuri ennen $t_i$:tä. Edelleen uskottavuusfunktioksi saadaan

$$L = \prod_{i=1}^{k} h(t_i)^{c_i}(1-h(t_i))^{n_i-c_i} \tag{25}$$

ja näin ollen log-uskottavuusfunktio on muotoa

$$l = \sum_{i=1}^{k} c_i \log h(t_i) + \sum_{i=1}^{k}(n_i - c_i)\log(1-h(t_i)), \tag{26}$$

josta saadaan ei-parametriseksi suurimman uskottavuuden estimaattoriksi

$$\widehat{h}(t_i) = \frac{\text{(muutostapahtumien havaittu määrä } t_i\text{:ssä)}}{\text{(muutostapahtumalle alttiiden määrä } t_i\text{:ssä)}} = \frac{c_i}{n_i}. \tag{27}$$

Eloonjäämisfunktio voidaan saadun tuloksen perusteella esittää muodossa

$$\widehat{S}(t) = \prod_{i|t_i \le t}(1 - \frac{c_i}{n_i}), \tag{28}$$

ja jos kaikissa pisteissä $t_i$ on korkeintaan yksi havaittu muutostapahtuma, edelleen muodossa

$$\widehat{S}(t) = \prod_{i|t_i \le t}(\frac{n-i}{n-i+1})^{b_i}, \tag{29}$$

jossa $n$ on havaintojen lukumäärä ja $b_i$ on 1, jos $i$:s havainto on järjestetyssä aineistossa sensuroimaton ja muuten nolla.

Koska $\widehat{h}(t_i)$ voidaan tulkita suhteelliseksi frekvenssiksi saadaan

$$\text{var}(\widehat{h}(t_i)) = \frac{\widehat{h}(t_i)(1-\widehat{h}(t_i))}{n_i} \tag{30}$$

ja tällöin $\log \widehat{S}(t) = \sum_{i|t_i \le t} \log(1-\widehat{h}(t_i))$:llä on asymptoottinen varianssiestimaattori

$$\widehat{\text{var}}[\log \widehat{S}(t)] = \sum_{i|t_i \le t}(1-\widehat{h}(t_i))^2 \text{var}(1-\widehat{h}(t_i)) = \sum_{i|t_i \le t} \frac{c_i}{n_i(n_i-c_i)} \tag{31}$$

eli Greenwoodin kaavana tunnettu asymptoottinen varianssiestimaattori $\widehat{S}(t)$:lle on

$$\widehat{\text{var}}[\widehat{S}(t)] = \widehat{S}^2(t) \sum_{i|t_i \le t} \frac{c_i}{n_i(n_i-c_i)}. \tag{32}$$

Tämän kaavan suora soveltaminen isoille ja pienille t:n arvoille saattaa johtaa luottamusrajoihin, jotka eivät pysy välissä $[0, 1]$. Tämä ongelma voidaan välttää tekemällä muunnos $\log[-\log \widehat{S}(t)]$, joka voi saada mitä tahansa arvoja. Tällöin asymptoottinen hajonta on muotoa

$$\widehat{s}(t) = \frac{\sqrt{\sum_{i|t_i \leq t} \frac{c_i}{n_i(n_i - c_i)}}}{\sum_{i|t_i \leq t} \log[\frac{c_i}{n_i(n_i - c_i)}]} \tag{33}$$

ja 95%:n luottamusväliksi $\widehat{S}(t)$:lle saadaan $\widehat{S}(t)^{e^{\pm 1.96 \widehat{s}(t)}}$. Hajonnat voidaan laskea vaihtoehtoisesti myös bootstrap-tekniikkaa käyttäen (ks. esim. Akritas 1986; Efron 1981).

Usein on tarpeellista esittää eloonjäämisfunktion keskeisiä ominaisuuksia tiiviissä muodossa sopivia tunnuslukuja käyttäen. Esimerkiksi muutostapahtumaan kuluvan ajan mediaani ja muut fraktiilit voidaan määritellä yhteyden

$$\frac{S(y_p)}{S(x)} = p \tag{34}$$

kautta, jossa $x$ on tarkasteluajanhetki ja $y_p$ ajanhetki, johon on $100p$ %:n mahdollisuus välttää muutostapahtuma. Mediaanille saadaan tarvittaessa määrättyä myös luottamusrajat (ks. esim. Brookmeyer & Crowley 1982; Reid 1981).

Keskimääräisen muutostapahtumaan kuluvan ajan voidaan osoittaa olevan yhtenevä eloonjäämisfunktion alla olevaan pinta-alaan (vrt. kaavan (21) toisen rivin viimeinen muoto pisteessä nolla). Sensuroidussa tapauksessa estimaattori keskimääräiselle ajalle, joka vielä viivytään tilassa, jossa ollaan hetkellä $x$, saadaan sijoittamalla tulokset (22) ja (28) kaavaan (21), jolloin estimaattori tulee muotoon

$$\widehat{r}(x) = \frac{\sum_{i|t_i \geq x} \{t_i (\prod_{j|t_j \leq t_i} (1 - \frac{c_j}{n_j})) \frac{c_j}{n_j}\}}{\prod_{i|t_i \leq x} (1 - \frac{c_i}{n_i})} - x. \tag{35}$$

Toinen vaihtoehto keskimääräisen viipymisajan estimointiin sensuroidussa tapauksessa on yksinkertaisesti rajoittaa maksimaalinen viipymisaika sopivaksi luvuksi, jonka jälkeen sensuroinnin takia tuntemattomaksi jäänyt häntätodennäköisyys ei enää vedä estimaattoria kohti ääretöntä. Tällöin keskimääräiselle viipymisajalle voidaan laskea myös hajonta (ks. esim. Kaplan & Meier 1958). Keskimääräisen viipymisajan estimaatti

on yleensä varsin herkkä valitulle maksimiajalle, joten estimaattiin on suhtauduttava riittävällä varauksella.

### 2.2.4 Eloonjäämisfunktioiden vertailuista

Kun eri ryhmille ja tiloille on estimoitu eloonjäämis- tai hasardifunktioita, on yleensä kiinnostavaa verrata niitä toisiinsa. Periaatteessa samanlaisia ongelmia kohdataan myös silloin, kun halutaan verrata estimoitua jakaumaa johonkin teoreettiseen malliin. Intuitiivisesti selkeää ja eksploratiiviseen analyysiin hyvin soveltuvaa on käyttää vertailuun graafisia tarkasteluja. Tällöin voidaan yksinkertaisesti piirtää eri ryhmien eloonjäämis- tai hasardifunktioita samaan kuvaan ja vertailla niiden ominaisuuksia. Estimaattoreille kannattaa laskea luottamusvälejä, jotta todelliset erot olisi helpompi havaita. Myös jakaumaa kuvaavat tunnusluvut, kuten keskiarvo ja mediaani, voivat antaa hyödyllistä tukea silmämääräisille tarkasteluille.

Jos aineistossa ei ole puutteellisia havaintoja, eloonjäämisfuktioiden vertailuun on mahdollista käyttää tavallisia ei-parametrisia testejä. Sensuroituja havaintoja sisältävälle aineistolle voidaan käyttää ei-parametrista Tarone-Ware-testiperhettä (Tarone & Ware 1977). Tämän testiperheen jäsenet palautuvat Wilcoxonin tai Kruskal-Wallisin testeihin aineistoissa, joissa ei ole sensuroituja havaintoja. Testiperhe voidaan johtaa suoraan $k$:n ryhmän samanaikaiseen vertailuun soveltuvaksi, jolloin kahden ryhmän vertailuun sopivat testisuureet saadaan $k$:n arvolla kaksi.

Oletetaan, että erilliset muutostapahtumien ajankohdat on järjestetty eli $t_1 < t_2 < \cdots < t_m$, jolloin hetkellä $t_i$ aineistosta voidaan muodostaa seuraavanlainen $2 \times k$-taulukko,

| | Ryhmä | | | | | | |
|---|---|---|---|---|---|---|---|
| | **1** | **2** | **...** | **j** | **...** | **k** | **Yhteensä** |
| Tapahtumat | $c_{1i}$ | $c_{2i}$ | ... | $c_{ji}$ | ... | $c_{ki}$ | $c_i$ |
| Alttiit | $n_{1i}$ | $n_{2i}$ | ... | $n_{ji}$ | ... | $n_{ki}$ | $n_i$ |

jossa $n_{ji}$ on muutostapahtumalle alttiiden yksilöiden määrä ja $c_{ji}$ muutostapahtumien määrä ryhmässä $j$ hetkellä $t_i$. Nollahypoteesin ($H_0 =$ ”yhtenevät eloonjäämisfuktiot”) voimassaollessa saadaan epäsuoralla vakioinnilla $c_{ji}$:n odotusarvoksi $E(c_{ji}) = \frac{n_{ji}c_i}{n_i}$.

Tällöin näyttö nollahypoteesia vastaan ryhmässä $j$ saadaan muotoon

$$S_{wj} = \sum_{i=1}^{m} w_i[c_{ji} - E(c_{ji})], \tag{36}$$

jonka odotusarvo on nollahypoteesin pätiessä nolla ja johon liittyvät kovarianssit ovat muotoa

$$V_{wjk} = \sum_{i=1}^{m} w_i^2 \frac{c_i(n_i - c_i)n_{ji}}{(n_i - 1)n_i}(\delta_{jk} - \frac{n_{ki}}{n_i}), \tag{37}$$

jossa $\delta_{jk}$ on $1$, jos $j = k$ ja muuten nolla. Määritellään, että vektori $\mathbf{S}'_w = (S_{w1}, ..., S_{wj})$ ja että $\mathbf{V}_w$ on $r \times r$-matriisi, jonka $(j, k)$:s alkio on muotoa $V_{wjk}$. Nyt testisuure

$$\chi_w^2 = \mathbf{S}'\mathbf{V}_w^{-1}\mathbf{S} \tag{38}$$

on nollahypoteesin pätiessä jakautunut kuten $\chi^2$-jakauma $(k - 1)$:llä vapausasteella, jos painofunktio on valittu järkevällä tavalla. Esimerkiksi valinnalla $w_i = n_i$ saadaan yleistetty Kruskal-Wallis-testi (toiselta nimeltään Gehan-Breslow-testi), valinnalla $w_i = 1$ log-rank-testi (Cox-Mantel-testi) ja valinnalla $w_i = \sqrt{n_i}$ Tarone-Ware-testi.

Yleistetyssä Kruskal-Wallis-testissä (kahden ryhmän tapauksessa puhutaan yleistetystä Wilcoxonin testistä) ”alkupään” havainnot saavat enemmän painoa ja näin ollen testi sopii erityisesti lyhyiden viipymisaikojen poikkeavuuden tutkimiseen. Log-rank-testi puolestaan painottaa jokaista havaintoa samalla tavalla ja on sen takia herkempi paljastamaan poikkeamia myös pitkien viipymisaikojen osalta. Siten testejä voidaan käyttää näiden ominaisuuksien perusteella pelkän poikkeamien todentamisen lisäksi myös poikkeamien luonteen määrittämiseen.

Testiperheen määritelmästä (36) havaitaan, että ne perustuvat suoraan erojen summaamiseen eivätkä itseisarvoihin tai neliösummiin. Näin ollen testejä on tarkoituksenmukaista käyttää vain, jos erot ovat jatkuvasti samansuuntaisia, sillä muutoin negatiiviset ja positiiviset poikkemat kumovat helposti toisensa. (Vrt. Breslow 1970; Cox 1972; Gehan 1965a; 1965b; Le 1997, 98-109; Tarone & Ware 1977.)

## 2.3 Markov-malleista

Yksilöt voivat liikkua dynaamisessa järjestelmässä annettujen oletusten puitteissa mielivaltaisesti ja näin ollen yksittäisten tilojen mallintamisen sijaan järjestelmää pitää pystyä käsittelemään ja mallintamaan kokonaisuutena. Aikaisemmin todettiin, että järjestelmiä pystytään varsin yleispätevästi kuvaamaan eri tiloihin liittyvien varantojen ja virtaamien avulla. Oletetaan, että järjestelmä pystytään kuvaamaan käyttämällä $k$:ta tilaa ja merkitään tilan $i$ varantoa hetkellä $T$ $n_i(T)$:llä. Kaikkien tilojen varannot hetkellä $T$ voidaan esittää yhtenä (rivi)vektorina muodossa

$$\mathbf{n}(T) = (n_1(T), n_2(T), ..., n_k(T)).$$

Pelkkä hetkittäisten varantojen kuvaaminen ei riitä järjestelmän mallintamiseen, joten formalisoidaan virtaamien kuvaaminen olettamalla, että jokaisella yksilöllä on ajasta riippumaton, tunnettu todennäköisyys tehdä mikä tahansa siirtymä. Yksilön siirtymätodennäköisyydet eri tilojen välillä voidaan esittää matriisina $\mathbf{P}$, jossa jokainen alkio $p_{ij}$ on todennäköisyys, että tarkasteluajanjakson alussa tilassa $i$ oleva yksilö on tarkasteluajanjakson lopussa tilassa $j$.

Niin sanotussa avoimessa järjestelmässä, jossa järjestelmään tulevat ja siitä poistuvat yksilöt pitää ottaa huomioon, kannattaa varata yksi tila kuvaamaan järjestelmästä poistuneita yksilöitä, jolloin matriisi $\mathbf{P}$ saadaan ositettua seuraavalla tavalla:

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{S} & \mathbf{w}' \\ \hline \mathbf{0} & 1 \end{array}\right] = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1,k-1} & w_1 \\ s_{21} & s_{22} & \cdots & s_{2,k-1} & w_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{k-1,1} & s_{k-1,2} & \cdots & s_{k-1,k-1} & w_{k-1} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix},$$

jossa $s_i$ on todennäköisyys siirtyä tilasta $i$ tilaan $j$ ja $w_i$ on todennäköisyys, että tarkasteluajanjakson alussa tilassa $i$ ollut yksilö on tarkasteluajanjakson lopussa poistunut järjestelmän piiristä eli siirtynyt tilaan $k$. Koska jokaisen yksilön on joko pysyttävä tilassa missä on, siirryttävä toiseen tilaan tai poistuttava järjestelmästä, rivisummat ovat $\sum_{j=1}^{k-1} s_{ij} + w_i = 1$ kaikilla $i = (1, ..., k-1)$. Tila $k$ kuvaa järjestelmästä poistuneita, jotka eivät palaa sen piiriin uudestaan eli todennäköisyys siirtyä tilasta $k$ johonkin toiseen tilaan

on $0$ ja todennäköisyys pysyä tilassa $k$ on $1$. Järjestelmästä poistuneita yksilöitä kuvaavia tiloja voi olla useampiakin, jos halutaan eritellä poistumissyitä. Järjestelmän ulkopuolelta saapuvia yksilöitä ei tarvitse ottaa huomioon matriisin $\mathbf{P}$ muotoilussa. Oletetaan sen sijaan, että hetkellä $T$ saapuvien yksilöiden yhteismäärä $R(T)$ noudattaa tunnettua jakaumaa ja että saapuneet yksilöt joutuvat eri tiloihin todennäköisyyksillä $r_1, r_2, ..., r_k$ $(\sum_{j=1}^{k} r_i = 1)$. Merkitään tätä niin sanottua täydennystodennäköisyysvektoria $\mathbf{r} = (r_1, r_2, ..., r_k)$:llä.

Matriisi $\mathbf{P}$ on ei-negatiivisia elementtejä sisältävä $k \times k$-neliömatriisi, jonka rivit summautuvat ykkösiksi. Tällaista matriisia kutsutaan stokastiseksi matriisiksi. Määritellään edelleen $k$-alkioinen rivivektori $\mathbf{p}(0)$, jonka alkiot ovat muotoa $p_i(0) = n_i(0)/N(0)$, jossa $N(0) = \sum_{j=1}^{k} n_j(0)$ kaikilla $i = 1, 2, ..., k$. Näin ollen $p_i(0)$ kertoo, kuinka suuri osuus järjestelmässä olevista yksilöistä on tilassa $i$ $(i = 1, 2, ..., k)$ ajanhetkellä $0$. Nyt siirtymämatriisi $\mathbf{P}$ ja alkutilaa kuvaava vektori $\mathbf{p}(0)$ määrittelevät niin sanotun Markovin ketjun.

### 2.3.1 Markovin ketjut

Markovin ketjun voidaan tulkita olevan stokastinen prosessi eli sarja toisistaan riippuvia satunnaismuuttujia $\{X_n, n = 0, 1, 2, ...\}$, jossa kaikille tiloille $i_0, i_1, ..., i_{n-1}, i, j$ ja kaikille $n \geq 0$

$$\begin{aligned} p_{ij} &= P\{X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}, X_n = i\} \\ &= P\{X_{n+1} = j \mid X_n = i\}. \end{aligned} \tag{39}$$

Yleisesti määritelmän (39) täyttäviä stokastisia prosesseja kutsutaan Markov-prosesseiksi. Määritelmän mukaan Markov-prosesseissa tulevan tapahtuman todennäköisyys riippuu vain siitä tilasta, jossa ollaan, eli siirtymisen todennäköisyys on riippumatonta prosessin aikaisemmasta historiasta. Jos siirtymätodennäköisyys $P\{X_{n+1} = j \mid X_n = i\}$ on lisäksi riippumaton $n$:stä, sanotaan siirtymätodennäköisyyksien olevan aika-homogeenisia. Seuraavassa rajoitutaan yksinkertaisuuden vuoksi tarkastelemaan aika-homogeenisia Markov-prosesseja, vaikka osa tuloksista yleistyy ei-homogeenisiinkin tapauksiin (ks. esim. Vassiliou 1998).

Oletetaan edellä mainittu alkutilaa kuvaava vektori $\mathbf{p}(0)$ ja siirtymämatriisi $\mathbf{P}$ tunnetuiksi. Merkitään $\mathbf{p}(0)$:n alkioita $p_i(0)$:lla ja $\mathbf{P}$:n alkioita $p_{ij}$:llä $(i = 1, ..., k;$ $j = 1, ..., k)$. Nyt

$$\begin{aligned}
& P\{X_0 = i_0, X_1 = i_1, ..., X_n = i_n\} \\
= \; & P\{X_n = i_n \mid X_0 = i_0, ..., X_{n-1} = i_{n-1}\} P\{X_0 = i_0, ..., X_{n-1} = i_{n-1}\} \\
= \; & p_{i_{n-1} i_n} P\{X_0 = i_0, ..., X_{n-1} = i_{n-1}\} \\
= \; & p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} P\{X_0 = i_0, ..., X_{n-2} = i_{n-2}\} \\
& \vdots \\
= \; & p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \cdots p_{i_0 i_1} P\{X_0 = i_0\} \\
= \; & p_{i_{n-1} i_n} p_{i_{n-2} i_{n-1}} \cdots p_{i_0 i_1} p_{i_0}(0)
\end{aligned}$$

eli prosessi on täysin määrätty, kun alkutila ja siirtymätodennäköisyydet tunnetaan.

Olkoon $p_{ij}^n$ todennäköisyys, että prosessi liikkuu tilasta $i$ tilaan $j$ täsmälleen $n$:llä askeleella eli formaalisti

$$p_{ij}^n = P\{X_{n+m} = j \mid X_m = i\}.$$

Tämä ehdollinen todennäköisyys on siis kaikkien $n$:n pituisten tilasta $i$ tilaan $j$ johtavien ”polkujen” todennäköisyyksien summa. Erityisesti $p_{ij}^1 = p_{ij}$ ja $p_{ij}^2 = \sum_k p_{ik} p_{kj}$. Edelleen kaikille $r \leq n$ pätee niin sanottu Chapman-Kolmogorov-yhtälö

$$p_{ij}^n = \sum_k p_{ik}^r p_{kj}^{n-r}, \tag{40}$$

sillä prosessin on oltava jossain tilassa hetkellä $r$ ja tällöin ketjusäännön mukaisesti

$$\begin{aligned}
p_{ij}^n &= P\{X_n = j \mid X_0 = i\} \\
&= \sum_k P\{X_n = j, X_r = k \mid X_0 = i\} \\
&= \sum_k P\{X_n = j \mid X_r = k, X_0 = i\} P\{X_r = k \mid X_0 = i\} \\
&= \sum_k p_{kj}^{n-r} p_{ik}^r.
\end{aligned}$$

Jos $\mathbf{P}^n$:llä merkitään $n$:n askeleen siirtymätodennäköisyyksien $p_{ij}^n$ muodostamaa matriisia, niin yhtälön (40) mukaan $\mathbf{P}^n = \mathbf{P}^r \mathbf{P}^{n-r}$. Soveltamalla tulosta toistuvasti

todetaan, että $\mathbf{P}^n$ saadaan kertomalla yhden askeleen siirtymämatriisi $\mathbf{P}$ itsellään $n$ kertaa.

Edellä $\mathbf{p}(0)$:lla merkittiin yksilöiden osuuksia eri tiloissa ajanhetkellä $0$. Merkitään nyt yleisemmin $\mathbf{p}(T) = (p_1(T), ..., p_k(T))$:llä yksilöiden (odotettuja) osuuksia eri tiloissa hetkellä $T(T = 0, 1, ...)$. Koska osuudet summautuvat ykköseksi, ne voidaan tulkita myös todennäköisyyksiksi olla kyseisissä tiloissa. Todennäköisyys olla tilassa $j$ hetkellä $1$, kun ollaan tilassa $i$ hetkellä $0$, voidaan esittää muodossa

$$P\{X_0 = i, X_1 = j\} = P\{X_1 = j \mid X_0 = i\}P\{X_0 = i\},$$

jolloin kokonaistodennäköisyyden kaavaa soveltamalla

$$P\{X_1 = j\} = \sum_i P\{X_1 = j \mid X_0 = i\}P\{X_0 = i\}$$

eli toisilla merkinnöillä saadaan yhteys

$$p_j(1) = \sum_i p_{ij}p_i(0). \tag{41}$$

Lisäksi soveltamalla yhtälöä (40) ja vaihtamalla summausjärjestystä

$$\begin{aligned} p_j(T+1) &= \sum_i p_{ij}^{T+1}p_i(0) = \sum_i \left( p_i(0) \sum_k p_{ik}^T p_{kj} \right) \\ &= \sum_k \left( p_{kj} \sum_i p_{ik}^T p_i(0) \right) = \sum_k p_{kj}p_k(T), \end{aligned} \tag{42}$$

joten tuloksista (41) ja (42) seuraa, että kaikilla $T$ $(T = 0, 1, ...)$

$$p_j(T+1) = \sum_i p_{ij}p_i(T) \tag{43}$$

eli matriisimuodossa:

$$\mathbf{p}(T+1) = \mathbf{p}(T)\mathbf{P}. \tag{44}$$

Käyttämällä yhteyttä (44) toistuvasti yhtälö saadaan muotoon

$$\mathbf{p}(T) = \mathbf{p}(0)\mathbf{P}^T \tag{45}$$

eli komponenttimuodossa:

$$p_j(T) = \sum_i p_{ij}^T p_i(0). \tag{46}$$

Markovin ketjujen yleisessä teoriassa (ks. esim. Feller 1968, 372-427; Ross 1970, 61-84) suuri osa kiinnostuksesta kohdistuu ketjujen rakenteen ja ominaisuuksien määrittämiseen siirtymämatriisia $\mathbf{P}$ käyttäen. Edellä esitettyyn järjestelmään liittyvä siirtymämatriisi oli mahdollista osittaa seuraavalla tavalla:

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{S} & \mathbf{w}' \\ \hline \mathbf{0} & 1 \end{array}\right], \tag{47}$$

jossa $\mathbf{w} = (w_1, ..., w_k)$. Osituksesta nähdään, että viimeisessä, järjestelmästä poistuneita kuvaavassa tilassa, siirtymisiä voi tapahtua vain samaan tilaan eli tila on absorboiva. Oletetaan, että $\mathbf{S}$ on muotoa, jossa tilojen välillä on jonkinlainen yhteys, ja että ainakin yksi $\mathbf{w}$:n alkio on nollasta poikkeava. Tällöin Markovin ketjujen teorian kielenkäytössä $\mathbf{P}$:n kaikki tilat viimeistä lukuunottamatta ovat transientteja ja viimeinen tila on absorboivana myös toistuva (recurrent). Tällaisessa absorboivan Markovin ketjun tapauksessa suurin osa kiinnostavasta informaatiosta saadaan pääteltyä niin sanotusta fundamentaalimatriisista $(\mathbf{I} - \mathbf{S})^{-1}$ (ks. esim. Bartholomew 1982, 61-66).

Mallin konkreettisessa soveltamisessa on kiinnostavampaa tarkastella todennäköisyyksien sijaan varantoja eri ajanhetkillä. Oletetaan ensin, että järjestelmään, jossa on $N$ yksilöä, ei tule uusia yksilöitä. Olkoon tilan $j$ varanto ajanhetkellä $T$ $n_j(T)$ ja olkoon tilasta $i$ tilaan $j$ siirtymisten määrä hetkien $T$ ja $T+1$ välisenä aikana $n_{ij}(T)$. Näistä oletuksista seuraa, että

$$n_j(T+1) = \sum_{i=1}^{k} n_{ij}(T). \tag{48}$$

Kun lasketaan yhtälön (48) molempien puolien odotusarvot

$$E\left[n_j(T+1)\right] = Np_j(T+1) = E\left[\sum_{i=1}^{k} n_{ij}(T)\right] = N\sum_{i=1}^{k} p_{ij}p_i(T), \tag{49}$$

päädytään tulokseen (43) lisättynä vakiokertoimella. Tämä tarkoittaa, että aiemmin saatuja tuloksia käyttäen pystytään määräämään tilojen odotetut varannot eri ajanhetkinä, kunhan tunnetaan siirtymämatriisi $\mathbf{P}$ sekä varannot hetkellä $0$.

Järjestelmän piiriin saapuvat yksilöt kuvattiin aikaisemmin saapuvien yhteismääränä ja todennäköisyyksinä joutua eri tiloihin, joten eri tiloihin päätyvien määrät noudattavat multinomijakaumaa. Jos järjestelmän ulkopuolisten yksilöiden merkitään olevan tilassa

0, niin kuhunkin tilaan $j$ $(j = 1, ..., k)$ saapuvien odotettu määrä on $\overline{n}_{0j}(T) = R(T)r_j$. Tällöin siis laskemalla yhteen järjestelmässä olevat ja siihen tulevat yksilöt saadaan määrättyä tilojen $j$ $(j = 1, ..., k)$ odotetut varannot eri ajanhetkinä $T$ $(T = 0, 1, ...)$ yhtälöstä

$$\overline{n}_j(T+1) = \sum_{i=1}^{k} \overline{n}_i(T)p_{ij} + \overline{n}_{0j}(T) \tag{50}$$

eli matriisimuodossa

$$\overline{\mathbf{n}}(T+1) = \overline{\mathbf{n}}(T)\mathbf{P} + \overline{\mathbf{n}}_0(T). \tag{51}$$

Koska järjestelmässä otetaan nyt huomioon myös saapuvat yksilöt, ei mielivaltaisen ajanhetken varantojen määrääminen onnistu suoraan kaavan (45) sovelluksena. Sen sijaan soveltamalla kaavaa (51) toistuvasti saadaan:

$$\overline{\mathbf{n}}(T) = \overline{\mathbf{n}}(0)\mathbf{P}^T + \sum_{\tau=0}^{T-1} \overline{\mathbf{n}}_0(T-\tau)\mathbf{P}^\tau, \tag{52}$$

jossa $\mathbf{P}^0$ on yksikkömatriisi $\mathbf{I}$.

Jos $\overline{\mathbf{n}}_0(T)$ esitetään muodossa $R(T)\mathbf{r}$, havaitaan jälkimmäisen termin muuntuvan muotoon $\mathbf{r}\sum_{\tau=0}^{T-1} R(T-\tau)\mathbf{P}^\tau$, jossa summa on geometrinen matriisisarja (ks. esim. Schwartz 1961, 152-155), jos $R(T)$ oletetaan ajasta riippumattomaksi vakioksi $R$. Koska

$$\begin{aligned} & (\mathbf{I}-\mathbf{P})(\mathbf{I}+\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^T) \\ = \; & \mathbf{I}+\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^T-\mathbf{P}-\mathbf{P}^2-\cdots-\mathbf{P}^T-\mathbf{P}^{T+1} \\ = \; & \mathbf{I}-\mathbf{P}^{T+1}, \end{aligned}$$

niin

$$\begin{aligned} \sum_{\tau=0}^{T-1}\mathbf{P}^\tau &= \mathbf{I}+\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^{T-1} \\ &= (\mathbf{I}-\mathbf{P})^{-1}(\mathbf{I}-\mathbf{P})(\mathbf{I}+\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^{T-1}) \\ &= (\mathbf{I}-\mathbf{P})^{-1}(\mathbf{I}-\mathbf{P}^T) \end{aligned}$$

sillä edellytyksellä, että käänteismatriisi $(\mathbf{I}-\mathbf{P})^{-1}$ on olemassa. Voidaan osoittaa, että matriisi $(\mathbf{I}-\mathbf{P})$ on ei-singulaarinen, jos ja vain jos neliömatriisin $\mathbf{P}$ kaikki ominaisarvojen itseisarvot ovat pienempiä kuin yksi (ks. esim. Graybill 1983, 98-99).

Nyt $\mathbf{P}$ on siirtymämatriisina stokastinen matriisi eli sen alkiot ovat ei-negatiivisia ja rivit summautuvat ykkösiksi. Kun $\mathbf{1}$:llä merkitään ykkösistä koostuvaa pystyvektoria, niin $\mathbf{P}'\mathbf{1} = \mathbf{1}$, joten ominaisarvojen määritelmän perusteella $\mathbf{P}$:n transpoosilla ja näin ollen $\mathbf{P}$:llä on ominaisarvona ykkönen.

Jos palataan tarkastelemaan $\mathbf{P}$:n ositusta (47), niin voidaan todeta järjestelmästä poistuneita kuvaavan tilan $k$ olevan ominaisuuksiltaan muista tiloista poikkeava, sillä absorboivana tilana sen varanto kasvaa rajatta. Vaikka järjestelmästä poistuneiden määrää voi jossain tapauksissa olla kiinnostavaa seurata, kaikki järjestelmää oleellisesti kuvaava tieto sisältyy $k-1$:tä ensimmäistä tilaa kuvaavaan varantovektoriin $\mathbf{n}_s(T)$, täydennysvektoriin $\mathbf{r}_s$ sekä osituksessa esiintyvään matriisiin $\mathbf{S}$. Matriisin $\mathbf{S}$ kaikki alkiot ovat ei-negatiivisia ja rivisummat alle ykkösiä. Matriisia nimitetään substokastiseksi, jos sen kaikki alkiot ovat ei-negatiivisia ja rivisummat korkeintaan ykkösiä. Tämän määritelmän perusteella jokainen stokastinen matriisi on myös substokastinen ja jokainen substokastinen matriisi voidaan laajentaa stokastiseksi lisäämällä matriisiin absorboiva tila. Näin ollen on ilmeistä, että esitetty stokastisia matriiseja koskeva teoria päätee myös substokastisille matriiseille ilman olennaisia muutoksia (Feller 1968, 400-401).

Matriisin $\mathbf{S}$ kaikki tilat ovat transientteja ja sen indusoima suunnattu verkko on vahvasti kytketty (strongly connected), joten matriisi $\mathbf{S}$ on redusoimaton (irreducible). Toisin sanoen jokaiseen tilaan voidaan päätyä jokaisesta muusta tilasta. (Vrt. Feller 1968, 390-392; Ortega 1987, 217-218).

Koska substokastinen matriisi $\mathbf{S}$ on redusoimaton ja sen alkioiden itseisarvojen muodostamat rivisummat ovat alle ykkösiä, niin $\mathbf{S}$:n kaikki ominaisarvojen itseisarvot ovat pienempiä kuin yksi (ks. esim. Ortega 1987, 224) ja tällöin

$$\overline{\mathbf{n}}_s(T) = \overline{\mathbf{n}}_s(0)\mathbf{S}^T + R\mathbf{r}_s(\mathbf{I}-\mathbf{S})^{-1}(\mathbf{I}-\mathbf{S}^T). \tag{53}$$

Järjestelmää mallinnettaessa saattaa joissain tapauksissa sen niin sanottu rajakäyttäytyminen olla erityisen kiinnostuksen kohteena. Matemaattisessa mielessä tällöin ollaan kiinnostuneita jonon $\{\overline{\mathbf{n}}(T)\}$ raja-arvosta eli tavoitteena on määrätä $\lim_{T\to\infty}\overline{\mathbf{n}}(T)$. Raja-arvon olemassaolo riippuu käytännössä kokonaan järjestelmään

tulevista yksilöistä, joten jos $\{R(T)\}$ ei jossain vaiheessa asetu vakioksi $R$, varannot joko kasvavat rajatta tai vähenevät nollaan. Oletetaan $R$ vakioksi kuten edellä ja tarkastellaan kaavan (53) käyttäytymistä, kun $T \to \infty$.

Koska matriisi $\mathbf{S}$ voidaan diagonalisoida muotoon $\mathbf{S} = \mathbf{B}\Lambda\mathbf{B}^{-1}$, jossa $\Lambda$ on $\mathbf{S}$:n ominaisarvojen muodostama lävistäjämatriisi ja $\mathbf{B}$ on ei-singulaarinen neliömatriisi sarakevektoreinaan $\mathbf{S}$:n ominaisarvoihin liittyvät ominaisvektorit, niin $\mathbf{S}^T = \mathbf{B}\Lambda^T\mathbf{B}^{-1}$ (ks. esim. Patovaara 1995, 240). Edellä todettiin $\mathbf{S}$:n kaikkien ominaisarvojen itseisarvojen olevan pienempiä kuin yksi, joten kaikille ominaisarvoille $\lambda_i$ $(i = 1, ..., k-1)$ pätee $\lim_{T\to\infty} \lambda_i^T = 0$ eli $\lim_{T\to\infty} \Lambda^T = \mathbf{0}$ ja edelleen $\lim_{T\to\infty} \mathbf{S}^T = \mathbf{0}$. Näin ollen jonon $\{\overline{\mathbf{n}}(T)\}$ raja-arvo on

$$\lim_{T\to\infty} \overline{\mathbf{n}}_s(T) = R\mathbf{r}_s(\mathbf{I} - \mathbf{S})^{-1}. \tag{54}$$

Tähän mennessä käsitelty teoria mahdollistaa kaikkien tilojen odotettujen varantojen laskemisen millä tahansa ajanhetkellä $T$ $(T = 1, 2, ...)$, mutta malliin liittyvästä epävarmuudesta ei ole mitään tietoa. Määrätään hajontojen selvittämiseksi varantojen kovarianssit tilanteessa, jossa järjestelmään ei tule uusia yksilöitä:

$$\begin{aligned}
& \operatorname{cov}\left[n_j(T+1), n_l(T+1)\right] \\
= \; & E\left[n_j(T+1)n_l(T+1)\right] - En_j(T+1)En_l(T+1) \\
= \; & E\left[\sum_{i=1}^{k} n_{ij}(T) \sum_{i'=1}^{k} n_{i'l}(T)\right] - E\left[\sum_{i=1}^{k} n_{ij}(T)\right] E\left[\sum_{i'=1}^{k} n_{i'l}(T)\right] \\
= \; & \sum_{i=1}^{k}\left(\sum_{i'=1}^{k} E\left[n_{ij}(T)n_{i'l}(T)\right]\right) - \sum_{i=1}^{k}\left(\sum_{i'=1}^{k} E\left[n_{ij}(T)\right] E\left[n_{i'l}(T)\right]\right) \\
= \; & \sum_{i=1}^{k}\sum_{i'=1}^{k}\left(E\left[n_{ij}(T)n_{i'l}(T)\right] - E\left[n_{ij}(T)\right] E\left[n_{i'l}(T)\right]\right).
\end{aligned} \tag{55}$$

Mallin oletuksien perusteella $n_i(T)$:n yksilön siirtymiset tilasta $i$ tilaan $j$ $(j = 1, ..., k)$ ovat jakautuneet multinomisesti todennäköisyyksillä $(p_{i1}, p_{i2}, \ldots, p_{ik})$, joten ehdollisten odotusarvojen nojalla

$$En_{ij}(T) = E_{n_i(T)}\left(E\left[n_{ij}(T) \mid n_i(T)\right]\right) = E\left[n_i(T)p_{ij}\right] \tag{56}$$

sekä

$$E_{n_i(T),n_{i'}(T)}\left(E\left[n_{ij}(T)n_{i'l}(T) \mid n_i(T), n_{i'}(T)\right]\right) = E\left[n_i(T)n_{i'}(T)p_{ij}p_{i'l}\right], \tag{57}$$

kun $i \neq i'$ ja muuten

$$E\left[n_{ij}(T)n_{il}(T) \mid n_i(T)\right] = E\left[n_i(T)\left[n_i(T)-1\right]p_{ij}p_{il} + \delta_{jl}n_i(T)p_{ij}\right], \tag{58}$$

jossa $\delta_{jl} = 1$, jos $j = l$ ja muuten nolla. Sijoittamalla tulokset (56), (57) ja (58) kaavaan (55) saadaan

$$\begin{aligned}\text{cov}\left[n_j(T+1), n_l(T+1)\right] &= \sum_{i=1}^{k}\sum_{i'=1}^{k} p_{ij}p_{i'l}\text{cov}\left[n_i(T), n_{i'}(T)\right] \\ &\quad + \sum_{i=1}^{k}\left(\delta_{jl}p_{ij} - p_{ij}p_{il}\right)En_i(T)\end{aligned} \tag{59}$$

eli matriisimuodossa

$$\mathbf{V}(T+1) = \mathbf{P}'\mathbf{V}(T)\mathbf{P} + \lceil\overline{\mathbf{n}}(\mathbf{T})\mathbf{P}\rfloor - \mathbf{P}'\lceil\overline{\mathbf{n}}(T)\rfloor\mathbf{P}, \tag{60}$$

jossa $\mathbf{V}$ on $k \times k$-kovarianssimatriisi ja $\lceil\cdot\rfloor$ tarkoittaa diagonaalimatriisia. Näin ollen on johdettu palautuskaava ajanhetken $T$ odotettujen varantojen ja niiden kovarianssien sekä ajanhetken $T+1$ varantojen kovarianssien välille. Koska kovarianssit ovat ajanhetkellä $T = 0$ nollia, saadaan ne laskettua mille tahansa ajanhetkelle kaavoista (43) ja (59). (Vrt. Bartholomew 1981, 112-113; 1982, 20-21; Harris 1963, 36-37; Pollard 1966; Vassiliou & Gerontidis 1985.)

Oletetaan järjestelmään tulon olevan stokastisesti riippumatonta järjestelmän sisäisistä liikkeistä, jolloin

$$\begin{aligned}&\text{cov}\left[n_j(T+1), n_l(T+1)\right] \\ &= \text{cov}\left[\sum_{i=1}^{k} n_{ij}(T), \sum_{i=1}^{k} n_{il}(T)\right] + \text{cov}\left[n_{0j}(T+1), n_{0l}(T+1)\right]\end{aligned} \tag{61}$$

ja matriisimuodossa

$$\mathbf{V}(T+1) = \mathbf{P}'\mathbf{V}(T)\mathbf{P} + \lceil\overline{\mathbf{n}}(\mathbf{T})\mathbf{P}\rfloor - \mathbf{P}'\lceil\overline{\mathbf{n}}(T)\rfloor\mathbf{P} + \mathbf{V}_r. \tag{62}$$

Ainoa poikkeama tuloksesta (59) on siis viimeinen kovarianssitermi, jonka muoto riippuu sisääntulon stokastisesta luonteesta ja sisääntulotilan määräytymismenetelmästä.

Edellä tehtiin mallin luonteeseen sopivat oletukset, joissa hetkellä $T$ saapuvien yksilöiden yhteismäärä $R(T)$ noudattaa tunnettua jakaumaa ja saapuneet yksilöt joutuvat eri tiloihin todennäköisyyksillä $r_1, r_2, ..., r_k$ $(\sum_{j=1}^{k} r_i = 1)$. Sisääntulijoiden odotetut määrät saadaan tällöin laskettua yhteydestä $\overline{R}(T)\mathbf{r}$, jossa $\overline{R}(T) = ER(T)$. Kovarianssit ovat tällöin muotoa

$$\text{cov}\left[n_{0j}(T), n_{0l}(T)\right] = E\left[n_{0j}(T)n_{0l}(T)\right] - En_{0j}(T)En_{0l}(T),$$

josta ehdollistamalla $R(T)$:n suhteen

$$En_{0j}(T) = E_{R(T)}\left(E\left[n_{0j}(T) \mid R(T)\right]\right) = E\left[R(T)r_j\right]$$

sekä

$$\begin{aligned} E\left[n_{0j}(T)n_{0l}(T)\right] &= E_{R(T)}(E[n_{0j}(T)n_{0l}(T) \mid R(T)]) \\ &= E\left[R(T)\left[R(T)-1\right]r_j r_l + \delta_{jl}R(T)r_j\right]. \end{aligned}$$

Nyt

$$\begin{aligned} \text{cov}\left[n_{0j}(T), n_{0l}(T)\right] &= E\left[n_{0j}(T)n_{0l}(T)\right] - En_{0j}(T)En_{0l}(T) \\ &= E\left[R(T)\left[R(T)-1\right]r_j r_l + \delta_{jl}R(T)r_j\right] - E[R(T)]^2 r_j r_l \\ &= E\left[R(T)^2]r_j r_l - E[R(T)]^2 r_j r_l - E[R(T)]r_j r_l + E[\delta_{jl}R(T)r_j\right] \\ &= \text{var}[R(T)]r_j r_l + \overline{R}(T)(\delta_{jl}r_j - r_j r_l). \end{aligned}$$

Jos sisääntulijoiden määrä on ajan suhteen homogeenista, $R(T)$ ei riipu $T$:stä eli merkitään sitä tällöin $R$:llä. Niin empiirisessä kuin teoreettisessakin mielessä on järkevää olettaa, että $\{R(T)\}$ on jono riippumattomia Poisson-jakautuneita satunnaismuuttujia vakiokeskiarvolla $\overline{R}$. Tällöin var$\{R(T)\}$ on Poisson-oletuksen nojalla $E[R(T)] = \overline{R}(T)$ eli aikahomogeenisena $\overline{R}$. Tarkastelemalla kovarianssimatriisin rajakäyttäytymistä, kun $T \to \infty$, havaitaan kovarianssien lähestyvän nollaa ja varianssien puolestaan odotettuja varantoja. Voidaankin osoittaa, että varannot ovat asymptoottisesti Poisson-jakautuneita (Pollard 1967). Jakaumakonvergenssin nopeus riippuu pitkälti siitä, kuinka nopeasti järjestelmässä alkuhetkellä olevat yksilöt poistuvat sen piiristä. Tulos yleistyy tietyin ehdoin myös ei-homogeenisiin tilanteisiin (ks. esim. Mehlmann 1977; Vassiliou 1982).

### 2.3.2 Semi-Markov-malli

Edellä esitetyillä Markovin ketjuilla siirtymätodennäköisyydet tilasta toiseen ovat ajasta riippumattomia vakioita, joten siirtymäajat ovat geometrisesti jakautuneita. Tätä ominaisuutta voidaan pitää Markov-mallin perustavanlaatuisena heikkoutena, sillä tilojen välisten siirtymien hasardifunktio on käytännössä hyvin harvoin mallin edellyttämä vakiofunktio. Semi-Markov-malli on yksinkertaisen Markov-mallin yleistys, jossa pelkkien siirtymätodennäköisyyksien lisäksi otetaan huomioon myös tilassa ennen siirtymää vietettävä aika.

Semi-Markov-malli voidaan tulkita stokastiseksi prosessiksi, jonka siirtymät tilojen välillä tapahtuvat Markovin ketjun mukaisesti, mutta jossa jokaisessa tilassa vietetty aika ennen siirtymää on satunnaismuuttuja. Olkoon $Q_{ij}(T)$ todennäköisyys, että tilaan $i$ siirtymisen jälkeen prosessi siirtyy seuraavaksi tilaan $j$ viimeistään ajanhetkellä $T$. Tällöin on oltava kaikille $i, j = 1, 2, ..., k$ & $T \geq 0$

$$Q_{ij}(T) \geq 0$$

ja

$$\sum_{k=0}^{\infty} Q_{ik}(\infty) = 1.$$

Merkitään nyt

$$p_{ij} = Q_{ij}(\infty)$$

sekä jos $p_{ij} > 0$, niin

$$F_{ij}(T) = \frac{Q_{ij}(T)}{p_{ij}}.$$

Jos $p_{ij} = 0$, niin olkoon $F_{ij}(T)$ nolla.

Näin ollen $p_{ij}$ edustaa todennäköisyyttä sille, että prosessi, joka on juuri siirtynyt tilaan $i$ tulee siirtymään seuraavaksi tilaan $j$ ja $F_{ij}(T)$ edustaa ehdollista todennäköisyyttä, että siirtyminen tapahtuu ajassa $T$, kun tiedetään, että prosessi on juuri siirtynyt tilaan $i$ ja tulee siirtymään sieltä seuraavaksi tilaan $j$. Toisin sanoen siirtymien voidaan olettaa tapahtuvan kahdessa vaiheessa. Kun siirrytään tilaan $i$, seuraavan tilan valinta riippuu

siirtymätodennäköisyyksistä $p_{ij}$, ja edelleen, kun seuraavaksi tilaksi tulee $j$, odotusajalla ennen siirtymää on kertymäfunktio $F_{ij}(\cdot)$.

Olkoon $J_0$ prosessin alkutila ja kun $n \geq 1$, tarkoittakoon $J_n$ prosessin tilaa välittömästi sen jälkeen, kun on tapahtunut $n$ siirtymää. Tällöin prosessi $\{J_n, n = 0, 1, 2, ...\}$ on Markovin ketju siirtymätodennäköisyyksillä $p_{ij}$. Tälläistä prosessia kutsutaan upotetuksi Markovin ketjuksi (embedded Markov chain). Olkoon edelleen $N_i(T)$ siirtymien lukumäärä tilaan $i$ aikavälillä $(0, T]$ ja määritellään

$$N(T) = \sum_{i=0}^{\infty} N_i(T). \tag{63}$$

Jos $Z(T)$:llä merkitään prosessin tilaa hetkellä $T$, niin $Z(T) = J_{N(T)}$ ja tällöin stokastista prosessia $\{Z(T), T \geq 0\}$ nimitetään semi-Markov-prosessiksi. (Ks. esim. Pyke 1961a; 1961b; Ross 1970, 85-188.)

Koska viipymisajalla tilassa $i$ on kertymäfunktio $F_{ij}(T)$ ehdolla, että seuraava siirtyminen on tilaan $j$, on sillä myös eloonjäämisfunktio $S_{ij}(T)$, tiheysfunktio $f_{ij}(T)$ ja hasardifuktio $h_{ij}(T)$. Tällöin hetkellinen lähtötaajuus tilasta $i$ hetkellä $T$ on

$$h_i(T) = \sum_{i \neq j} h_{ij}(T). \tag{64}$$

Nyt yhteyden (14) perusteella

$$S_i(T) = e^{-\int_0^T h_i(u)du} \tag{65}$$

on todennäköisyys, että yksilö on hetkellä $T$ edelleen tilassa $t$, joten todennäköisyys siirtyä tilasta $i$ tilaan $j$ on

$$p_{ij} = \int_0^{\infty} h_{ij}(T) S_i(T) dT = \int_0^{\infty} h_{ij}(T) e^{-\int_0^T \sum_{i \neq j} h_{ij}(u)du} dT. \tag{66}$$

Vastaavasti ehdollinen tiheysfunktio $f_{ij}(T)$ saadaan muotoon

$$f_{ij}(T) = \frac{h_{ij}(T) S_i(T)}{\int_0^{\infty} h_{ij}(T) S_i(T) dT} = \frac{h_{ij}(T) e^{-\int_0^T \sum_{i \neq j} h_{ij}(u)du}}{\int_0^{\infty} h_{ij}(T) e^{-\int_0^T \sum_{i \neq j} h_{ij}(u)du} dT} \tag{67}$$

eli $p_{ij}$ ja $f_{ij}(T)$ voidaan esittää pelkkää hasardifunktiota $h_{ij}(T)$ käyttäen. Näin ollen semi-Markov-prosessi on täysin määrätty, kun hasardifunktiot $h_{ij}(T)$ $(i, j = 1, 2, ..., k)$ tunnetaan.

Semi-Markov-prosessissa kiinnitetyn välin siirtymätodennäköisyydet riippuvat järjestelmän tilasta välin alkupisteessä. Oletetaan, että yksilö on juuri siirtynyt tilaan $i$. Olkoon $P_{ij}(T)$ todennäköisyys, että yksilö on tilassa $j$ ajan $T$ kuluttua. Jos $j \neq i$, niin yksilön on pitänyt kokea ainakin yksi siirtymä aikavälillä $(0, T)$. Oletetaan, että ensimmäinen siirtymä tapahtuu tilaan $h$ välillä $(t, t+\delta)$, joten yksilön täytyy siirtyä edelleen tilaan $j$ välillä $(t, T)$. Jos $j = i$, on lisäksi mahdollista todennäköisyydellä $S_i(T)$, että yksilö pysyy tilassa $i$ koko aikavälin $(0, T)$. Laskemalla todennäköisyydet yli $h$:n ja $t$:n saadaan

$$\begin{aligned} P_{ij}(T) &= \sum_{h \neq i} \int_0^T h_{ih}(t) S_i(t) P_{hj}(T-t) dt + \delta_{ij} S_i(T) \\ &= \sum_{h \neq i} \int_0^T p_{ih} f_{ih}(t) P_{hj}(T-t) dt + \delta_{ij} \sum_{n \neq i} p_{in} S_{in}(T), \end{aligned} \tag{68}$$

jossa $\delta = 1$, kun $i = j$ ja muuten $0$. Olkoot $\mathbf{P}_S(T) = \{P_{ij}(T)\}$, $\mathbf{g}(T) = \{p_{ij} f_{ij}(T)\}$ ja $\mathbf{H}(T) = \lceil \sum p_{in} S_{in}(T) \rfloor$ $k \times k$-matriiseja, jolloin yhtälö (68) saadaan matriisimuotoon

$$\mathbf{P}_S(T) = \int_0^T \mathbf{g}(t) \mathbf{P}_S(T-t) dt + \mathbf{H}(T). \tag{69}$$

(Vrt. Bartholomew ym. 1991, 177-178; Mehlmann 1979; Weiss & Zelen 1965.)

Voidaan osoittaa, että yhtälö (52) yleistyy tapaukseen, jossa aika on jatkuva muuttuja (ks. esim. Bartholomew 1982, 114-115). Se saa tällöin muodon

$$\overline{\mathbf{n}}(T) = \mathbf{n}(0)\mathbf{P}(T) + \mathbf{r} \int_0^T R(x) \mathbf{P}(T-x) dx, \tag{70}$$

joka kelpaa semi-Markov-mallin varantojen määräämiseen, jos $\mathbf{P}(T)$:nä käytetään yhtälöä (69). Matemaattisesti täsmällinen ratkaisu saadaan käyttämällä Laplace-muunnosta (ks. esim. Widder 1945), jolloin

$$\mathbf{n}^*(s) = \mathbf{n}(0)\mathbf{P}^*(s) + \mathbf{r} R^*(s) \mathbf{P}^*(s)$$

ja

$$\mathbf{P}^*(s) = \mathbf{g}^*(s)\mathbf{P}^*(s) + \mathbf{H}^*(s)$$

eli

$$\mathbf{P}^*(s) = [\mathbf{I} - \mathbf{g}^*(s)]^{-1} \mathbf{H}^*(s)$$

ja

$$\mathbf{n}^*(s) = [\mathbf{n}(0) + \mathbf{r}R^*(s)]\{[\mathbf{I} - \mathbf{g}^*(s)]^{-1}\mathbf{H}^*(s)\}, \tag{71}$$

jossa * tarkoittaa Laplace-muunnosta (vrt. Bartholomew 1982, 107-108 & 130). Vaikka yhtälö (71) näyttää varsin yksinkertaiselta, sitä ei saada suljettuun muotoon kuin yksinkertaisissa erikoistapauksissa (ks. esim. McClean 1978; Taylor ym. 1998b). Muuten tuloksia voidaan saada käyttämällä numeerisia menetelmiä (ks. esim. Papadopoulou & Vassiliou 1999; Pérez-Ocón ym. 1999) tai simulaatiomenetelmiä, kuten tämän tutkimuksen rinnakkaistuotteena syntyvässä raportissa on kuvattu (Sund 2000).

Käytännössä aineisto on aina diskreettiä, joten integraali yhtälössä (69) voitaisiin korvata summalla ja laskea todennäköisyydet tämän jälkeen rekursiivisesti. Toinen vaihtoehto on formuloida semi-Markov-malli uudelleen analysointiin paremmin soveltuvaan muotoon.

Yksi vaihtoehto esittää semi-Markov-malli on yleistää niin sanottua kilpailevien riskien mallia (competing risk model). Kilpailevien riskien malli on elinaika-analyysin yleistys, jossa otetaan huomioon, että toteutuvia tapahtumia voi olla useanlaisia pelkän ”kuoleman” sijaan (ks. esim. Cox & Oakes 1984, 142-155; Elandt-Johnson & Johnson 1980, 269-341). Toisin sanoen kilpailevien riskien malleissa lähtötilasta voidaan päätyä yhteen useista toisilleen vaihtoehtoisista tiloista. Semi-Markov-mallissa on useita tiloja, joista kaikista (järjestelmästä poistuneita kuvaavia tiloja lukuunottamatta) on mahdollista päätyä muihin tiloihin. Näin ollen semi-Markov-mallin jokaista tilaa voidaan ajatella kilpailevien riskien mallin lähtötilana, ja siten semi-Markov-malli on mahdollista kuvata täysin joukolla eri tilojen välisiä kilpailevien riskien malleja. Tällainen semi-Markov-mallin formulointi auttaa ehkä intuitiivisella tasolla ymmärtämään, miksi malli redusoituu ehdollisten hasardifunktioiden määräämiseksi, mutta ennen kaikkea lähestymistapa tarjoaa pitkälle kehitettyjä menetelmiä mallin estimointiin.

Toinen mahdollisuus on istuttaa semi-Markov-malli Markovin ketjujen teoriakehykseen formuloimalla siirtymämatriisi $\mathbf{P}$ sopivalla tavalla. Perusideana on yksinkertaisesti jakaa jokainen tila $i$ siivuihin siellä vietetyn ajan mukaan. Tämän idean sovelluksia on

tutkimuskohteesta riippuen nimitetty kumulatiivisen jatkuvuuden (cumulative inertia) malleiksi (ks. esim. McGinnis 1968; Henry ym. 1971), aikaspesifeiksi (duration specific) Markov-malleiksi (ks. esim. Shinnar & Stewman 1978) tai vain laajennetuiksi (expanded) Markov-malleiksi (ks. esim. Von Korff 1978, 120-129).

Valittavan aikasiivun pituus riippuu suoraan aineiston ominaisuuksista. Yleisesti ottaen jokaisessa tilassa aika-akseli tulee jakaa yhtä pitkiin väleihin $\tau_p = [t_p, t_{p+1})$ $(p = 0, 1, ..., l-1)$ siten, että kunkin yksilön jokainen havaittu siirtymä kuuluu täsmälleen yhteen väliin ja että yhdessä välissä tapahtuu korkeintaan yksi saman yksilön siirtymä. Jotta malli pysyisi äärellisenä, oletetaan, että välillä $\tau_l = [t_l, \infty)$ siirtymätodennäköisyydet ovat vakioita. Nyt todennäköisyys siirtyä tilan $i$ aikasiivusta $p$ tilan $j$ aikasiivuun $0$ on $h_{ij}(\tau_p)$ ja todennäköisyys pysyä tilassa $i$ eli siirtyä aikasiivusta $p$ aikasiivuun $p+1$ on

$$h_{ii}(\tau_p) = 1 - \sum_{i \neq j} h_{ij}(\tau_p), \tag{72}$$

jossa $h_{ij}(T)$ on tilasta $i$ tilaan $j$ siirtymisen ehdollinen hasardifunktio. Nyt siirtymämatriisi $\mathbf{P}$ voidaan esittää muodossa:

$$\mathbf{P} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{k1} & \cdots & \mathbf{A}_{kk} \end{bmatrix}, \tag{73}$$

jossa

$$\mathbf{A}_{ii} = \begin{bmatrix} 0 & h_{ii}(\tau_0) & 0 & 0 & \cdots & 0 \\ 0 & 0 & h_{ii}(\tau_1) & 0 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & h_{ii}(\tau_{l-2}) & 0 \\ 0 & 0 & 0 & \cdots & 0 & h_{ii}(\tau_{l-1}) \\ 0 & 0 & 0 & \cdots & 0 & h_{ii}(\tau_l) \end{bmatrix}$$

ja

$$\mathbf{A}_{ij} = \begin{bmatrix} h_{ij}(\tau_0) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{ij}(\tau_{l-1}) & 0 & \cdots & 0 \\ h_{ij}(\tau_l) & 0 & \cdots & 0 \end{bmatrix}.$$

Alkuperäisten tilojen odotetut varannot saadaan yksinkertaisesti laskemalla kutakin tilaa laajennetussa mallissa vastaavien aikasiivujen odotetut varannot yhteen. Varianssit pystytään puolestaan määräämään yhteydestä

$$\text{var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \text{var}\left(Y_i\right) + 2\sum_{i=1}^{n}\sum_{j>i} \text{cov}\left(Y_i, Y_j\right) = \mathbf{1}\Sigma\mathbf{1}', \tag{74}$$

jossa $\mathbf{1} = (1, ..., 1)$ on ykkösistä koostuva rivivektori ja $\Sigma$ on tarvittava osa laajennetun mallin kovarianssimatriisia.

Semi-Markov-mallin siirtymämatriisikehitelmä paisuu nopeasti varsin suuriulotteiseksi, jolloin sen suora soveltaminen vaikeutuu. Approksimaatiomielessä olisi houkuttelevaa luokitella alkuperäinen aineisto karkeampiin aikayksiköihin ja muodostaa pieniulotteisempi siirtymämatriisi. Tämä tuottaa valitettavasti hankaluuksia teoreettisessa mielessä, sillä Markovin ketjun tiloja aggregoimalla saatu uusi Markovin ketju ei yleisesti ottaen ole enää Markov-prosessi (ks. esim. Morrison ym. 1971; Thomas & Barr 1977). Ongelman vaikutusten arviointi on tärkeää mallin käyttökelpoisuuden kannalta ja siihen palataan validiteettitarkastelujen yhteydessä (ks. kohta 2.3.4).

### 2.3.3 Semi-Markov-mallin ei-parametrinen estimointi

Semi-Markov-mallin parametrien estimoinnissa aineiston puutteellisuudet aiheuttavat täsmälleen samanlaisia ongelmia kuin pelkkien eloonjäämisfunktioidenkin tapauksessa. Semi-Markov-mallin siirtymämatriisikehitelmä johtaa ehdollisten hasardifuktioiden diskretisointiin, joten ei-parametriset estimointimenetelmät ovat erityisen käyttökelpoisia ja mahdollistavat mallin estimoinnin ilman oletuksia hasardien funktionaalisista muodoista.

Voidaan osoittaa, että Kaplan-Meier-estimaattori yleistyy varsin suoraviivaisesti kilpailevien riskien mallien ja siten myös semi-Markov-mallien estimointiin (ks. esim. Kalbleisch & Prentice 1980, 168-169; Lagakos ym. 1978; Ouhbi & Limnios 1996). Ainoana erona alkuperäiseen estimaattoriin on siirtymiselle alttiina olevien yksilöiden poikkeava väheneminen muihin tiloihin siirtymisen takia.

Tarvittava ei-parametrinen hasardifunktio on tällöin viipymisajoilla $t_p$ $(p = 1, 2, ..., l)$ muotoa

$$\widehat{h}_{ij}(t_p) = \frac{c_{ij}(t_p)}{n_i(t_p)}, \tag{75}$$

jossa $c_{ij}(t_p)$ on ajassa $t_p$ tilasta $i$ tilaan $j$ $(i \neq j)$siirtyvien määrä ja $n_i(t_p)$ hetkellä $t_p$ siirtymiselle alttiina olevien yksilöiden määrä. Lisäksi hasardifunktio on nolla niinä viipymisaikoina, joina havaittuja siirtyjiä ei ole. Tarvittaessa eloonjäämisfunktio saadaan estimoitua vastaavasti kuin yhtälöön (28) johtavat tarkastelut osoittavat. Estimaattoria voidaan edelleen yleistää tilanteisiin, jossa aineistossa on myös oikealta sensurointia monimutkaisemmilla tavoilla rajoitettuja havaintoja (ks. esim. McClean & Devine 1995; McClean & Gribbin 1987; 1991; McClean & Montgomery 1999; McClean ym. 1998; Vardi 1982).

Periaatteessa myös aktuaarimatematiikassa ja demografiassa käytetyt menetelmät tapahtumatauluineen ovat käyttökelpoisia semi-Markov-mallin estimoinnissa (ks. esim. Bartholomew ym. 1991, 184-197; Elandt-Johnson & Johnson 1980, 294-341; Von Korff 1978, 138-167), joten sopiva menetelmä kannattaa valita aineiston ominaisuuksien perusteella.

Tarkastellaan esimerkiksi semi-Markov-mallin siirtymämatriisikehitelmän tapauksessa syntyvää tilannetta. Kehitelmässä jokainen tila jaettiin siivuihin siellä vietetyn ajan mukaisesti. Ajatellaan siirtymämatriisia erillisten tilojen ja siivujen sijaan luokkina $C_i$ $(i = 1, 2, ..., t(\mathbf{P}))$, jossa $t(\mathbf{P})$ on neliömatriisin $\mathbf{P}$ tyyppi, jotta yksilö pystytään ”paikallistamaan” yhdellä parametrilla. Valitaan nyt kohortiksi tietyn ajanjakson $[u_0, u_{l+1})$ aikana järjestelmän piirissä olleet yksilöt, joiden kaikkien luokka tunnetaan tarkastelun alkamishetkenä ja joita on seurattu täydellisesti vähintään tarkastelun loppuhetkeen saakka.

Jotta tarkastelu olisi ylipäätään mielekästä, oletetaan, että valittu ajanjakso saadaan jaettua siirtymämatriisihajotelmaa vastaaviin väleihin. Merkitään nyt hetkellä $u_p$ $(p = 0, 1, ..., l)$ luokassa $C_i$ olevien yksilöiden määrää $Z_i^p$:llä ja luokkaan $C_j$ siirtyvien määrää $L_{ij}^p$:llä. Intuitiivisesti päättelemällä on varsin selvää, että kussakin luokassa $C_i$ on tarkasteluajanjakson aikana yhteensä $Z_i = \sum_{p=1}^{l} Z_i^p$ altistumista siirtymälle ja $L_{ij} = \sum_{p=1}^{l} L_{ij}^p$ havaittua siirtymää luokkaan $C_j$, joten hasardi siirtyä luokasta $C_i$ luokkaan $C_j$ on

$$h_{C_iC_j} = \frac{L_{ij}}{Z_i}. \tag{76}$$

### 2.3.4 Mallin validiteetti

Markov-mallilla saatuihin tuloksiin saattaa liittyä monista virhelähteistä peräisin olevaa epävarmuutta, jonka merkitsevyys riippuu mallin käyttötarkoituksista. Virhelähteet voidaan jakaa karkeasti kolmeen luokkaan: satunnaisvirheisiin, estimointivirheisiin ja spesifikaatiovirheisiin (Bartholomew 1975). Satunnaisvirhe on mallin stokastisista piirteistä johtuvaa tilastollista epävarmuutta. Tällaista mallin luonteeseen kuuluvaa epävarmuuden määrää pystytään mittamaan hajontojen avulla. Estimointivirhe puolestaan koskee mallin tuntemattomien parametrien määrittämistä aineistosta ja liittyy läheisesti spesifikaatiovirheisiin, sillä on täysin mahdollista sovittaa Markov-malli sopivaan aineistoon täysin mekaanisella tavalla, vaikka mallin oletukset eivät olisikaan voimassa.

Ensimmäinen askel mallin soveltamisessa on sen huolellinen suunnittelu ja sovittaminen teoreettiseen viitekehykseen (ks. sivu 9). Mallin muotoilemisen jälkeen on hyödyllistä tarkastella ja testata sitä koskevia oletuksia. Kolmanneksi mallin antamia tuloksia kannattaa verrata vastaaviin havaittuihin arvoihin. Näiden kolmen ”säännön” luova ja vuorovaikutteinen käyttäminen auttanee välttämään pahimmat virheet mallin soveltamisessa (vrt. Bartholomew ym. 1991, 114-119).

Markovin ketjujen teoriassa oletetaan, että siirtymätodennäköisyydet eivät riipu ajasta eivätkä yksilön aikaisemmasta historiasta (mallin puitteissa). Lisäksi siirtymätodennäköisyyksien tulee olla samat kaikille yksilöille ja yksilöiden täytyy olla toisistaan riippumattomia. Periaatteessa oletuksia voidaan koetella erilaisilla testeillä. Tällöin on olennaista, että nolla- ja vastahypoteesit formuloidaan tarkoituksenmukaisella tavalla ja että tuloksia tarkastellaan ”voimakkuusvarauksella”. Esimerkiksi liian pienten ja toisaalta liian isojen aineistojen käyttö saattaa testistä riippuen johtaa jopa tulkinnallisesti harhaanjohtaviin tuloksiin. (Ks. esim. Anderson & Goodman 1957; Billingsley 1961; Chatfield 1973; Denny & Yakowitz 1978; Kay 1986; Kelton & Kelton 1984; McFarland 1970; Yakowitz 1976).

Homogeenisuusoletukset niin ajan kuin ryhmän suhteen ovat mallin kannalta oleellisimpia ja helpoimmin testattavissa olevia. Yleensä jo silmämääräiset tarkastelut eri ajankohtien tai ryhmien aineistoille estimoitujen eloonjäämis- tai hasardifunktioiden välillä paljastavat oleelliset erot, mutta tarvittaessa luottamusvälien ja edelleen formaalimpien testien käyttö on tietysti mahdollista (ks. kohta 2.2.4). Homogeenisuuden testausta voidaan suorittaa Markovin ketjujen tapauksessa myös virtaamien multinomisuuteen (ks. esim. Sales 1971), siirtymämatriisin potensseihin (ks. esim. Bartholomew 1982, 33-38) tai lineaarisesti muuttuviin siirtymätodennäköisyyksiin (ks. esim. De Stavola 1988) perustuen. Yleisesti ottaen testaus pitää tehdä erikseen jokaiselle tilalle, joten tarvittavien testien määrä nousee tilojen lukumäärän kasvaessa nopeasti ja toisaalta samojen testien tulokset saattavat poiketa toisistaan eri tiloissa. Heterogeenisuuden tunnistaminen sekä huomioonottaminen on ongelmallisuudestaan huolimatta kuitenkin ensiarvoisen tärkeää järkevien tulosten aikaansaamiseksi (ks. esim. Ugwuowo & McClean 2000).

Muiden oletusten testaaminen on astetta hankalampaa, mutta käytäntö on osoittanut pienten poikkeamien vaikutuksen olevan varsin vähäistä mallin hyödyntämisen kannalta. Käytännön soveltaminen on nostattanut ylipäänsä kritiikkiä testauksen relevanttiudesta, sillä ”huonokin” malli voi olla käyttötarkoitukseensa aivan riittävä. Tämä käytännön vapaamielisyys on johtanut Markov-mallien varsin villiin ja joskus myös oletuksia rikkovaan soveltamiseen eri aloilla. Mallin toimivuuden ”yleisesti hyväksyttynä” testinä

on käytetty sen kykyä tuottaa toivottuja tuloksia eli aineiston rajoissa on kokeiltu, kuinka hyvin mallilla tehdyt ennusteet osuvat kohdalleen. Jos mallin suorituskyky osoittautuu huonoksi, ei sen ennustuskykyynkään voi luottaa. Mallin huonolle suorituskyvylle keksitään yleensä joku teoreettinen selitys, joka johtaa mallin uudelleen formulointiin (usein laajentamiseen) ja toivottavasti myös parempiin tuloksiin. Tilojen lisääminen kuitenkin kasvattaa samalla varantojen ja virtaamien määriä, joten jossain vaiheessa niiden arvoja ei enää pystytä estimoimaan riittävällä tarkkuudella. Mallin muodostaminen on näin ollen tasapainoilua realistisuuden ja aineiston mahdollisuuksien välimaastossa, sillä malli on vain niin hyvä kuin sen antamat ennusteet. (Bartholomew ym. 1991, 114-117.)

Mallin tuottamien ennusteiden osuvuutta voidaan testata yksinkertaisesti vertaamalla niitä havaittuihin arvoihin esimerkiksi kuvallisin keinoin. Vertailu voidaan haluttaessa formalisoida $\chi^2$-yhteensopivuustestiksi (ks. esim. Sales 1971; McClean & Gribbin 1991; Taylor ym. 1998b).

# 3 Hoitoilmoitusjärjestelmä

Hoitoilmoitusjärjestelmä on Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskuksen (Stakes) StakesTieto-tulosalueen ylläpitämä järjestelmä, jonka ulospäin näkyvänä tuotoksena ovat valtakunnalliset sosiaali- ja terveydenhuollon hoitoilmoitusrekisterit. Rekisterit sisältävät yksilötason tietoja sosiaali- ja terveydenhuollon laitoshoidosta, asumispalveluista, säännöllisestä kotihoidosta sekä päiväkirurgiasta. Rekisteröitäviä muuttujia ovat muun muassa asiakkaan tunnistetiedot, ikä, sukupuoli, kotikunta, hoitoaika, diagnoosi- ja toimenpidetiedot sekä tiedot hoitoon lähettävästä tahosta, hoitopaikasta sekä jatkohoitopaikasta.

Tiedot kerätään asiakaskohtaisesti hoitojakson päättyessä sekä vuoden lopussa laitoksissa sisällä olevista potilaista. Tiedot ilmoittaa palvelun tuottaja. Vuoden aikana kertyneisiin hoitoilmoituksiin sisältyvät tiedot lisätään vuosittain hoitoilmoitusrekisteriin. Terveydenhuollon rekisteri valmistuu (rekisterivuotta seuraavan) toukokuun lopussa ja sosiaalihuollon rekisteri alkusyksystä.

Diagnoosien eli tautien tai terveydenhuoltopalvelujen käytön syiden luokituksena on Suomessa ollut käytössä vuosina 1987-1995 kansainvälinen tautiluokitus ICD-9 (Lääkintöhallitus 1986). Vuodesta 1996 lähtien käytössä on ollut tautiluokitus ICD-10 (Stakes 1995; 1999a).

Toimenpideluokituksena käytettiin vuosina 1983-1995 Sairaalaliiton julkaiseman toimenpidenimikkeistön mukaista luokitusta (Sairaalaliitto 1983). Vuonna 1996 otettiin käyttöön Nomeskon (Pohjoismainen lääketieteellinen tilastokomitea) leikkausluokituksen suomalainen versio (Stakes 1996a).

Hoitoilmoitusjärjestelmä on vuosien 1993-1996 kehitysvaiheen ja muutosten jälkeen edennyt jossain määrin vakiintuneeseen tilaan. Tämän kehityksen aikana hoitoilmoitus on muodostunut sosiaali- ja terveydenhuollon tilastojärjestelmän keskeiseksi osaksi. Hoitoilmoituksen painopiste on edelleen laitoshoidon seurannassa, mutta varovaisia

laajennuksia avohoitopalvelujen suuntaan on jo tehty, sillä hoidon painopiste on yhä enenevässä määrin siirtymässä laitoshoidon ulkopuolelle. (Stakes 1996b.)

Terveydenhuollosta hoitoilmoitustietoja on kerätty vuodesta 1994 ja sosiaalihuollosta sekä päiväkirurgiasta vuodesta 1995 lähtien. Terveydenhuollon hoitoilmoitukset kattavat sairaaloiden ja terveyskeskusten vuodeosastohoidon ja sosiaalihuollon ilmoitukset puolestaan vanhainkodit ja niitä vastaavat yksiköt, kehitysvammaisten laitoshoidon, päihdehuollon laitoshoidon sekä muut ympärivuorokautista hoitoa antavat toimintayksiköt. Sosiaali- ja terveydenhuollon kotihoidon laskenta tehdään joka toinen vuosi marraskuun lopussa. Laskentaan kuuluvat hoito- ja palvelusuunnitelman piirissä olevat henkilöt: vähintään kerran viikossa kotipalveluja tai kotisairaanhoidon tai näitä molempia palveluja saavat henkilöt sekä vastaavantasoisia palveluja saavat palveluasuntojen, tukiasuntojen ja kuntoutuskotien henkilöt. (Ks. esim. Stakes 1999b, 4.)

Nykyistä hoitoilmoitusjärjestelmää edeltänyt lääkintöhallituksen ylläpitämä poistoilmoitusrekisteri kattaa vuosien 1967-1993 terveydenhuollon poistoilmoitukset päättyneistä hoitojaksoista. Poistoilmoitusten keräämisen lisäksi on suoritettu myös erillisiä potilaslaskentoja pitkäaikaishoidon ja psykiatrisen hoidon osalta, koska tietyt potilasryhmät vaihtuvat varsin hitaasti. Poistoilmoitusjärjestelmän teknisessä toteutuksessa, tietosisällössä ja keruukäytännöissä on vuosien mittaan tapahtunut lukuisia muutoksia (ks. esim. Härö 1980; Nenonen 1997; Ruotsalainen 1991, 42-43). Poistoilmoitusjärjestelmää edeltäneet erilaiset sairaalakäyttöä koskevat tilastot ulottuvat vuoteen 1841 saakka (ks. esim. Gissler 1999a; Nikiforov 1984, 1-3). Varhaisimmat terveydenhuollon tilastot koskivat piiri- ja kunnanlääkäreiden toimintaa sekä tartuntatautitilannetta eri puolilla maata, ja tilastoja on olemassa jo 1700-luvun puolelta. Sosiaalihuollon jatkuvaluonteinen tiedonkeruu alkoi 1860-luvulla köyhäinhoidon kertomuksista (Hämäläinen 1998).

Terveydenhuollon valtakunnallisen henkilörekisterilain (L471/1987, 387/94, 630/95, A476/87, 431/94) nojalla Stakes voi luovuttaa omista rekistereistään tietoja terveydenhuollon toimintaa, tautien ennaltaehkäisyä tai hoitoa koskevaa tai näihin liittyvää tieteellistä tutkimusta varten edellyttäen, että tietojen luovutus ja käyttö ei

vaaranna rekisteröidyn yksityisyyden suojaa, hänen etuaan ja oikeuksiaan eikä valtion turvallisuutta (Hämäläinen 1998; vrt. Kleemola 1998a; 1998b). Hoitoilmoitusrekisterin tietoja ehdotetaankin käytettäväksi perinteisen aggregaattitason tilastotuotannon lisäksi hoidon peittävyyden, palvelurakenteen, kustannusten, hoitoonohjausprosessin sekä potilasvirta-analyysin kuvaamiseen (Stakes 1999b, 3).

Vuosina 1994-1995 Stakesin yksilötason rekisteritietojen tutkimuskäyttöön annettiin yhteensä 125 lupaa, joista 40 % koski hoitoilmoitusrekisteriä. Luvista yli 80 % myönnettiin Stakesin ulkopuoliseen tutkimukseen. Terveydenhuollon aineistoja käytettiin huomattavasti enemmän kuin sosiaalihuollon aineistoja (Hämäläinen 1998). Tutkimusluparekisterin mukaan vuosien 1996-1999 aikana yksilötason lupa-asioita oli käsittelyssä lähes 500. Näistä noin 30 % koski hoitoilmoitusrekisteriä; vuonna 1999 31 hankkeessa käytettiin hoitoilmoitusrekisterin tietoja.

Hoitoilmoitusrekisterin tutkimuskäytössä voidaan erottaa ainakin kolme tyypillistä käyttömuotoa. Suuressa osassa tapauksista rekisteriä käytetään hakemistona varsinaiseen potilasaineistoon toimipaikoissa, jolloin tarkoituksena on vain usein rajata aineistoa haluttuun tautiin tai toimenpiteeseen. Toinen tapa on tarkastella rekisterin tietoja ja niiden välisiä suhteita sekä vuorovaikutusta käyttäen tarvittaessa apuna erilaisia aggregaattitason tietoja. Kolmas vaihtoehto on yhdistää hoitoilmoitusrekisterin tietoja jonkun muun yksilökohtaisia tietoja sisältävän tietolähteen kanssa. (Vrt. Koskinen ym. 1998.)

## 3.1 Hoitoilmoitusrekisterin luotettavuus ja kattavuus

Palvellakseen käyttötarkoitustaan on hoitoilmoitustietojen oltava riittävän kattavia ja tietosisällöltään luotettavia. Ennen vuotta 1980 poistoilmoitusrekisterin kattavuudesta tai sen tietojen oikeellisuudesta ei ollut saatavilla selvityksiä. Tämä vaikeutti ja rajoitti tietojen käyttöä, sillä tarkistusten suorittaminen jäi tietojen käyttäjien tehtäväksi (Ojala 1980; Poikolainen 1980).

Useiden potentiaalisten virhelähteiden vuoksi hoitoilmoitusaineisto ei ole täysin virheetöntä. Hoitoilmoitusaineiston tarkkuuteen vaikuttavia mahdollisia ongelmatekijöitä

ovat ainakin lääkärin tai muun hoitohenkilön tekemät tulkinnat hoidettavasta potilaasta (esimerkiksi diagnosointi ei ole aina yksiselitteistä), huolellisuus ja kattavuus potilaan tietojen kirjaamisessa sekä erilaiset käsittelyvaikeudet tietojen siirrossa hoitoilmoitusrekisteriin (Nikiforov 1984, 15; Poikolainen 1983). Myös rekisterin tietosisällön ja luokitusten muuttuminen aiheuttaa runsaasti ongelmia (ks. esim. Nenonen 1997; 1999).

Hoitoilmoitusrekisterin tietojen luotettavuutta voidaan tarkastella eri tavoin. Selvät epäloogisuudet ja mahdottomuudet pystytään löytämään helposti jo ennen tietojen viemistä rekisteriin. Esimerkiksi Hilmo-oppaan (Stakes 1996b) liitteessä 2 on annettu nykyään käytettäviä hoitoilmoitustietoja koskevia yksityiskohtaisia tarkistussääntöjä. Stakesista on muutaman viime vuoden ajan ollut lisäksi saatavissa tarkistukset automaattisesti tekevä tietokoneohjelma lähdekoodeineen. Tarkistusmahdollisuuksista huolimatta ennen vuotta 1970 virheellisiä henkilötunnuksia oli noin 15 %:ssa poistoilmoituksista (Pulkkinen & Kaprio 1980). Vuonna 1971 virheellisiä tietoja oli noin 11 %:ssa ja vuonna 1981 6 %:ssa ilmoituksia (Koskenvuo ym. 1985). Vuoden 1982 rekisteritiedoissa oli edelleen noin 5 % virheellisiä henkilötunnuksia (Salmela & Koistinen 1987). Loogisuustarkistuksia otettiinkin enenevästi käyttöön vasta automaattisen tietojenkäsittelyn yleistymisen myötä, ja virheiden määrä väheni oleellisesti vuoden 1986 jälkeen (Mähönen 1993, 44). Nykyään loogisesti virheellistä tietoa ei pääse käytännössä lainkaan rekisteriin asti.

Toinen tapa tarkastella hoitoilmoitustietojen luotettavuutta on verrata kutakin ilmoitusta vastaavan potilaan sairauskertomukseen. Tällöin siis oletetaan sairauskertomuksessa olevien tietojen olevan totuudenmukaisia. Poikolainen (1982; 1983) vertasi potilasotosten vuosia 1974-1975 koskevia rekisteritietoja vastaaviin sairauskertomuksiin viidessä alkoholinkäyttöön liittyvässä diagnoosiryhmässä. Osoittautui, että yhtäpitäviä tietoja oli hoitopaikoissa sekä syntymäajoissa 98 %, saapumispäivissä 96 %, lähtöpäivissä 94 %, asuinkunnissa 93 % ja päädiagnooseissa 91 %. Muiden verrattujen tietojen osalta yhtäpitäviä tietoja oli alle 90 %. Sairaalatyypeittäin tarkasteltuna mielisairaaloiden tiedot oli Poikolaisen mukaan täytetty parhaiten. Niissä sairauskertomusta vastaavia tietoja oli ilmoituksissa keskimäärin 9,31 kymmenestä. Hakkarainen (1982) tarkasteli otoksen

avulla, kuinka moni poistoilmoitusrekisterin mukaan vuonna 1970 ensimmäistä kertaa mielisairaalaan otetuista potilaista oli myös sairauskertomuksen mukaan ensikertalainen. Osoittautui, että 6 % ensikertalaisiksi oletetuista potilaista ei todellisuudessa ollut ensikertalaisia ja että virheiden kokonaismäärä oli noin 8 %.

Uudemmassa tutkimuksessa tarkasteltiin potilasotoksen avulla vuoden 1986 jälkipuoliskon tietoja sairaaloiden ja terveyskeskusten vuodeosastohoidossa olevien osalta (Aro ym. 1990; Keskimäki & Aro 1991). Tutkimuksen mukaan sairauskertomukset ja rekisteritiedot olivat päädiagnoosin osalta yhteneviä keskimäärin 95 %:ssa tapauksista. Tautipääryhmittäin päädiagnoosin yhdenmukaisuus oli mielenterveyden häiriöissä 98 % ja verenkiertoelinten sairauksissa 94 %. Suurista tautiryhmistä vain kasvainten päädiagnoositietojen yhdenmukaisuus jäi alle 90 %:n. Myös yleisimpien yksittäisten päädiagnoosikoodien yhdenmukaisuus oli hyvä: skitsofreniassa ja synnytyksissä 99 %, sydäninfarktissa ja alkoholismissa 98 % sekä aivoveritulpassa 95 %. Muita hyvään yhtäpitävyyteen yltäneitä tietoja olivat tulo- ja lähtöpäivämäärät (96 %), erikoisala (91 %), hoitojakson päättyminen kuolemaan (87 %) sekä psykiatrian sairaansijalle tulon tapa (86 %). Rekisteritietojen luotettavuutta heikensivät selvimmin toimenpide-, tapaturma- ja sivudiagnoositietojen puutteelliset merkinnät.

Luotettavuuden lisäksi tietojen käyttökelpoisuuteen olennaisesti vaikuttava tekijä on kattavuus. Rekisterin kattavuutta voidaan tutkia vertaamalla palvelua tuottavien yksiköiden vuosikertomuksissa ilmoitettuja hoitojaksojen ja hoitopäivien määriä hoitoilmoitusaineiston vastaaviin tietoihin. Mainittua tekniikkaa käyttävissä tarkasteluissa on arvioitu vuosien 1967-1977 ja 1982 rekisteritietojen kattavuuden olevan hoitojaksojen osalta noin 95 % (Nikiforov 1984; Salmela & Koistinen 1987). Tarkasteluissa osoittautui myös, että kattavuudessa oli vaihtelua niin alueittain kuin sairaalatyypeittäinkin. Erityisesti terveyskeskuksissa hoitoilmoitusten kattavuus oli pienempää ja virheiden määrä suurempaa kuin sairaaloissa keskimäärin. Poikolainen (1983) arvelee terveyskeskusten resurssipulan olevan yksi merkittävä syy epätarkkuuksien esiintyvyyteen. Yksittäisten toimipisteiden kohdalla esiintyvät kattavuusaukot eivät kuitenkaan yleensä oleellisesti vääristä suurempien alueiden suhteellista potilas- ja hoitorakennetta kuvaavia indikaattoreita.

Hoitoilmoitusrekisterin kattavuutta ja luotettavuutta on mahdollista tarkastella myös vertaamalla sen tietoja muiden rekisterien tietoihin. Pulkkisen ja Kaprion (1980) mukaan poistoilmoitusrekisterin ja kuolinsyyrekisterin yhdistäminen paljasti muun muassa poistoilmoitusrekisterin heikon kattavuuden sairaalassa kuolleitten osalta sekä kuolinsyydiagnoosien epävarmuuden. Heliövaara ym. (1984) vertasivat Kansaneläkelaitoksen autoklinikkatutkimuksen sairaalakäytön seurantatietoja vuosien 1973-1977 poistoilmoitusrekisterin tietoihin. Seurantatietojen mukaan sydäninfarktin vuoksi sairaalahoidossa olleista potilaista noin 85 %:lle oli tehty myös poistoilmoitus. Aivohalvauksien osalta määrä oli noin 82 %. Mähönen (1993) tarkasteli tutkimuksessaan poistoilmoitusrekisterin luotettavuutta vuosina 1983-1990 sepelvaltimotaudin kehityssuuntien seurannassa vertaamalla poistoilmoitusrekisterin tietoja tapauskohtaisesti FINMONICA-sydäninfarktirekisteriin. Tutkimuksen mukaan poistoilmoitusrekisteristä puuttui 4-13 % niistä sydäninfarktikohtauksista, jotka oli rekisteröity sydäninfarktirekisteriin sairaalassa hoidetuiksi. Selityksenä puuttuviin tietoihin oli joko se, että kliinisenä diagnoosina oli käytetty jotain muuta kuin sepelvaltimotaudin diagnoosia, tai että henkilötunnukset olivat virheellisiä. Poistoilmoitusrekisterin diagnoosi vastasi kliinistä diagnoosia noin 95-prosenttisesti.

Hoitoilmoitusjärjestelmän laajentaminen sosiaalisektorille on herättänyt eettistä keskustelua eri intressiryhmissä (Nouko-Juvonen 1999). Lisäksi osassa palveluja toteuttavista organisaatioista hoitoilmoitustietojen tuottaminen koetaan sosiaalihuollon perinteisistä käytännöistä poikkeavaksi (Aaltonen 1996). Sosiaalihuollon osalta tietojen antamisessa olikin vielä vuonna 1998 selviä kunta- ja palveluntuottajakohtaisia eroja. Aineistoa käsitellessä on otettava huomioon, että rekisterin mukainen palvelujen kehitys ei välttämättä kuvaa todellista muutosta palvelujen käytössä vaan eroja tietojen ilmoittamisessa. Aineiston kattavuus on suuntaa-antavien, muihin tilastoihin perustuvien arvioiden mukaan hyvä vain kehitysvammalaitosten osalta (vrt. Nouko-Juvonen 1997); tosin kuntayhtymien osalta asiakasmäärä on niissäkin hieman todellista pienempi. Toisaalta kehitysvammaisten autettujen asumispalvelujen luvut ovat 30-40 % liian pienet. Vanhainkotien ja vastaavien yksiköiden tiedot ovat hieman vajaita; erityisesti yksityisen sektorin osuus on puutteellista. Päihdehuollon laitosten osalta hoitoilmoitusaineisto on hoitopäivien määrän perusteella 20 % pienempi kuin muiden tilastojen mukaan.

Asumispalvelujen asiakkaista ei löydy kattavuuden arviointiin sopivia vertailutietoja. (Kauppinen 1999, 6-7.)

Sosiaalihuollon hoitoilmoituksissa diagnoosit ilmoitetaan vain, jos lääkäri on määritellyt ne diagnoosiluokitus ICD-10:n mukaisesti ja jos tieto niistä on saatavissa (Stakes 1999a, 24). Vuonna 1998 diagnooseja ilmoitettiinkin vain pienestä osasta asiakkaita. Asiakaslaskennassa diagnoosi oli noin 30 %:lla kehitysvamma- ja päihdehuollon laitosten asukkaista ja lähes 20 %:lla vanhainkotien asukkaista. Asumispalveluissa diagnooseja oli merkitty alle 10 %:ssa ilmoituksista. Päättyneiden hoitojaksojen osalta ilmoitettujen diagnoosien osuus oli vanhainkodeissa ja päihdehuollon laitoksissa samaa suuruusluokkaa asiakaslaskennan kanssa; kehitysvammalaitoksissa diagnoosimerkintöjä oli 20 %:lla ja asumispalveluissa ei käytännössä lainkaan. (Kauppinen 1999, 8.)

Hoitoilmoitusrekisterin tiedoissa on myös muita ongelmia, joita kohdataan erityisesti epidemiologisessa tutkimuksessa. Ensinnäkin on muistettava, että rekisterin peittävyys rajoittuu lähes ainoastaan laitoshoitoon. Toiseksi on huomioitava, että poistoilmoitusrekisterin diagnoosit perustuvat kliiniseen diagnostiikkaan, jossa ei käytetä tarkoin määriteltyjä kriteereitä mutta jossa diagnoosi perustuu usein huomattavasti kattavampaan tietoon potilaasta. Erillisten tauti- ja toimenpiderekisterien merkitys puolestaan perustuu tarkkaan laadunvalvontaan sekä täsmällisiin diagnostisiin kriteereihin. Epidemiologisessa tutkimustyössä vakioimattoman diagnoosin käytöllä saattaa olla merkitystä varsinkin kehityssuuntien seurannassa, koska kliinisessä työssä käytetyt diagnostiset kriteerit voivat myös muuttua ajan myötä. (Mähönen 1993, 123.)

Mähösen (1993, 128-130) mukaan poistoilmoitusrekisterin ja sydäninfarktirekisterin diagnostinen yhteensopivuus oli hyvä. Kliinisissä diagnoosikäytännöissä oli kuitenkin havaittavissa alueellisia eroja. Isohanni ym. (1997) vertailivat poistoilmoitusrekisterin skitsofreniadiagnoosien ja operationaalisten DSM-III-R-diagnoosien yhtäpitävyyttä kohorttitutkimuksessa ja havaitsivat, että 52 %:lla DSM-III-R-kriteerin täyttävistä potilaista oli hoitoilmoitusrekisterissä diagnoosina skitsofrenia. Diagnoosikoodit oli siirretty luotettavasti poistoilmoitusrekisteriin: virheellisesti koodattuja tietoja oli vain

1,6 %. Lisäksi peittävyys osoittautui hyväksi, sillä kokonaan laitoshoidon ulkopuolella hoidettuja skitsofreenikkoja löytyi kaikista skitsofreenikoista vain 3 %. Cannonin ym. (1998) tutkimuksen mukaan skitsofrenian rekisteridiagnoosi oli Helsingissä syntyneistä potilaista koostuvassa pienehkössä otoksessa yhtenevä haastattelijan tekemän DSM-III-R-diagnoosin kanssa 92 %:lla potilaista. Mäkikyrön ym. (1998) tutkimuksessa 93.2 % skitsofreniadiagnoosilla hoitoilmoitusrekisteriin merkityistä yhden kunnan potilaista täytti myös skitsofrenian DSM-III-R-kriteerin. Suvisaari (1999, 99-100) toteaa skitsofreniadiagnoosien reliabiliteettitutkimusten osoittavan, että Suomessa psykiatrit käyttävät kliinisessä työssään varsin tiukkoja kriteerejä skitsofreniadiagnooseille ja että poistoilmoitusrekisterissä olevia diagnooseja voidaan pitää varsin luotettavina.

Diagnoosikoodiston muutos ICD-9:stä ICD-10:een vuonna 1996 on vaikuttanut diagnoosijakaumiin, mikä on otettava huomioon tarkasteltaessa muutoksia edellisiin vuosiin (Lounamaa ym 1997). Luokituksesta johtuvat ongelmat tulevat esiin lähinnä psykiatrisissa diagnooseissa, alkoholiin liittyvissä ongelmissa sekä tapaturmissa. Skitsofrenian diagnoosiryhmän osalta koodiston muutos ei näytä kuitenkaan vääristäneen merkittävästi potilaiden tai hoitopäivien määriä; psykiatrisissa diagnooseissa suurimmat ongelmat ilmenevät neuroosi-diagnoosiryhmässä (Nenonen ym. 2000).

Yleisesti ottaen diagnoosiryhmittäisissä tarkasteluissa yksittäisten potilaiden diagnoosien pysyvyys aiheuttaa erityisongelmia. Potilaiden eri hoitojaksojen väliset diagnoosien vaihtelevuudet ovat tulkinnanvaraisia ja vaatisivat tapahtumakohtaista tarkastelua tai muita erityistoimenpiteitä. Käytännössä merkitsevät ongelmat ilmenevätkin hoitoilmoitusrekisterissä selkeimmin asiakaslaskentojen ja vastaavien poistoilmoituksien diagnoosien eroina. Laskentadiagnoosi perustuu lyhytaikaisemmalle potilaan seurannalle kuin vastaava poistoilmoitus, joten periaatteessa laskentadiagnoosi on epävarmempi kuin poistodiagnoosi. Näin ollen pelkkiin laskentatietoihin perustuviin diagnoosiryhmätarkasteluihin tulee suhtautua varauksella.

Nykyään tiedot hoitoilmoitusrekisteriin saadaan monissa tapauksissa suoraan laitosten tietojärjestelmistä, joita useimmissa sairaaloissa käytetään myös kuntalaskutuksen perustana. Näin aineiston oikeellisuus on myös laitoksen taloudellinen etu;

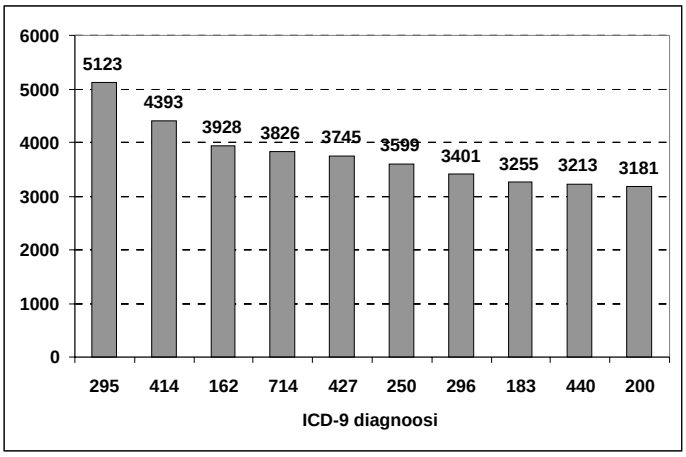
oikeellisuutta seuraavat myös laskun maksajat eli kunnat. Etenkin DRG-pohjaiset hinnoittelujärjestelmät ovat lisänneet aineiston tarkkuusvaatimuksia, koska epälooginen diagnoosi–toimenpide-pari aiheuttaa käynnin luokittelemisen hinnaltaan huokeampaan ”kaatoluokkaan” ja näin pienentää palvelun tuottaneen yksikön saamaa maksua palveluksistaan.

Yhteenvetona hoitoilmoitusrekisterin luotettavuudesta voidaan todeta, että rekisteritiedoilla ja sairauskertomuksilla on hyvä vastaavuus, mikäli tiedot vain tulevat rekisteröidyiksi. Tietojen puuttuminen heikentää kuitenkin oleellisesti rekisterin käyttökelpoisuutta. Rekisterin laadun parantamiseksi onkin ehdotettu monenlaisia toimenpiteitä: selkeämpää ohjeistusta, hoitoilmoitustietojen käsittelyyn osallistuvien henkilöiden koulutusta, luotettavuuden toistuvaa arviointia, diagnoosinumeroiden käytön seurantaa, yhteistoimintaa eri tautirekisterien kanssa, tietosisällön ajoittaista tarkistusta, nopeaa palautetta ja aineiston käyttökelpoiseksi tekevää raportointia. (Aro ym. 1990; Keskimäki & Aro 1991; Mähönen 1993, 121-124; Poikolainen 1983; Salmela & Koistinen 1987.)

Nykyisellään hoitoilmoitusrekisterin aineisto on varsin käyttökelpoista terveydenhuollon ja sosiaalihuollon laitosten osalta. Harvinaisempia sairauksia tai potilasryhmiä tarkastellessa on kuitenkin otettava huomioon, että hoitoilmoitusrekisterin tiedot saattavat erilaisista syistä johtuen olla hyvinkin harhaanjohtavia. Näin ollen on välttämätöntä varmistaa hoitoilmoitusaineiston soveltuvuus tutkimuksen tietolähteeksi erikseen kunkin ilmiön kohdalla.

# 4 Skitsofreenikoiden potilaspopulaatio

Tämän tutkimuksen aineistoa valittaessa päätettiin lähtökohtaisesti rajoittua terveydenhuollon hoitoilmoitusrekisteristä helposti identifioitavan sairausryhmän koko maan tasoisiin tietoihin. Päämääränä oli löytää hoitoilmoitustietojen puolesta mahdollisimman ongelmaton ja luotettava potilaiden osajoukko, jonka muutostapahtumat olisivat luonteeltaan mallinnukseen soveltuvia. Yhtenä tärkeänä soveltuvuuden kriteerinä pidettiin hoidon toistuvuutta saman diagnoosin puitteissa. Asiantuntijoiden kanssa käytyjen keskustelujen perusteella syntyneiden ehdotusten todentamiseksi laskettiin kolmen merkin päädiagnooseittain ICD-9-tautiluokituksen (Lääkintöhallitus 1986) toistuvaa hoitoa saaneiden potilaiden lukumäärät, jonka jälkeen arvioitiin mikä esille tulleista diagnooseista olisi esimerkkiaineistoksi sopivin. Kuviossa 3 on esitetty diagnoosiryhmittäin niiden potilaiden määriä, joilla on samalla päädiagnoosilla enemmän kuin kolme hoitojaksoa vuosien 1994-1998 aikana. Tämän suhteellisen lyhyen ajanjakson aikana hoitoilmoitusrekisterin mukaan toistuvaa hoitoa vaativia sairauksia ovat siis muun muassa mielenterveyden häiriöt (diagnoosikoodit 295; 296), sydän- ja verisuonitaudit (414; 427; 440), syövät (162; 183; 200), nivelreuma (714) sekä diabetes (250).




Kuvio 3. Vähintään kolme kertaa laitoksessa vuosien 1994-1998 aikana olleiden potilaiden määriä diagnoosiryhmittäin

Suoritettujen tarkastelujen perusteella skitsofreeniset häiriöt (295) vaikuttivat tämän tutkimuksen tarpeisiin parhaiten soveltuvilta, sillä hoidon toistuvuuden lisäksi tietojen luotettavuus ja kattavuus on aikaisemmissa tutkimuksissa osoittautunut varsin hyväksi (ks. kohta 3.1). Valintaa puoltava tekijä oli myös se, että Hakkaraisen (1985; 1989) psykiatrisen potilasväestön muutostapahtumia kuvaava teoreettinen malli oli suoraan sovellettavissa skitsofreenisiin häiriöihin liittyvien muutostapahtumien malliksi.

Näin ollen hoitoilmoitusrekisteristä poimittiin kaikki poisto- ja asiakaslaskentailmoitukset vuosilta 1994-1998, joissa ICD-9-tautiluokituksen mukainen kolmen merkin päädiagnoosi oli 295 eli skitsofreeninen häiriö. ICD-9 luokituksen diagnoosikoodia 295 vastaavat vuodesta 1996 asti voimassa olleessa ICD-10-luokituksessa diagnoosit F20.0-F20.9, F23.1, F23.2, F25.0-25.2, F25.8 sekä F25.9 (Lääkintöhallitus 1986; Stakes 1995; Stakes 1999a). Mainituilla ehdoilla hoitoilmoitusrekisteristä löytyi 95 705 tietuetta, joista jokainen sisälsi potilaan salatun henkilötunnuksen, hoitolaitoskoodin, syntymävuoden, sukupuolen sekä tulo- ja poistopäivän.

Ei ole välttämättä itsestään selvää, minkälaista populaatiota kolmen merkin päädiagnoosiehdolla poimittu otos hoitoilmoitusrekisteristä kuvaa. Esimerkiksi psykiatrista potilasväestöä kuvaavat aineistot on usein rajattu joko toimipaikan tai erikoisalan mukaan, ja toisaalta epidemiologisessa tutkimuksessa aineiston poiminta hoitoilmoistusrekisteristä saatetaan mieluummin suorittaa henkilötunnuksen kuin diagnoosin perusteella. Yksittäiseen diagnoosiin rajoituttaessa ei pystytä havaitsemaan potilaiden muilla diagnooseilla saamaa hoitoa, joten jonkinlainen täydennyspoiminta henkilötunnuksien perusteella voi olla hyvinkin tarkoituksenmukaista. Yksinkertaisuuden vuoksi tässä tutkimuksessa ei ole suoritettu mitään täydennyspoimintoja, joten käytetty otos kuvaa puhtaasti hoitoilmoitusten perusteella havaittavaa skitsofreenisia häiriöitä koskevaa terveydenhuollon laitoskäyttöä sekä hoitokäytäntöjä. Yksittäiseen diagnoosiin rajoittumisesta näyttäisi tässä tapauksessa seuraavan vain vähäisiä ongelmia (ks. sivu 54). Tutkimusaineisto on myös esimerkki hallinnollisen kategorisoinnin tuottamasta keinotekoisesta mutta sinänsä käyttökelpoisesta luokituksesta (vrt. kohta 5.1).

Koska rekisteristä poimittiin sekä poistoilmoitukset että asiakaslaskennan tiedot, osa havainnoista sisälsi päällekkäisiä tietoja. Myös pelkistä poistoilmoituksista löytyi sisäkkäisiä sekä limittäisiä hoitojaksoja. Tyypillisesti sisäkkäinen hoitojakso oli lyhyt käynti toisessa laitoksessa pidemmän hoitojakson puitteissa. Limittäisyyksistä puolestaan ainakin osa johtunee vaihtelevista kirjaamiskäytännöistä.

Kun aineistosta poistettiin päällekkäisyydet, yhdistettiin limittäiset sekä välittömästi toisiaan seuraavat hoitojaksot ja pidettiin huolta siitä, että siivotusta aineistosta löytyi asiakaslaskentapäivinä jokainen kyseisessä laskennassa skitsofreenisten häiriöiden päädiagnoosilla mukana ollut potilas, jäi jäljelle 74 431 havaittua hoitojaksoa. Kaksi saman potilaan peräkkäistä hoitojaksoa luokiteltiin välittömästi toisiaan seuraaviksi, jos seuraava jakso alkoi edellisen jakson poistopäivänä tai sitä seuraavana päivänä.

Hoitojaksojen yhdistämiseen johtavan vuorokausimäärän valitseminen on ongelmallista, sillä periaatteessa se vaatisi tapauskohtaista tarkastelua ja mahdollisesti myös tietoja kunkin potilaan laitosten ulkopuolella saamasta hoidosta. On myös varsin yleistä, että potilaat kirjataan poistetuiksi myös lomien ajaksi, jolloin poistaminen tai sisäänottaminen ei välttämättä tarkoita muutosta potilaan hoidon tarpeessa. Myös laitoksesta toiseen siirtäminen aiheuttaa teknisen, tietyssä mielessä turhan ilmoituksen. Korkeila ym. (1998) ovat edellä mainituista syistä määritelleet hoitojaksot psykiatrisissa sairaaloissa erillisiksi, jos poiston ja sitä seuraavan sisäänoton välinen aika on vähintään kahdeksan vuorokautta. Von Korff (1978, 27) on puolestaan vastaavanlaisessa aineistossa yhdistänyt hoitojaksot, jos poissaolo laitoksesta on kestänyt korkeintaan 30 vuorokautta. Pitkää yhdistämisperiodia käytettäessä aineistosta alkaa kuitenkin häviätä hallinnollisessa mielessä oleellista informaatiota erilaisista hoitokäytännöistä, joten tässä tutkimuksessa teknisetkin poistot on otettu mukaan tarkasteluihin.

Asiakaslaskennoista peräisin olevissa tiedoissa poistopäivä on usein tuntematon eli havainnot ovat sensuroituja. Periaatteessa asiakaslaskentaa seuraavan vuoden aikana tehdyistä poistoilmoituksista tai viimeistään seuraavasta asiakaslaskennasta tulisi löytyä tietoa kunkin potilaan kohtalosta. Näin ollen vain vuoden 1998 asiakaslaskentatiedoissa tulisi olla sensuroituja havaintoja. Tästä huolimatta vuosien 1994-1997 laskentatiedoissa

on skitsofreenisen häiriön diagnoosilla kaiken kaikkiaan 655 potilasta, joista ei kyseisen laskennan jälkeen löydy poistoilmoitusta eikä myöhempää laskentatietoa samalla diagnoosilla. Toisin sanoen vuosittain noin 3-4 %:lta potilaista puuttuu edellistä laskentaa vastaavalla diagnoosilla kirjattu hoitoilmoitus. Hoitoilmoitusrekisteristä ei ole erikseen tarkistettu, kuinka suurelle osalle potilaista löytyy eri diagnoosilla kirjattu ilmoitus. Ilman tarkistus- ja täydennyspoimintoja aineistossa on yhteensä 3 613 poistopäivän osalta puutteellista havaintoa eli sensuroitujen havaintojen osuus on 4.9 % kaikista havainnoista. On merkillepantavaa, että aineistossa ilmenevä sensuuri ei ole satunnaista vaan johtuu hoitoilmoitusrekisterin tietojen keruun luonteesta eli sensurointiajankohdat ovat kiinteitä ja vastaavat asiakaslaskentapäiviä. Sensurointiaikojen kiinteys johtaa joissain tapauksissa matemaattisesti hankalammin käsiteltäviin tai ominaisuuksiltaan huonompiin estimaattoreihin (ks. esim. Akritas 1986; Efron 1981; Reid 1981).

Aineiston luotettavuuden kannalta on varsin erikoista, että poistoilmoitusten perusteella lasketut skitsofreenisten häiriöiden potilaspopulaatiot ovat asiakaslaskentapäivinä säännönmukaisesti miltei 10 % suurempia kuin pelkät asiakaslaskentatiedot antaisivat odottaa. Nämä ”ylimääräiset” potilaat ovat sijoittuneina varsin tasaisesti eri laitoksiin, joten mitään systemaattista laskentailmoitusten laitoskohtaista katoa ei ole havaittavissa. Luultavasti suurimmalle osalle näistä potilaista on asiakaslaskennassa annettu skitsofreenisesta häiriöistä poikkeava päädiagnoosi. Vaikka kyseisten potilaiden asiakaslaskentadiagnooseja ei olekaan erikseen varmistettu hoitoilmoitusrekisteristä, voidaan todeta havaitun ilmiön olevan sopusoinnussa aikaisempien tutkimustulosten (ks. sivu 49) kanssa, joiden mukaan skitsofreniadiagnoosin antamisessa käytetään varsin tiukkoja kriteerejä: yleensä laskentapäivänä potilaasta tiedetään vähemmän kuin poistopäivänä, joten diagnoosin kanssa ollaan varovaisia. Käytännössä voidaan siis olettaa, että myös vuoden 1998 asiakaslaskentatiedot sisältävät liian vähän skitsofreenisten häiriöiden potilaita.

Rajoittumalla hoitoilmoitusrekisterin tietoihin kaikkia populaation todellisuudessa kokemia tapahtumia, kuten kuolemaa tai ulkomaille muuttamista, ei pystytä havaitsemaan kattavasti. Tässä tutkimuksessa käytetty aineisto sisältää hoitoilmoistusrekisterin tiedoistakin ainoastaan skitsofreenisten häiriöiden päädiagnoosilla saadun hoidon, joten

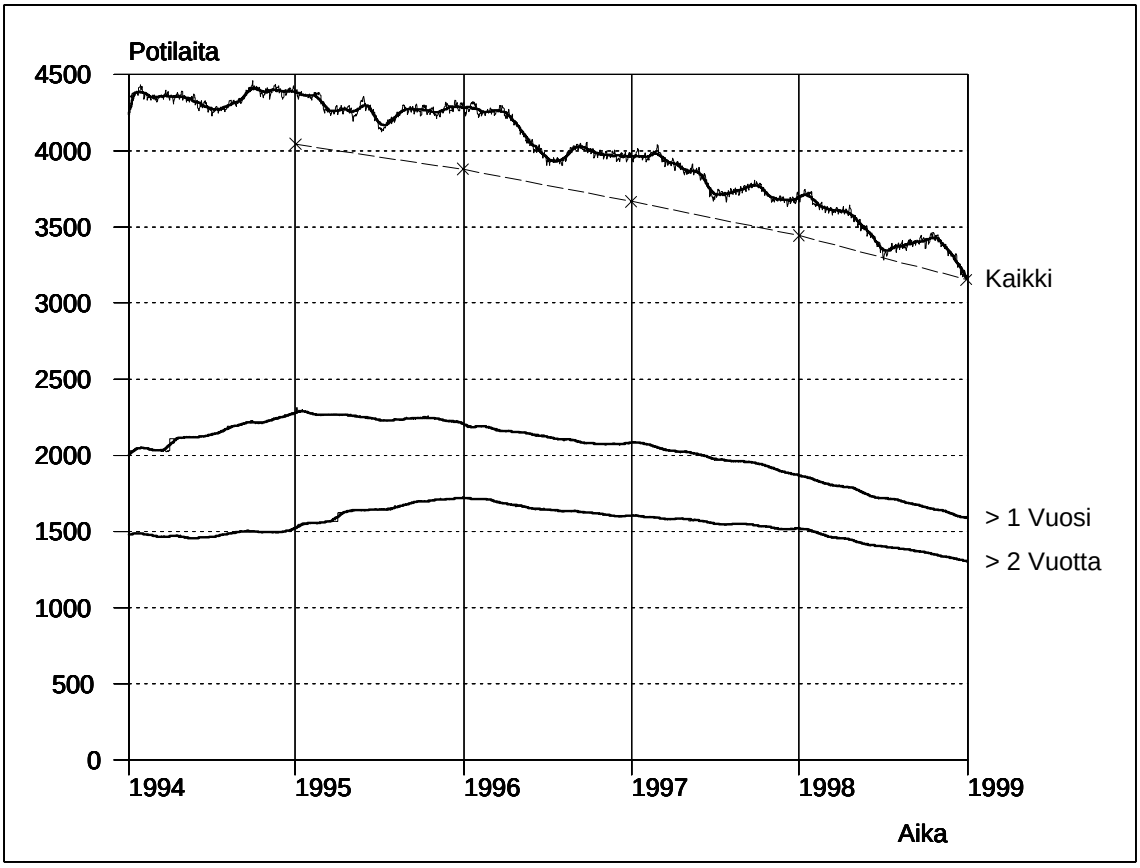
sopiva täydennyspoiminta eri rekistereistä saattaisi antaa tarpeellista lisäinformaatiota. Puutteet eivät kuitenkaan muodosta tämän tutkimuksen puitteissa mainittavaa ongelmaa, kunhan niiden olemassaolon mahdollisuus tiedostetaan.

## 4.1 Skitsofreenisten häiriöiden potilaspopulaation ominaisuuksia

Potilaspopulaatioita voidaan kuvailla niiden koon perusteella eli tarkastelemalla kuinka paljon potilaita järjestelmän eri tiloissa kullakin ajanhetkellä on. Yleensä on myös tarkoituksenmukaista hahmotella kyseessä olevan potilaspopulaation rakennetta ryhmittelemällä potilaat pienempiin ryhmiin sopivien yksilökohtaisten ominaisuuksien perusteella. Esimerkiksi laitoksessa olevien potilaiden populaation koko ja rakenne muuttuvat ajan myötä jatkuvana prosessina, kun potilaita otetaan laitokseen ja poistetaan laitoksesta. Nämä monista eri syistä tapahtuvat siirtymät eli transitiot laitoksessa ja sieltä poissa olemisen välillä ilmentävät laitospopulaation dynamiikkaa (Von Korff 1978, 26). Tässä kappaleessa tarkastellaan laitospopulaation kokoa ja rakennetta niiltä osin, kuin se on tarpeellista dynamiikkaa kuvaavan mallin rakentamiseksi.

### 4.1.1 Laitosvaranto

Hoitoilmoitusrekisterin tiedoista pystytään suoraan määrittämään potilaiden päivittäinen laitosvaranto poimimalla aineistosta ne hoitojaksot, joissa tarkastelupäivä osuu tulo- ja poistopäivän väliin. Käytetyssä tutkimusaineistossa hankaluuksia aiheuttivat vain vuosien 1994-1997 sensuroidut havainnot. Seuraavia tarkasteluja varten sensuroitujen havaintojen poistopäivien oletettiin olevan tasaisesti jakautuneita sensurointihetkeä seuraavan vuoden ajalle, jonka jälkeen puuttuvat tiedot oli helppo korvata oletuksen mukaisilla (satunnaisilla) poistamispäivillä. Puutteellisten tietojen pienestä määrästä johtuen ne eivät oleellisesti vääristä havaittuja trendejä. Yksittäisen päivän kohdalla sensuroiduista havainnoista johtuva virhe on huonoimmassa mahdollisessa tapauksessa 3 %:n luokkaa. Tällöin kaikkien sensuroitujen havaintojen oikean poistamispäivän pitäisi olla laskentapäivää seuraava päivä.




Kuvio 4. Laitoksissa läsnäolevien skitsofreniapotilaiden varannot päivittäin sekä asiakaslaskentojen perusteella

Kuviossa 4 on esitetty skitsofreniapotilaiden päivittäiset laitosvarannot. Kuviossa on koko laitosvarannon lisäksi esitetty sekä vuoden että kaksi vuotta yhtäjaksoisesti laitoksessa olleiden päivittäiset määrät, joiden perusteella on mahdollista arvioida pitkäaikaispotilaiden osuutta kaikista laitoksessa olevista potilaista tarkasta pitkäaikaispotilaan määritelmästä riippumatta. Myös asiakaslaskentojen mukaiset liian pienet varannot on luotettavuustarkastelussa mainitun ongelman havainnollistamiseksi merkitty kuvioon.

Päivittäisiä varantoja on tasoitettu 30 päivän liukuvalla keskiarvolla satunnaisvaihtelun vähentämiseksi. Kuviosta voidaan havaita vuoden 1996 jälkeinen laitosvarannon varsin tasainen pieneneminen yli neljästä tuhannesta potilaasta alle 3 500 potilaaseen. Varantojen vaihtelussa on myös vuoden sisällä tapahtuvaa vaihtelua: kesäkuukausina potilasmäärät

| Vuosi | Hoidettuja Potilaita | Prevalenssi per 1000 as. kesk.määr. vuoden aikana | Hoitopäiviä | Hoitopäiviä per hoid.potilas | Alkaneita hoitojaksoja | Alk.hoitojaksoja per hoid.potilas |
|---|---|---|---|---|---|---|
| 1994 | 11009 | 2.16 | 1584570 | 116 | 13624 | 1.24 |
| 1995 | 10778 | 2.11 | 1531586 | 113 | 13601 | 1.26 |
| 1996 | 10870 | 2.12 | 1452694 | 104 | 13938 | 1.28 |
| 1997 | 10412 | 2.02 | 1360336 | 94 | 14441 | 1.39 |
| 1998 | 10112 | 1.96 | 1235673 | 85 | 14519 | 1.44 |

Taulukko 1. Skitsofreenisten häiriöiden diagnoosilla hoidettujen potilaiden määrät, hoitopäivät ja alkaneet hoitojaksot vuosina 1994-1998

ovat pienimmillään. Vähintään vuoden laitoksessa olleita potilaita on noin puolet koko varannosta. Näiden pitkäaikaispotilaiden määrä seurailee koko varannon vuositasoista trendiä ilman ”kesäkuoppia” eli vuoden sisäinen vaihtelu kohdistuu kokonaisuudessaan lyhytaikaispotilaisiin.

### 4.1.2 Hoitokerrat

Taulukossa 1 on esitetty ICD-9-tautiluokituksen päädiagnoosi 295:n mukaisten hoidettujen potilaiden, hoitopäivien ja alkaneiden hoitojaksojen lukumäärät sekä tuhatta asukasta kohden suhteutetut potilasmäärät (prevalenssit) vuosina 1994-1998. Hoitoa saaneiden potilaiden ja hoitopäivien määrä on tänä aikana vähentynyt selvästi (vrt. Nenonen ym. 2000), mutta vuoden aikana alkaneiden hoitojaksojen määrä on puolestaan kasvanut voimakkaasti. Tämä kuvastaa hoitokäytäntöjen muutoksia: pitkäaikaisen laitoshoidon sijaan tarjotaan yhä enenevässä määrin toistuvaa lyhytaikaista hoitoa. Potilasmäärällä suhteutettuja tunnuslukuja tulkittaessa tulee kuitenkin ottaa huomioon, että niissä oletetaan hoitopäivien ja hoitojaksojen määrien jakautuneen tasaisesti potilaille. Todellisuudessa jakaumat ovat hyvinkin vinoja, sillä pitkäaikaispotilaat keräävät ison osan hoitopäivistä, vaikka kyseisillä potilailla on hyvin vähän alkaneita hoitojaksoja vuoden aikana. Prevalenssin pienentyminen kuvastaa hoidossa olleiden suhteellisen osuuden vähentymistä; saatu tulos seurailee 1990-luvun alun trendiä (ks. esim. Korkeila ym. 1998; Salokangas ym. 2000).

Potilailla on lähes poikkeuksetta lukuisia uusintakäyntejä laitoksessa. Koska käytetyssä aineistossa ei ollut tietoja siitä, monennestako hoitokerrasta kulloisessakin hoitojaksossa oli kyse, ei ensi kertaa laitokseen otettavia yksilöitä pystytty tunnistamaan yksiselitteisellä

tavalla. Hakkaraisen (1982; 1989, 77-80) mukaan Suomessa uusintakäynti tapahtuu käytännössä kolmen vuoden kuluessa edellisestä poistamisesta. Käytetyssä aineistossa tämä siis tarkoittaa, että vuoden 1997 jälkeen ensimmäistä kertaa laitokseen otettuja potilaita voidaan pitää ”todellisina” ensikertalaisina. Myös Korkeila ym. (1998) ovat käyttäneet tutkimuksessaan vastaavaa kriteeriä ensikertalaisille.

Vuonna 1997 ensikertalaisia oli yhteensä 2754 ja vuonna 1998 yhteensä 2357. Vastaavat insidenssit (määrät suhteutettuna riskiväestöön) ovat 0.54 ja 0.46 promillea. Vuonna 1998 sekä ensimmäistä että toista kertaa laitoksessa olleiden osuudet päivittäisestä laitosvarannosta pysyttelivät varsin vakaasti 20 %:n seutuvilla eli vähintään kolmatta kertaa laitoksessa olevien osuus päivittäisestä varannosta oli noin 60 %.

Vuosien 1994-1998 välisenä aikana laitoskäyntien määrä oli keskimäärin 3.2 käyntiä potilasta kohden. Noin 10 %:lla potilaista oli enemmän kuin kuusi hoitokertaa ja noin 5 % ylitti kymmenen hoitokerran rajan. Maksimissaan yhdellä potilaalla oli lähes sata laitoskäyntiä. Moninkertaiset uusijat voivat kerätä itselleen pitkän kokonaishoitoajan, vaikka yksittäiset hoitojaksot jäävätkin lyhyiksi. Jatkuvassa sairaalahoitokierteessä olevia potilaita kutsutaan pyöröovipotilaiksi (Hakkarainen 1982; Von Korff 1978, 70-84). Jatkuva hoitokierre voi johtua niin potilaasta (esimerkiksi alkoholiongelma (Hakkarainen 1989)) kuin hoitoa tarjoavasta organisaatiostakin (potilaita pyritään pitämään mahdollisimman paljon laitoshoidon ulkopuolella). Myös hoitokäytäntöjen viimeaikaiset suuntaukset heijastuvat käytäntöön toistuvina hoitojaksoina; esimerkiksi erilaiset tuki- ja seurantahoidot perustuvat toistuvaan lyhytaikaiseen palvelujen käyttöön. Osaltaan lyhyet hoitoajat johtuvat tietoisesta pyrkimyksestä välttää potilaiden institutionalisoitumista. Joidenkin potilaiden useat hoitojaksot saattavat kuitenkin myös viestiä sitä, ettei kaikille potilaille kyetä tarjoamaan tarkoituksenmukaista hoitopaikkaa kotiutuksen jälkeen (vrt. Nouko-Juvonen 1999). Mielenterveyden Keskusliiton toistuvan psykiatrisen hoidon tarpeen ehkäisyyn tähtäävän projektin väliraportin (Nouko-Juvonen 2000) mukaan ”pyöröovisyndrooma” johtuu sairaalahoitoa kevyempien vaihtoehtojen puutteesta: päivätoimintakeskuksista on huutava pula ja sairaalajakson jälkeen henkilö ”hukataan”, sillä hänelle ei ole luotu siteitä avohoidon palveluihin. Lisäksi ongelmaa pahentavat taloudelliset vaikeudet, virikkeiden ja seuran

puute, puutteelliset asunto-olosuhteet sekä voimavarojen puuttuessa kykenemättömyys huolehtia itsestään yksin ollessa. Usein myös jo olemassa olevista palveluista ei osata hakea tietoa ennen kuin ollaan sairaalahoidon tarpeessa. Salokangas ym. (2000) ovat tutkineet tarkemmin hoitojärjestelmän muutosten ja sairaalasta kotiuttamisen vaikutuksia skitsofreniapotilaiden elämään ja hoitotilanteeseen.

### 4.1.3 Ryhmittely ja ryhmien ominaisuudet

Kaiken kaikkiaan tutkimusaineistossa on hoitojaksotietoja 23 184:stä potilaasta. Potilaat on jaettu Hakkaraisen (1980; 1989, 21-30) ja Von Korffin (1978) tarkasteluja seuraillen iän ja sukupuolen mukaan kahdeksaan ryhmään. Ryhmiin jako on tehty käytännön syistä iän sijaan syntymävuoden perusteella siten, että kukin potilas kuuluu vain yhteen ryhmään tarkasteluajanjakson kuluessa tapahtuvasta ikääntymisestä huolimatta. Taulukossa 2 on esitetty tutkimusaineiston potilaiden ikä- ja sukupuolijakaumat käytetyllä ryhmäjaolla, joka tässä tapauksessa vastaa lähinnä ikäryhmitystä: alle 25-, 25-44-, 45-64- ja yli 64-vuotiaat. Nuorimmassa ikäryhmässä miesten osuus on huomattavasti naisten osuutta suurempi, kun taas vanhimmassa ryhmässä tilanne on päinvastainen. Aineiston nuorimman miespuolisen potilaan syntymävuosi on 1989 ja vanhimman puolestaan 1899. Naisilla vastaavat vuodet ovat 1984 ja 1896. Kaiken kaikkiaan alle 15-vuotiaana skitsofreniadiagnoosilla hoitoa saaneita potilaita oli aineistossa vain 18 (0.08 %), mikä johtunee siitä, ettei skitsofreniadiagnoosia yleisen käytännön mukaan anneta nuorille potilaille. Tautiluokitus ICD-10:n mukaan skitsofreniaan ei myöskään sairastuta yli 45-vuotiaana. Tästä huolimatta aineiston mukaan vuosina 1997-1998 ensimmäistä kertaa sisäänotetuista skitsofreenisten häiriöiden (ICD-9: 295) potilaista 51 % oli vähintään 45-vuotiaita. Tämä johtunee luokituskoodiston ja diagnoosikriteerien muutoksista (vrt. Suvisaari 1999, 21), kliinisistä käytännöistä sekä aineiston rajoituksista.

| Syntymävuosi | Mies | Nainen | Yhteensä |
|---|---|---|---|
| -1930 | 1001 (4.3 %) | 2465 (10.6 %) | 3466 (14.9 %) |
| 1931-1950 | 3875 (16.7 %) | 4104 (17.7 %) | 7979 (34.4 %) |
| 1951-1970 | 5739 (24.8 %) | 4261 (18.4 %) | 10000 (43.1 %) |
| 1971- | 1129 (4.9 %) | 610 (2.6 %) | 1739 (7.5 %) |
| Kaikki | 11744 (50.7 %) | 11440 (49.3 %) | 23184 (100 %) |

Taulukko 2. Skitsofreenisten häiriöiden diagnoosilla hoidetut potilaat vuosina 1994-1998 ikä- ja sukupuoliryhmittäin

| | | 31.12.1993 | | | | 31.12.1994 | | | | 31.12.1995 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | % | p% | pr | N | % | p% | pr | N | % | p% | pr |
| -1930 | Mies | 332 | 7.7 | 64 | 1.2 | 311 | 7.1 | 71 | 1.1 | 255 | 5.9 | 71 | 0.9 |
| | Nainen | 649 | 15.1 | 64 | 1.4 | 660 | 15.1 | 68 | 1.4 | 622 | 14.4 | 67 | 1.3 |
| | Yhteensä | 981 | 22.8 | 64 | 1.3 | 971 | 22.2 | 69 | 1.3 | 877 | 20.3 | 68 | 1.2 |
| 1931-1950 | Mies | 808 | 18.8 | 51 | 1.3 | 795 | 18.2 | 56 | 1.2 | 728 | 16.8 | 60 | 1.1 |
| | Nainen | 646 | 15.0 | 39 | 1.0 | 614 | 14.0 | 47 | 0.9 | 595 | 13.8 | 46 | 0.9 |
| | Yhteensä | 1454 | 33.8 | 46 | 1.1 | 1409 | 32.2 | 53 | 1.1 | 1323 | 30.6 | 54 | 1.0 |
| 1951-1970 | Mies | 1091 | 25.3 | 40 | 1.4 | 1111 | 25.4 | 50 | 1.5 | 1183 | 27.4 | 50 | 1.6 |
| | Nainen | 662 | 15.4 | 35 | 0.9 | 689 | 15.7 | 39 | 0.9 | 663 | 15.3 | 38 | 0.9 |
| | Yhteensä | 1753 | 40.7 | 38 | 1.2 | 1800 | 41.1 | 46 | 1.2 | 1846 | 42.7 | 46 | 1.2 |
| 1971- | Mies | 74 | 1.7 | 18 | 0.1 | 137 | 3.1 | 23 | 0.2 | 165 | 3.8 | 25 | 0.2 |
| | Nainen | 46 | 1.1 | 15 | 0.1 | 61 | 1.4 | 26 | 0.1 | 110 | 2.5 | 17 | 0.1 |
| | Yhteensä | 120 | 2.8 | 17 | 0.1 | 198 | 4.5 | 24 | 0.1 | 275 | 6.4 | 22 | 0.2 |
| Kaikki | Mies | 2305 | 53.5 | 47 | 0.9 | 2354 | 53.8 | 53 | 0.9 | 2331 | 53.9 | 53 | 0.9 |
| | Nainen | 2003 | 46.5 | 45 | 0.8 | 2024 | 46.2 | 51 | 0.8 | 1990 | 46.1 | 48 | 0.8 |
| | Yhteensä | 4308 | 100 | 46 | 0.8 | 4378 | 100 | 52 | 0.9 | 4321 | 100 | 51 | 0.8 |

| | | 31.12.1996 | | | | 31.12.1997 | | | | 31.12.1998 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | % | p% | pr | N | % | p% | pr | N | % | p% | pr |
| -1930 | Mies | 211 | 5.3 | 71 | 0.8 | 188 | 5.1 | 67 | 0.7 | 137 | 4.3 | 70 | 0.5 |
| | Nainen | 543 | 13.5 | 71 | 1.2 | 436 | 11.8 | 72 | 0.9 | 360 | 11.4 | 71 | 0.8 |
| | Yhteensä | 745 | 18.9 | 71 | 1.0 | 624 | 16.9 | 71 | 0.8 | 497 | 15.7 | 70 | 0.7 |
| 1931-1950 | Mies | 684 | 17.3 | 57 | 1.1 | 633 | 17.1 | 56 | 1.0 | 512 | 16.2 | 58 | 0.8 |
| | Nainen | 524 | 13.3 | 45 | 0.8 | 502 | 13.6 | 42 | 0.8 | 417 | 13.2 | 41 | 0.6 |
| | Yhteensä | 1208 | 30.6 | 52 | 0.9 | 1135 | 30.7 | 50 | 0.9 | 929 | 29.4 | 50 | 0.7 |
| 1951-1970 | Mies | 1075 | 27.2 | 54 | 1.4 | 994 | 26.9 | 52 | 1.3 | 876 | 27.7 | 53 | 1.2 |
| | Nainen | 639 | 16.2 | 43 | 0.9 | 582 | 15.8 | 42 | 0.8 | 519 | 16.4 | 41 | 0.7 |
| | Yhteensä | 1714 | 43.4 | 50 | 1.1 | 1576 | 42.7 | 49 | 1.1 | 1395 | 44.2 | 49 | 0.9 |
| 1971- | Mies | 171 | 4.3 | 33 | 0.2 | 246 | 6.7 | 28 | 0.3 | 229 | 7.3 | 29 | 0.3 |
| | Nainen | 112 | 2.8 | 27 | 0.1 | 113 | 3.1 | 31 | 0.1 | 107 | 3.4 | 22 | 0.1 |
| | Yhteensä | 283 | 7.2 | 31 | 0.2 | 359 | 9.7 | 29 | 0.2 | 336 | 10.6 | 27 | 0.2 |
| Kaikki | Mies | 2141 | 54.2 | 55 | 0.9 | 2061 | 55.8 | 52 | 0.8 | 1754 | 55.6 | 53 | 0.7 |
| | Nainen | 1809 | 45.8 | 51 | 0.7 | 1633 | 44.2 | 50 | 0.6 | 1403 | 44.4 | 47 | 0.5 |
| | Yhteensä | 3950 | 100 | 53 | 0.8 | 3694 | 100 | 51 | 0.7 | 3157 | 100 | 50 | 0.6 |

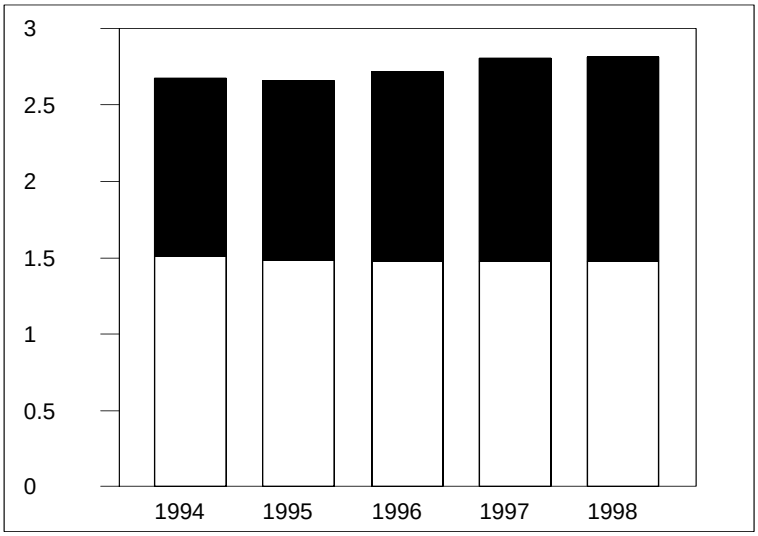
Taulukko 3. Skitsofreenisten häiriöiden potilaiden määrät, suhteelliset määrät, vähintään vuoden laitoksessa olleiden osuudet sekä prevalenssit 1000 asukasta kohden vuosien 1993-1998 viimeisinä päivinä

Ryhmien ominaisuuksien havainnollistamiseksi on taulukkoon 3 kerätty kunkin ryhmän koko, suhteellinen osuus laitoksessa olevista, vähintään vuoden laitoksessa olleiden osuus ja laitoksessa olemisen prevalenssi (tuhatta asukasta kohden) vuosien 1994-1998 viimeisinä päivinä. Taulukkoa tarkastellessa tulee ottaa huomioon, että ryhmien välisiä siirtymisiä ei tapahdu eli taulukon luvut kuvaavat myös kohorttien muutoksia eivätkä pelkkää laitosjärjestelmää.

Taulukon perusteella voidaan havaita, että nuorimman ikäryhmän (vuoden 1970 jälkeen syntyneet) osuus on kasvanut voimakkaasti ja vanhusten osuus on puolestaan pienentynyt nopeasti; muiden ryhmien osuudet ovat pysyneet kohtalaisen vakaina. Tämä voidaan tulkita niin, että nuorin ikäryhmä saavuttaa tarkasteluajanjakson kuluessa skitsofrenian puhkeamisen kannalta kriittisimmän iän ja vanhustenryhmässä puolestaan kuolevuus on suurta. Vähintään vuoden yhtäjaksoisesti laitoksessa olleiden potilaiden osuus ikäryhmän kaikista laitoshoitoa saaneista kasvaa varsin johdonmukaisesti iän lisääntyessä (vrt. Hakkarainen 1980, 35-37). Pitkäaikaispotilaiden osuus on pysytellyt tarkasteluajanjakson kuluessa noin 50 %:ssa, vaikka potilaiden absoluuttinen lukumäärä onkin selvästi vähentynyt. Laitoksessa olemisen prevalenssi vaikuttaisi hiukan pienentyneen tarkastelujakson aikana. Prevalenssi poikkeaa selvästi alaspäin nuorimmassa ikäryhmässä, mutta muiden ikäryhmien väliset erot ovat yllättävän pieniä; syntymävuoteen perustuvan ikäryhmittelyn sovittaminen vuoden viimeisten päivien väestötietoihin saattaa aiheuttaa ikäryhmittäisiin prevalensseihin vähäistä harhaa. Miehillä prevalenssit ovat vanhinta ikäryhmää lukuunottamatta naisia suurempia (vrt. Hakkarainen 1989, 126-127).

### 4.1.4 Volyymit ja ottamistaajuudet

Kuviossa 5 vuosien 1994-1998 aikana alkaneiden hoitojaksojen määrät on suhteutettu tuhatta asukasta kohden; pylvään alempi osa kuvaa potilaiden vuoden ensimmäisten sisäänottojen osuutta kaikista hoitojaksoista. Ottamistaajuus on kasvanut hiukan tarkasteluajanajakson kuluessa; vuoden ensimmäisten sisäänottojen ottamistaajuus on kuitenkin pysynyt vakiona.




Kuvio 5. Kaikki alkaneet hoitojaksot sekä potilaiden vuoden ensimmäiset hoitojaksot (valkoinen osa) 1000 asukasta kohden vuosina 1994-1998

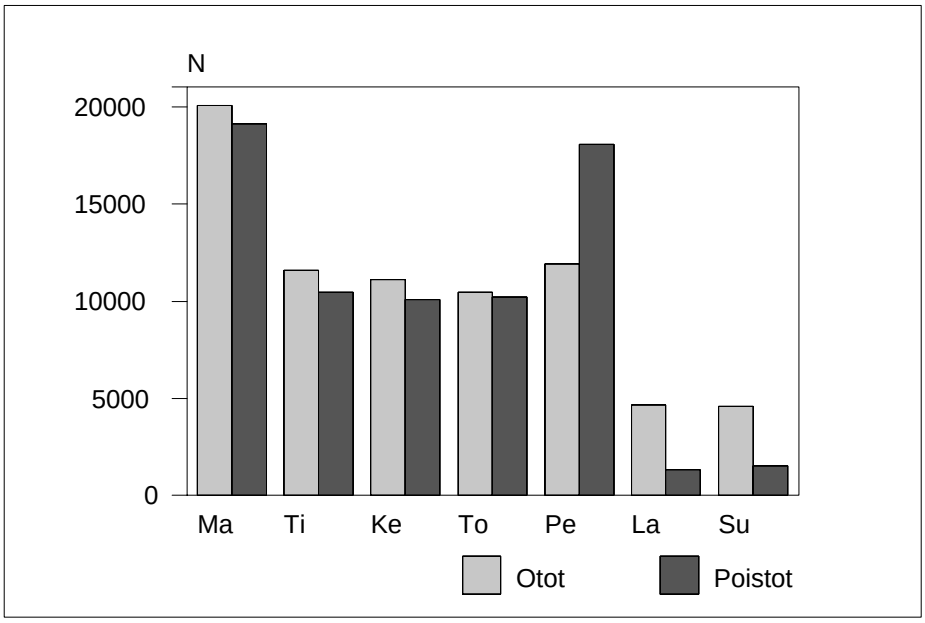
| | | 1994 | | 1995 | | 1996 | | 1997 | | 1998 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | 1. | N | 1. | N | 1. | N | 1. | N | 1. |
| | Mies | 330 | 228 | 424 | 292 | 486 | 333 | 776 | 486 | 789 | 499 |
| -1930 | Nainen | 178 | 121 | 218 | 167 | 300 | 200 | 397 | 234 | 480 | 260 |
| | Yhteensä | 508 | 349 | 642 | 459 | 786 | 533 | 1173 | 720 | 1269 | 759 |
| 1931 | Mies | 3603 | 1957 | 3698 | 1994 | 3792 | 2014 | 3955 | 2008 | 4002 | 1988 |
| - | Nainen | 2854 | 1571 | 2811 | 1560 | 2846 | 1502 | 3075 | 1540 | 3141 | 1562 |
| 1950 | Yhteensä | 6457 | 3528 | 6509 | 3554 | 6638 | 3516 | 7030 | 3548 | 7143 | 3550 |
| 1951 | Mies | 2570 | 1385 | 2564 | 1287 | 2556 | 1242 | 2462 | 1209 | 2388 | 1179 |
| - | Nainen | 2555 | 1474 | 2482 | 1398 | 2539 | 1397 | 2550 | 1379 | 2595 | 1389 |
| 1970 | Yhteensä | 5125 | 2859 | 5046 | 2685 | 5095 | 2639 | 5012 | 2588 | 4983 | 2586 |
| | Mies | 464 | 259 | 414 | 243 | 424 | 249 | 364 | 226 | 321 | 175 |
| 1971- | Nainen | 1070 | 682 | 990 | 646 | 995 | 638 | 862 | 527 | 803 | 538 |
| | Yhteensä | 1534 | 941 | 1404 | 889 | 1419 | 887 | 1226 | 753 | 1124 | 713 |
| | Mies | 6967 | 3829 | 7100 | 3816 | 7258 | 3838 | 7557 | 3929 | 7500 | 3859 |
| Kaikki | Nainen | 6657 | 3848 | 6501 | 3771 | 6680 | 3737 | 6884 | 3680 | 7019 | 3749 |
| | Yhteensä | 13624 | 7677 | 13601 | 7587 | 13938 | 7575 | 14441 | 7609 | 14519 | 7608 |

Taulukko 4. Laitokseen otettujen potilaiden volyymit ja ensimmäisten sisäänottojen määrät ikä- ja sukupuoliryhmittäin vuosina 1994-1998

Taulukossa 4 on esitetty vuosien 1994-1998 aikana alkaneiden hoitojaksojen määrät ikä- ja sukupuoliryhmittäin. Taulukon mukaan vuoden ensimmäisten hoitojaksojen osuus on hiukan pienentynyt vuodesta 1994. Myös tämä ilmentää hoitokäytäntöjen muuttumista kohti lyhytkestoisempaa, mutta useammin toistuvaa hoitoa. Vanhimmassa ja nuorimmassa ikäryhmässä vuoden ensimmäisten hoitojaksojen osuudet ovat suurempia kuin muissa ryhmissä. Vaikuttaisi myös siltä, että naisilla ensimmäisten hoitojaksojen osuudet ovat miehiä suurempia. Vähäisen hoidon toistuvuuden voidaan tulkita ilmentävän joko suurta tarvetta pitkäaikaishoitoon tai vaihtoehtoisesti hoidon tarpeettomuutta. Esimerkiksi vanhuksilla kuolema voi muuttaa pitkäaikaishoidon tarpeen hoidon tarpeettomuudeksi.

Sisäänotto- ja poistamisvolyymit eivät ole jakautuneet tasaisesti vuoden jokaiselle päivälle. Kuviossa 6 on esitetty kuinka hoitojaksojen alku- ja loppupäivät jakautuvat eri viikonpäiville. Viikonloppuna on niin sisäänottojen kuin poistojenkin osalta hiljaisinta, mutta ottojen ja poistojen vähäinen määrä kompensoituu perjantain ja maanantain poikkeavan suurilla volyymeilla. Myös laitokseen ensi kertaa tulevilla sisäänottojen viikonpäivittäinen jakauma on muodoltaan täysin samanlainen; hoidon tarve ei siis ensikertalaisillakaan liene kovin usein akuuttia hoitoa vaativaa. Syyt havaittuihin hoitoonotto- ja poistojakaumiin ovat ilmeisesti käytännöllis-hallinnollisia. Toisin sanoen maanantai muita yleisempänä sisäänottopäivänä saattaa johtua siitä, että viikonlopun aikana tulleet henkilöt sisäänkirjoitetaan vasta maanantaina esimerkiksi päivystysosastolta varsinaiselle osastolle siirron jälkeen. Toisaalta skitsofreenikko saattaa hyvin pärjätä viikonlopun omaistensa hoivissa kotona. Ilmiö näkyy aineistosta perjantain suurena poistomääränä ja vastaavasti maanantain suurena sisäänottomääränä. Lisäksi on yleisesti tunnettu tosiasia, että maanantaina sairaalaan pääsy onnistuu helpommin kuin viikonloppuna. Näin ollen hoidon alkua tai laitoksesta toiseen siirtoa voidaan käytännöllisistä syistä hyvin lykätä, jos potilas ei ole välittömän hoidon tarpeessa.

Päiväkohtaisissa sisäänotto- ja poistamisvolyymeissa on liikaa hajontaa säännönmukaisuuksien hahmottamiseksi. Volyymien aggregointi viikoille, kuukausien päiville sekä kuukausille ei paljastanut suuria epäsäännöllisyyksiä volyymeissa; jakaumat olivat yleisten loma-aikojenkin osalta hyvin tasaisia. Havaittavan piikin




Kuvio 6. Sisäänotot ja poistot viikonpäivittäin

sisäänottovolyymeihin kuitenkin tekivät kuukauden ensimmäinen päivä sekä vuoden ensimmäinen viikko.

Kuten jo aikaisemmin todettiin (ks. sivut 57-58), voidaan vuoden 1997 alun jälkeen laitokseen ensi kertaa tulleita pitää todellisina uusina potilaina. Hoitoilmoitusrekisterin tiedoista pystytään määrittämään, kuinka monta ensi kertaa hoitoon tulevaa potilasta kunakin päivänä laitokseen on otettu. Tässä tapauksessa molemmilta käyttökelpoisilta vuosilta (1997 ja 1998) saatiin 365 havaintoa kullekin ikä- ja sukupuoliryhmälle. Analyysien perusteella havaittiin, että päivässä saapuvien potilaiden määrien jakaumat olivat joka ryhmässä molempina vuosina hyvin samankaltaisia. Lisäksi osoittautui, että saapuvien potilaiden määrät noudattavat joka ryhmässä riittävällä tarkkuudella Poisson-jakaumaa. Taulukkoon 5 on koottu poisson-jakaumien estimoidut parametrit ja $\chi^2$-yhteensopivuustestien P-arvot. Poisson-oletus jää kyseenalaiseksi vain naisten toiseksi nuorimmalla ikäryhmällä; visuaalisten tarkastelujen perusteella Poisson-oletus on siinäkin tapauksessa järkevä.

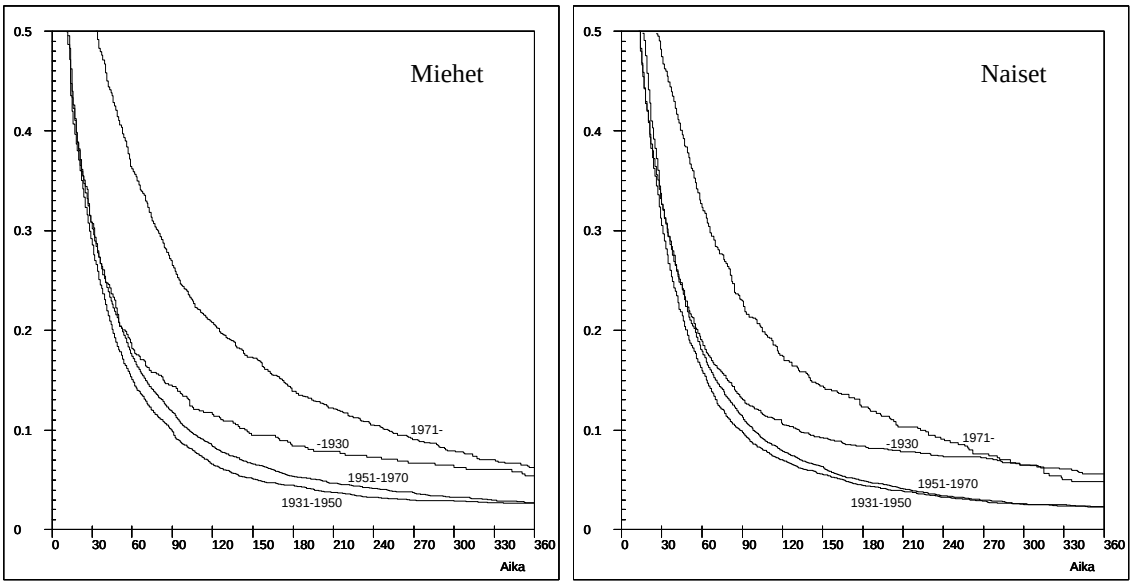
| | 1997 | | | | 1998 | | | |
|---|---|---|---|---|---|---|---|---|
| *N=365* | **Mies** | | **Nainen** | | **Mies** | | **Nainen** | |
| **Syntymävuosi** | $\lambda$ | $P$ | $\lambda$ | $P$ | $\lambda$ | $P$ | $\lambda$ | $P$ |
| -1930 | 0.30 | 0.45 | 0.61 | 0.17 | 0.19 | 0.79 | 0.64 | 0.01 |
| 1931-1950 | 1.10 | 0.39 | 1.34 | 0.07 | 0.98 | 0.70 | 1.13 | 0.17 |
| 1951-1970 | 1.74 | 0.03 | 1.38 | 0.02 | 1.50 | 0.13 | 1.17 | 0.01 |
| 1971- | 0.73 | 0.25 | 0.34 | 0.76 | 0.56 | 0.02 | 0.30 | 0.87 |

Taulukko 5. Sovitettujen Poisson-jakaumien parametrit ja yhteensopivuustestien P-arvot

### 4.1.5 Hoito- ja uusiutumisaikajakaumat

Aineiston pitkittäisulottuvuus lisää sen heterogeenisuutta. Kuten kuviosta 4 havaittiin on laitoshoidossa olevien skitsofreenikoiden määrä jatkuvasti vähentynyt. Suoritettujen tarkastelujen perusteella myös hoitoajoissa ilmenee vastaavaa säännönmukaisuutta; esimerkiksi eri vuosina alkaneiden hoitojaksojen hoitoaikajakaumaa kuvaavat Kaplan-Meier-estimaattorit poikkeavat järjestelmällisesti toisistaan. Vuosina 1996-1998 alkaneiden hoitojaksojen pituuden mediaani oli kuitenkin kunakin vuonna kaksi viikkoa, joten valtaosa hoitojaksoista on melko lyhyitä. Keskimääräinen hoitoaika rajoitettuna yhteen vuoteen (yli vuoden pituiset hoitojaksot on typistetty vuoden pituisiksi) oli vuonna 1996 49 vuorokautta, vuonna 1997 46 vuorokautta ja vuonna 1995 43 vuorokautta eli hoitoaikajakaumat ovat voimakkaasti oikealle vinoja. Edellä todetut Kaplan-Meier-estimaattoreiden väliset erot korostuvat vasta yli kahden viikon pituisten hoitojaksojen osalta. Hoitoaikojen systemaattisen lyhentymisen vuoksi hoitoaikajakaumien määrittämiseen käytetään seuraavissa tarkasteluissa vain vuosien 1997 ja 1998 tietoja. Molempien vuosien tietojen käyttäminen perusteltua, sillä kyseisten vuosien Kaplan-Meier-estimaattorit sisältyvät kumpikin toistensa 95 % luottamusväleihin. Jakaumien samankaltaisuutta testaavan yleistetyn Wilcoxonin testinkään mukaan merkitsevää eroa ei ole (P=0.77).

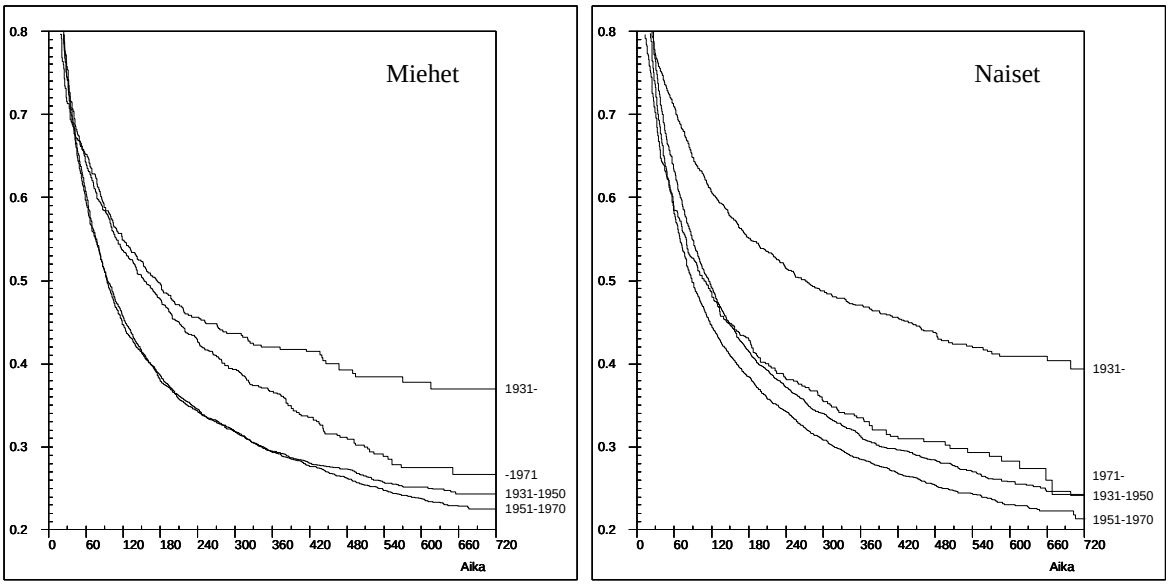
Kuviossa 7 on esitetty oleelliset osat ryhmittäisiä hoitoaikajakaumia kuvaavista Kaplan-Meier-estimaattoreista. Muissa paitsi nuorimmassa ikäryhmässä yli puolet hoitojaksoista jää alle kahden viikon mittaisiksi. Muutenkin laitoksesta poistuminen on hitainta nuorimmalla ikäryhmällä alle puolen vuoden pituisissa hoitojaksoissa, mutta hoidon pidetessä laitoksessa olevien osuus lähestyy muita ryhmiä. Vähintään vuoden pituisia




Kuvio 7. Kaplan-Meier-estimaattorin mukaiset todennäköisyydet viipyä laitoksessa tietty määrä päiviä

hoitojaksoja on vanhuksilla sekä nuorilla noin 6 % ja muilla noin 3 % kaikista ryhmän hoitojaksoista. Naisten hoitojaksot ovat yleisesti ottaen hiukan miesten hoitojaksoja lyhyempiä; suurimmillaan ero on nuorimmassa ikäryhmässä.

Koska skitsofreenikoilla on lähes poikkeuksetta useita uusintakäyntejä laitoksessa, on tarkoituksenmukaista analysoida myös hoidon uusiutumisen todennäköisyyksiä. Tarkastelemalla hoitoilmoitusrekisterin tietojen avulla potilaan kahden peräkkäisen hoitojakson loppu- ja alkupäiviä pystytään päättelemään hoitoajan lisäksi hoitojaksojen välinen niin sanottu hoidon uusiutumisen odotusaika. Kuviossa 8 on esitetty ryhmittäiset uusiutumisaikajakaumat Kaplan-Meier-estimaattoreiden avulla. Vanhuksista palaa uudestaan laitokseen kaikkein pienin osuus (noin 60 %), mikä johtunee ryhmän suuresta kuolleisuudesta. Myös nuorin ikäryhmä poikkeaa jälleen selvästi ’keski-ikäisistä’, joista yli 75 % joutuu uudestaan laitokseen kahden vuoden kuluessa poistamisesta. Naiset uusivat miehiä hitaammin.

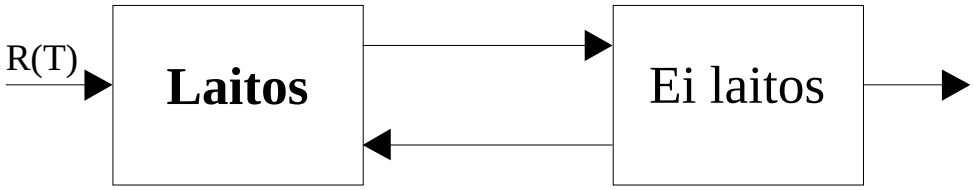



Kuvio 8. Kaplan-Meier-estimaattorin mukaiset todennäköisyydet joutua uudestaan laitokseen tietyn päivän kuluttua edellisen hoidon päättymisestä

## 4.2 Potilaspopulaatiota kuvaava dynaaminen malli

Tarkasteltavan potilaspopulaation dynamiikkaa ilmentävät tekijät on sidottava sopivalla tavalla toisiinsa, jotta olisi mahdollista muodostaa populaatiota kuvaava malli vuorovaikutusten kokonaisvaltaiseksi hahmottamiseksi. Skitsofreenisten häiriöiden potilaiden tapauksessa dynaamisen järjestelmän teoreettiseksi pohjaksi sopii Hakkaraisen (1985; 1989) esittämä psykiatrisen potilasväestön teoreettinen muutostapahtumakenttä. Hakkaraisen kenttä on kuitenkin tämän tutkimuksen tarpeita ja käytettävissä olevaa aineistoa ajatellen turhan monimutkainen, sillä esimerkiksi avohoidosta ei ole käytettävissä tarvittavia tietoja. Von Korffin (1978) esittämä malli on taas oletuksiltaan epärealistinen, sillä siinä ei oteta huomioon järjestelmään tulevia eikä siitä poistuvia yksilöitä riittävällä tarkuudella. Tässä tutkimuksessa käytettävä malli sijoittuu kahden edellämainitun ehdotelman välimaastoon ja sen verkkoesitys on kuviossa 9.

Mallissa on eritelty laitoksessa olevat ja sieltä poistetut potilaat. Mahdollisia virtaamia ovat järjestelmän ulkopuolelta laitokseen ensimmäisen kerran tulevat potilaat, laitoksesta poistettavat ja sinne uudestaan joutuvat potilaat sekä potilaat, jotka eivät




Kuvio 9. Kaksitilainen avoin malli

laitoksesta poistamisen jälkeen enää uusi hoitoa. Mallissa siis oletetaan, että hoito- ja uusiutumisaikajakaumat ovat riippumattomia hoito- tai poistokerrasta. Lisäksi laitoksesta kuoleman kaltaisista erityissyistä poistuvat potilaat on sisällytetty hoitoa uusimattomien potilaiden joukkoon ja järjestelmään tulevat uudet yksilöt joutuvat kaikki laitokseen.

Edellä kuvattu malli saadaan formalisoitua semi-Markov-malliksi luvussa 2 esitetyllä tavalla. Numeeristen tulosten saamiseksi semi-Markov-malli on tässä tutkimuksessa palautettu Markovin ketjuksi laajentamalla siirtymämatriisia (ks. kohta 2.3.2). Alustavien analyysien perusteella tarvittavat hasardifunktiot alkavat tasaantua vasta noin vuoden pituisen hoidon jälkeen. Näin ollen päiväkohtainen aineisto johtaa ulottuvuuksiltaan niin suuriin siirtymämatriiseihin, ettei niitä kannata käyttää nykyisillä mikrotietokoneilla ilman erikoistoimenpiteitä. Lisäksi päivätasoisesta aineistosta ei-parametrisesti estimoitujen hasardien satunnaisvaihtelu on varsin suurta etenkin pitemmillä hoitoajoilla, joten estimoinnissa olisi tarkoituksenmukaista hyödyntää sopivia tasoitusmenetelmiä (ks. esim. González-Mantega ym. 1996; Härdle 1991; Wu 1989).

Edellä mainituista syistä tämän tutkimuksen aineisto luokiteltiin uudestaan siten, että perusaikayksiköksi tuli päivän sijaan viikko, vaikka teoreettisessa mielessä siitä saattaa seurata hankaluuksia (ks. sivu 37). Alustavien analyysien perusteella päädyttiin siirtymämatriisiin, jossa laitostila jaettiin 50 aikasiivuun ja ei-laitostila 49 siivuun; lisäksi varattiin yksi siivu järjestelmästä poistuneita varten. Toisin sanoen muodostetussa mallissa poistamisen hasardi on 49 viikkoa vapaasti vaihteleva ja viikosta 50 alkaen vakio. Uusiutumisajan suhteen uusiminen on mahdollista 49 viikkoa, jonka jälkeen yksilöt tulkitaan poistuneiksi järjestelmästä. Tämä poistumisoletus on ristiriidassa uusien skitsofreniatapausten ”kolmen vuoden säännön” kanssa (ks. sivut 57-58), joten säännön rikkominen on huomioitava myös järjestelmään tulijoiden määrissä. Uuden oletuksen

| Syntymävuosi | Mies $\lambda$ | Nainen $\lambda$ |
|---|---|---|
| -1930 | 2.5 | 7 |
| 1931-1950 | 12.9 | 15.4 |
| 1951-1970 | 20.5 | 16.8 |
| 1971- | 6 | 3.1 |

Taulukko 6. Mallissa käytettyjen Poisson-jakaumien parametrit

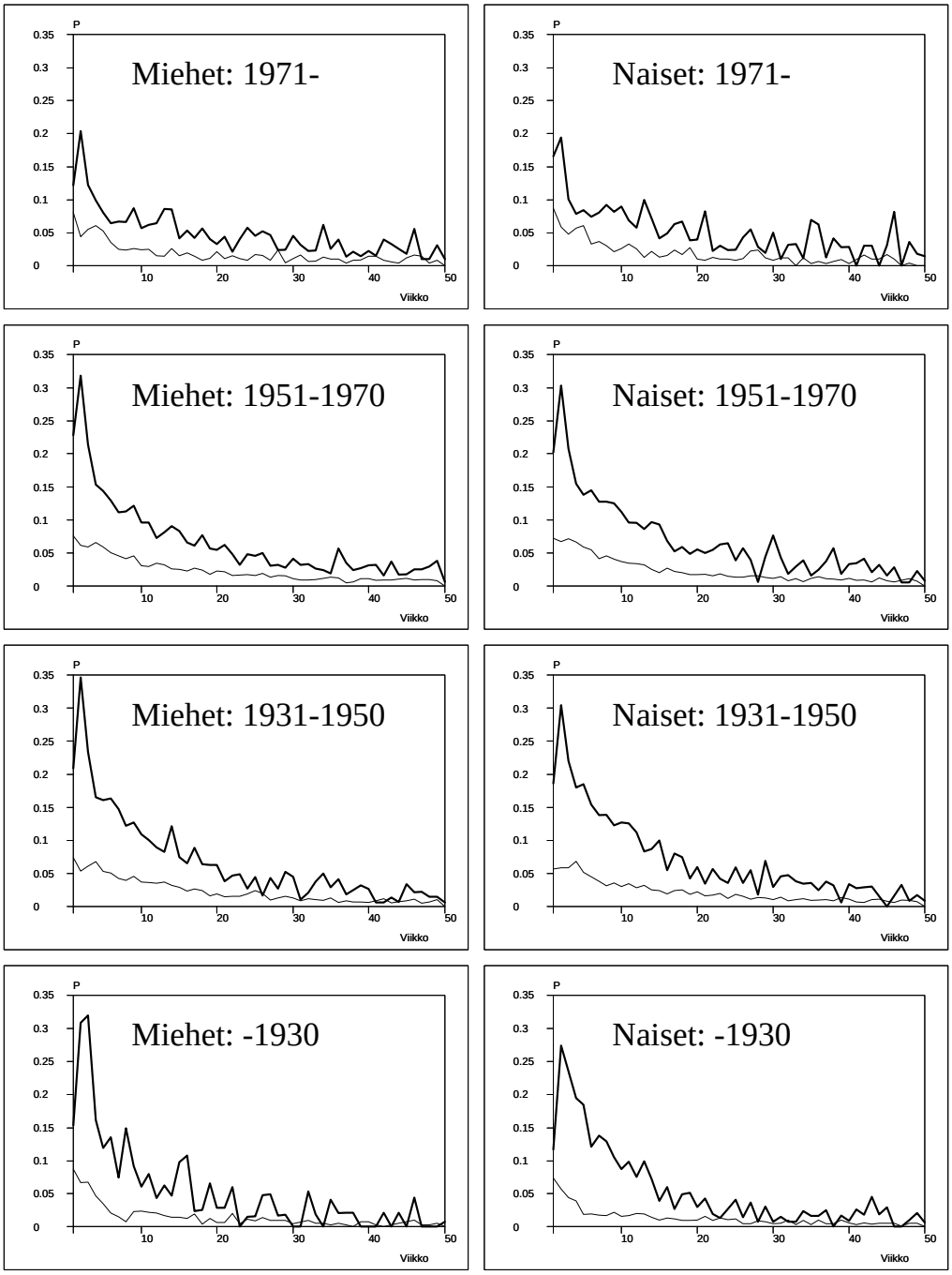
mukainen sisääntulijoiden lukumäärä noudatti viikkopohjaisessa aineistossa riittävällä tarkkuudella Poisson-jakaumaa jokaisessa ikä- ja sukupuoliryhmässä. Analyyseissa käytettyjen Poisson-jakaumien parametrit ovat taulukossa 6.

Loppujen lopuksi malli siis muodostuu kahdeksasta toisistaan riippumattomasta rinnakkaismallista (ikä- ja sukupuoliryhmät), joista jokaista kuvaa $100 \times 100$ -siirtymämatriisi sekä Poisson-jakautunut sisääntulijoiden määrä.

### 4.2.1 Mallin estimointi ja validiteetti

Mallille estimoitiin tarvittavat viikkokohtaiset hasardit käyttäen kohdassa 2.3.3 esiteltyjä menetelmiä. Sekä Kaplan-Meier- että kohorttiestimaattorit antoivat hyvin toisiaan vastaavat tulokset. Vähintään 50 viikkoa laitoksessa olleiden hasardi osoittautui mallin antamiin tuloksiin kaikkein herkimmin reagoivaksi parametriksi ja sen estimointi oli kätevintä suorittaa eri luokituksiin hyvin soveltuvalla kohorttiestimaattorilla.

Estimoidut hasardifunktiot on esitetty kuviossa 10. Paksumpi viiva kuvaa laitoksesta poistumisen hasardia ja ohuempi viiva puolestaan laitokseen takaisin joutumisen hasardia. Laitoksesta poistamisen hasardifunktiot saavat suurimmat arvonsa toisella viikolla, jonka jälkeen ne ovat muodoltaan laskevia. Toisin sanoen mitä kauemmin laitoksessa on oltu, sitä kauemmin siellä vielä tullaan olemaankin. Tulos on sopusoinnussa aikaisempien skitsofreniatutkimusten kanssa (ks. esim. Brown 1960; Reuter 1981; Reuter & Von Korff 1980). Vastaavasti hoitokertojen välisen odotusajan kasvaessa todennäköisyys joutua uudestaan laitokseen pienenee.

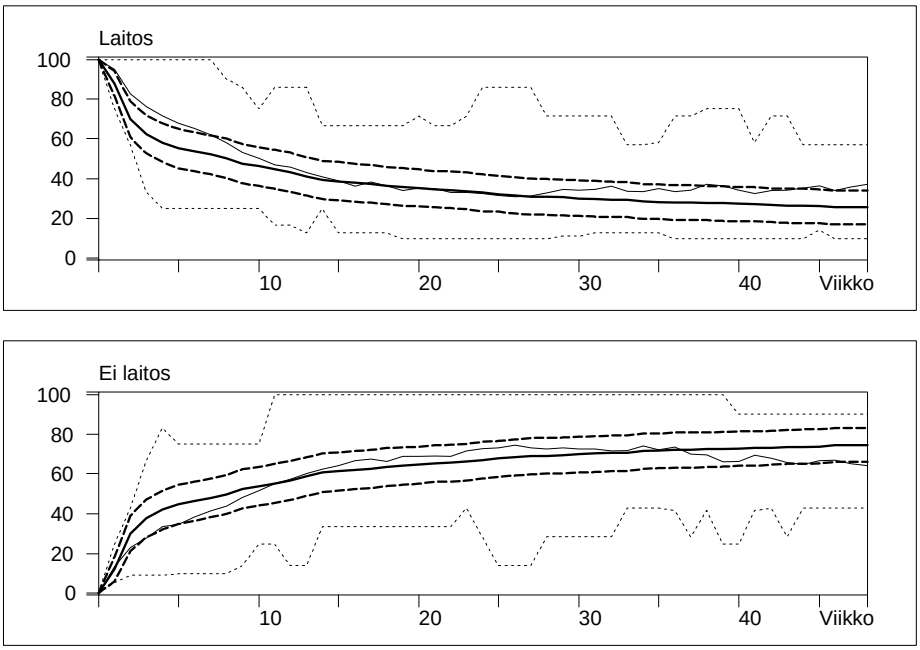



Kuvio 10. Ryhmittäiset hasardifunktiot laitoksesta poistumiselle (paksumpi viiva) sekä laitokseen takaisin joutumiselle

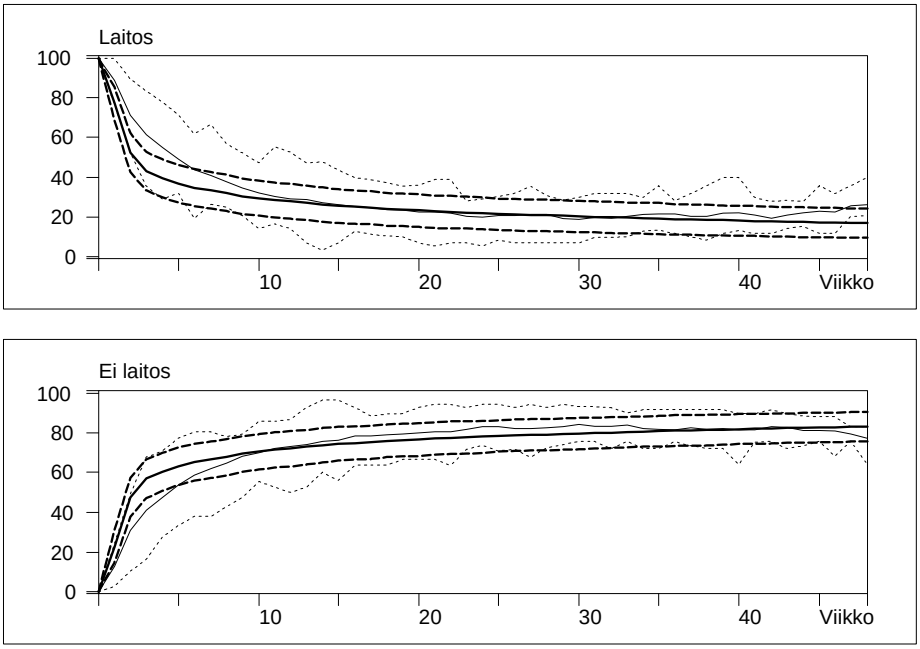
Muodostetun semi-Markov-mallin validiteetin tarkastelemiseksi laskettiin jokaiselle ikä- ja sukupuoliryhmälle mallin antama ennuste ja 95%:n luottamusvälit sadan potilaan kohortin liikkeistä 48 viikon pituiselle ajalle. Tämän jälkeen aineistoa käyttäen seurattiin vuoden 1997 jokaisen viikon laitokseen otettujen potilaskohorttia 48 viikkoa. Kunkin kohortin potilaiden määrä skaalattiin vastaamaan sataa potilasta, jolloin seurantaviikottaisista jakaumista pystyttiin laskemaan jokaisen viikon minimi, maksimi ja keskiarvo. Saadut tulokset on esitetty kuvioissa 11-18.

Kuvioista havaitaan, että mallin antama ennuste pysyy kohorteista laskettujen minimi- ja maksimiarvojen välissä kohtalaisen hyvin; vain ”keski-ikäisten” ryhmissä ennusteen luottamusväli karkaa alkuviikoilla havaittujen rajojen ulkopuolelle. Ensimmäisten viikkojen ennusteet ovat myös keskiarvojen suhteen systemaattisesti harhaisia. Muut poikkeamat keskiarvoista johtunevat skaalaamisen aiheuttamasta epätarkkuudesta pienten potilaskohorttien osalta, sillä mainittuja poikkeamia esiintyy vain kooltaan pienimmillä ikä- ja sukupuoliryhmillä.

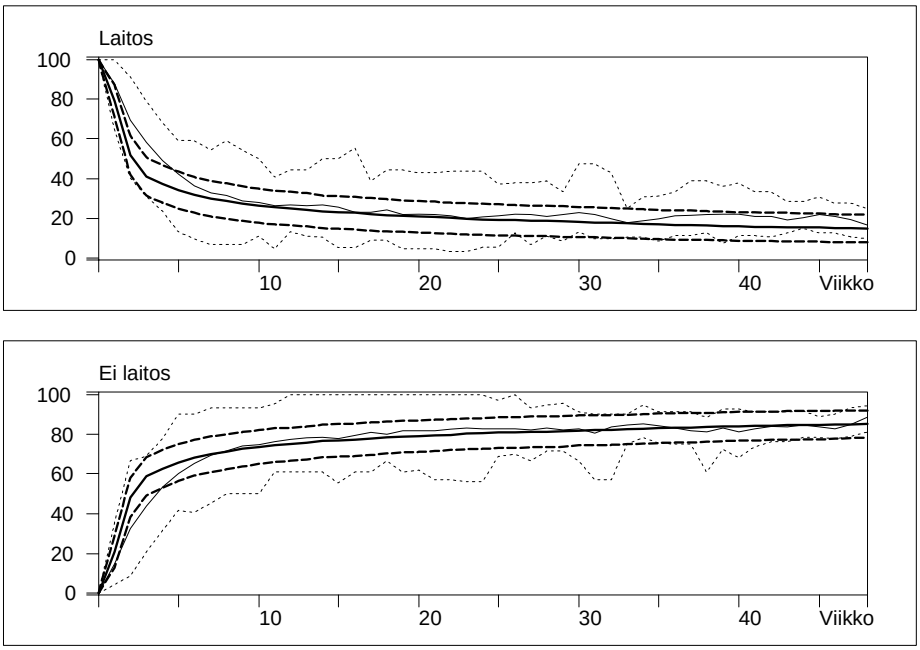
Yhteenvetona voidaan todeta, että malli ennustaa yksinkertaisuudestaan huolimatta populaation liikkeitä yllättävänkin hyvin. Havaittujen ensimmäisten viikkojen poikkeamien perusteella malli ei kuitenkaan pysty kuvaamaan toistuvaa erittäin lyhytaikaista hoitoa vaativien potilaiden liikkeitä täysin todellisuutta vastaavalla tavalla. Kyseisten potilaiden määrät ovat kuitenkin varsin pieniä ja ennusteen virhe tasoittuu hyvin nopeasti, joten malli vaikuttaa riittävän hyvältä käyttötarkoitukseensa.



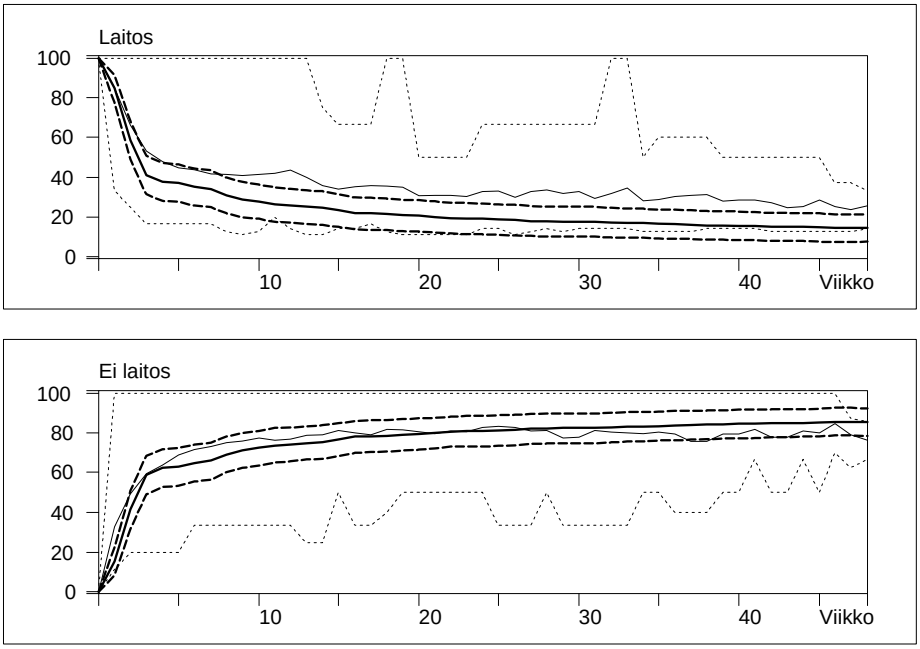

Kuvio 11. Semi-Markov-mallin kohorttiennusteet ja 95%:n luottamusvälit (paksummat viivat) sekä havaittujen kohorttien min, max ja keskiarvo; miehet 1971-



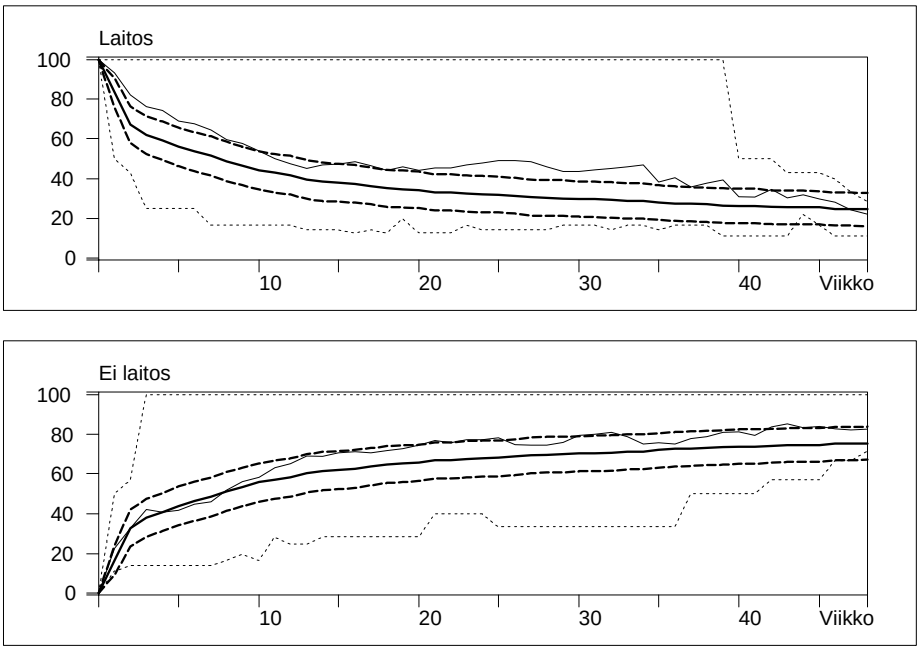

Kuvio 12. Semi-Markov-mallin kohorttiennusteet ja 95%:n luottamusvälit (paksummat viivat) sekä havaittujen kohorttien min, max ja keskiarvo; miehet 1951-1970



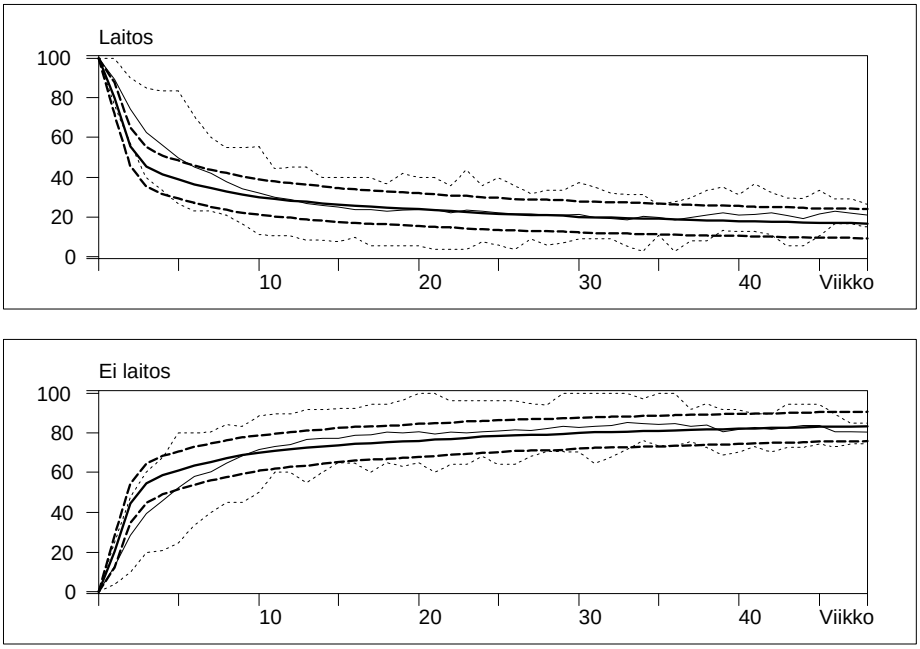

Kuvio 13. Semi-Markov-mallin kohorttiennusteet ja 95%:n luottamusvälit (paksummat viivat) sekä havaittujen kohorttien min, max ja keskiarvo; miehet 1931-1950



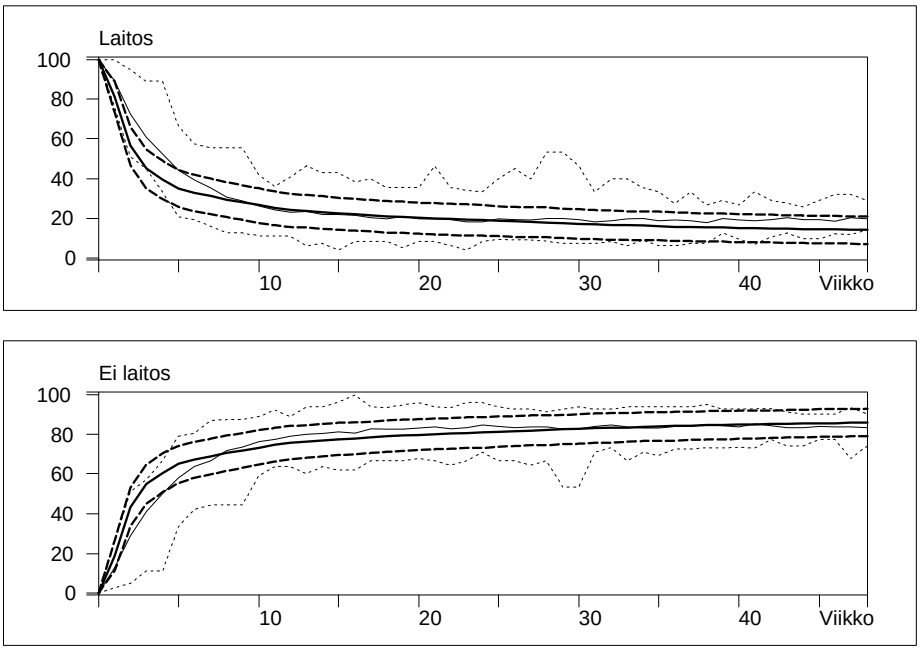

Kuvio 14. Semi-Markov-mallin kohorttiennusteet ja 95%:n luottamusvälit (paksummat viivat) sekä havaittujen kohorttien min, max ja keskiarvo; miehet -1930



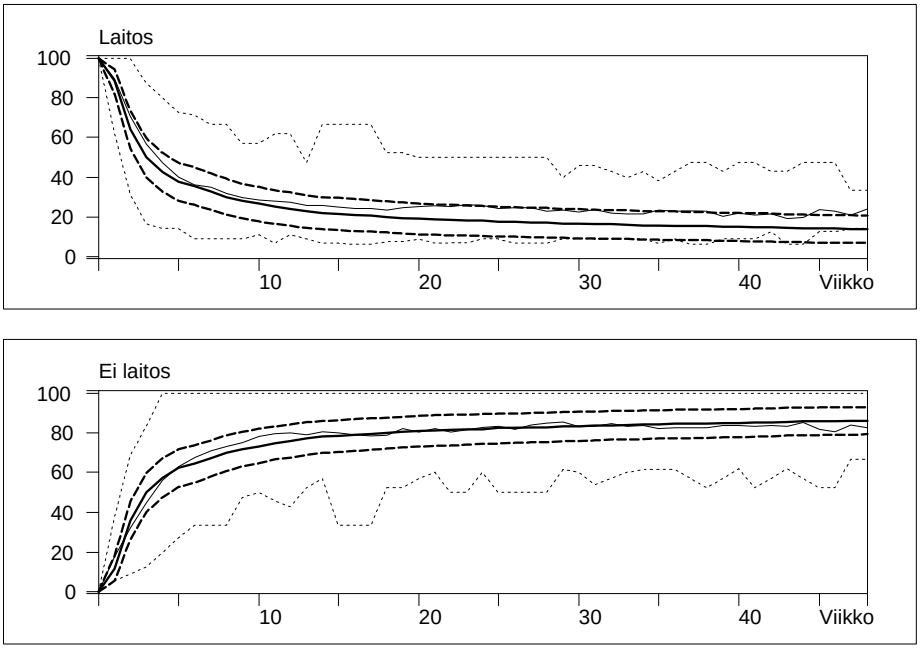

Kuvio 15. Semi-Markov-mallin kohorttiennusteet ja 95%:n luottamusvälit (paksummat viivat) sekä havaittujen kohorttien min, max ja keskiarvo; naiset 1971-




Kuvio 16. Semi-Markov-mallin kohorttiennusteet ja 95%:n luottamusvälit (paksummat viivat) sekä havaittujen kohorttien min, max ja keskiarvo; naiset 1951-1970



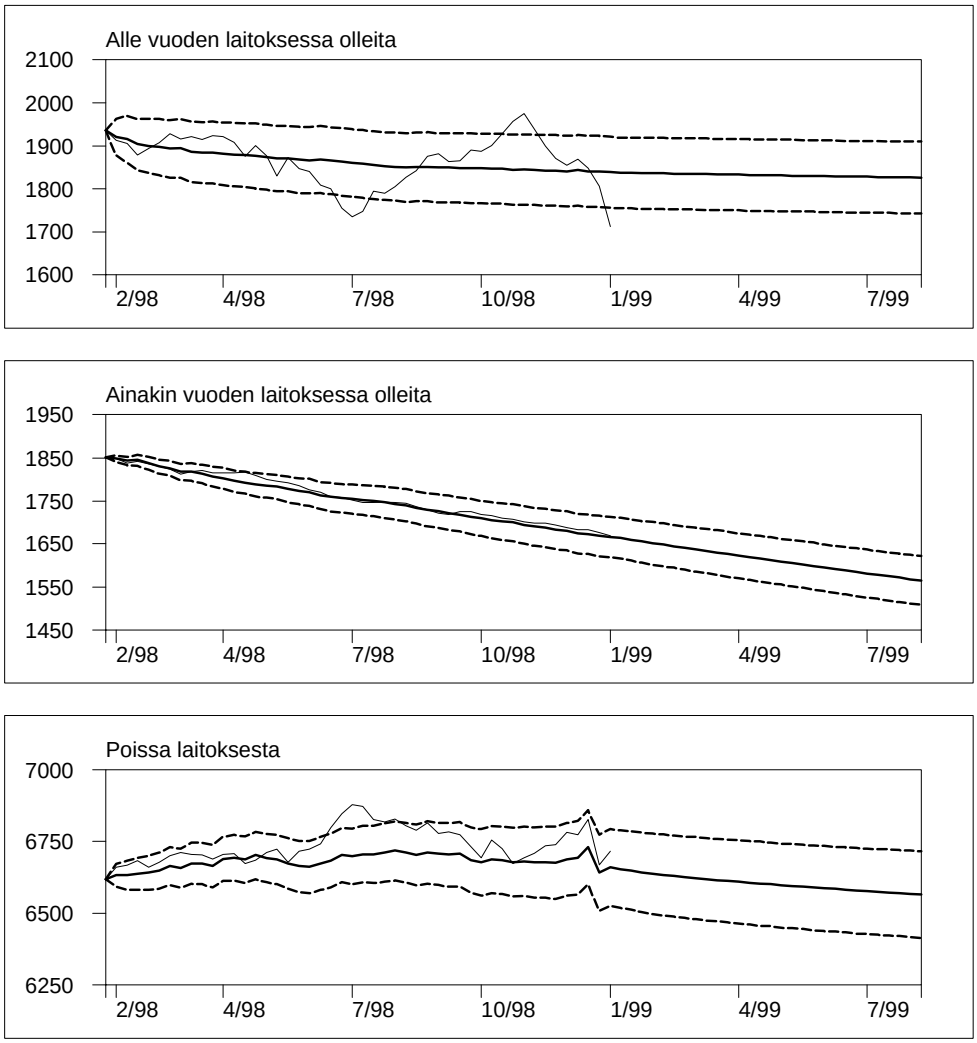

Kuvio 17. Semi-Markov-mallin kohorttiennusteet ja 95%:n luottamusvälit (paksummat viivat) sekä havaittujen kohorttien min, max ja keskiarvo; naiset 1931-1950




Kuvio 18. Semi-Markov-mallin kohorttiennusteet ja 95%:n luottamusvälit (paksummat viivat) sekä havaittujen kohorttien min, max ja keskiarvo; naiset -1930

Pelkkien ryhmittäisten kohorttitarkastelujen lisäksi on kiinnostavaa testata, kuinka hyviä ennusteita koko malli antaa tilanteessa, jossa järjestelmän alkutila on mielivaltainen ja sen piiriin tulevat yksilöt otetaan huomioon. Kuviossa 19 on esitetty mallin ennuste ja 95%:n luottamusvälit, kun alkutilaksi on valittu vuoden 1998 kolmas viikko. Kuvioon on piirretty myös todellisen populaation liikkeet. Laitoksessa olijat on jaettu kahteen osaan, joiden voidaan tulkita kuvaavan lyhyt- ja pitkäaikaisia potilaita. Teknisessä mielessä alle vuoden laitoksessa olleiden määrä on siirtymämatriisin laitostilan aikasiivuissa 1-49 olevien yksilöiden yhteenlaskettu määrä, ainakin vuoden laitoksessa olleiden määrä vastaa laitostilan aikasiivua 50 ja poissa laitoksesta on puolestaan ei-laitostilassa olevien yksilöiden yhteenlaskettu määrä. Järjestelmän piiristä poistuneiden yksilöiden määriä ei ole tässä tapauksessa pidetty kiinnostavana informaationa ja se on jätetty esittämättä. Kuten kuviosta havaitaan, mallin antama ennuste seuraa todella hyvin populaation havaittuja liikkeitä. Alle vuoden laitoksessa olleiden potilaiden populaation havaituissa varannoissa esiintyy kuitenkin runsaasti kausivaihtelua, joten mallin lyhyen aikavälin ennusteen osuvuus riippuu alkutilaksi valitusta viikosta.

Taulukossa 6 esitettyjä Poisson-jakaumien parametrejä käytettäessä lyhytaikaisten potilaiden varanto tasaantuu ajan myötä raja-arvoon 1 769 ja laitoksesta poissaolijoiden varanto puolestaan 6 217 potilaaseen. Molemmat raja-arvot ovat suhteellisen lähellä kuviossa esitetyllä aikavälillä olevia varantoja. Sisäänoton Poisson-oletuksen nojalla raja-arvojen hajonnat ovat arvojen neliöjuuret. Pitkäaikaisten potilaiden varannon tasaantuminen on huomattavasti hitaampaa raja-arvon ollessa 1 039. Jos toisin sanoen sisäänotettavien potilaiden määrä pysyy suhteellisen tasaisena, pitkäaikaispotilaiden määrä tulee nykyisillä hoitokäytännöillä vähenemään reiluun tuhanteen potilaaseen ja lyhytaikaisten määrä pysyy lähes samalla tasolla kuin vuonna 1998.




Kuvio 19. Semi-Markov-mallin mukainen varantojen ennuste ja 95%:n luottamusvälit sekä havaitut varannot

# 5 Pohdintaa

Tässä tutkimuksessa on kuvattu dynaamisten potilaspopulaatioiden mallintamiseen soveltuvia tilastollisia menetelmiä. Vaikka menetelmät ovat periaatteessa aineistosta riippumattomia, on ne pyritty sitomaan käytäntöön soveltamalla niitä hoitoilmoitusrekisteristä poimittuun aineistoon. Periaatteessa tässä tutkimuksessa ei ole kuitenkaan ole pyritty vastaamaan mihinkään yksittäiseen tutkimuskysymykseen, vaan tarkoituksena on ollut tarjota uusia metodologisia näköaloja rekisteriaineistojen entistä monipuolisempaan hyödyntämiseen.

Dynaamisten ilmiöiden mallintamista koskevassa luvussa on esitetty, kuinka tietynlaiseen muottiin sopivia ongelmanasetteluja sekä aineistoja pystytään käsittelemään analyyttisen mallin tasolla. Esitetty varannoista ja virtaamista muodostuvaa järjestelmää kuvaava muotti on varsin yleispätevä, eivätkä sen mahdollisuudet rajoitu tässä tutkimuksessa esitettyihin yksinkertaisiin malleihin. Terveydenhuollon sovelluksissa voidaan pelkän laitoshoidon kuvaamisen lisäksi esimerkiksi seurata hoitoketjuja, ottaa alueulottuvuus huomioon tai tarkastella yksilöiden järjestelmässä liikkuessaan keräämiä kustannuksia. Myös tutkimuksen kohteena olevien yksilöiden valinta on mielivaltaista; sairauden mukaan rajattujen potilaiden sijaan voidaan rajautua esimerkiksi tietyn sairaalan tai sairaanhoitopiirin potilaisiin.

Ongelmanasetteluun liittyvien ilmiöiden operationalisointi mallinnettavaksi järjestelmäksi on yleensä lähtökohtaisesti kaikkein vaativin ja eniten substanssitietoutta vaativa tehtävä. Järjestelmäpohjaisen mallin perusideoiden ja tavoitteiden ymmärtäminen on osoittautunut yleisesti ottaen helpoksi, mikä varmasti omalta osaltaan edistää niin ”kliinikon” ja ”tutkijan” kuin tiedon tuottajan ja käyttäjänkin välistä yhteisymmärrystä. Toisin sanoen on todennäköistä, että osapuolet pystyvät kommunikoimaan aikaisempaa tehokkaammin, kun asioista pystytään puhumaan yhteisellä kielellä (vrt. Dillon 1994).

Yhteisestä kielestä on kyse myös silloin, kun asioita esitetään matemaattisesti täsmällisellä tavalla. Eri tutkimusperinteissä samoista asioista saatetaan kuitenkin puhua eri nimillä ja käyttää nille erilaisia merkintöjä. Esimerkiksi perinteiset

demografiset menetelmät juontavat juurensa nykyaikaista todennäköisyysteoriaa edeltävältä ajalta, ja niiden toimintaperiaatteet ovat yleensä käsitettävissä kohtalaisen helposti myös intuitiivisella tasolla. Toisaalta näiden demografisten menetelmienkin matemaattisten ominaisuuksien ja yhteyksien täsmällinen esittäminen onnistuu alkuperäisia lähestymistapoja kätevämmin elinaika-analyysin merkinnöillä. Tämän tutkimuksen mallintamisen perusteita käsittelevässä kohdassa onkin osoitettu, kuinka kontekstiin erityisen hyvin sopivat demografiset käsitteet (vrt. Hakkarainen 1989) pystytään esittämään yhtäpitävästi elinaika-analyysin merkinnöillä.

Muilta osin mallintamista koskevassa luvussa on keskitytty ”sopivien” menetelmien tilastoteoreettiseen kuvaamiseen. Lisäksi pyrkimyksenä on ollut osoittaa, miten matemaattisilta ominaisuuksiltaan monimutkaistenkin mallien käyttö voi palautua hyvin yksinkertaisten menetelmien soveltamiseksi. Käsitellyt menetelmät onkin rajattu siten, että ne muodostavat toisiaan tukevan kokonaisuuden; kantavana teemana on järjestelmän kokonaisvaltainen mallintaminen sen osien kautta. Parametriset ja ei-parametriset lähestymistavat poikkeavat toisistaan niin paljon, että niiden voidaan tulkita muodostavan täysin omat kokonaisuutensa. Näin ollen tutkimuksen kuluessa päädyttiin irrottamaan parametrisia menetelmiä kuvaava osuus omaksi raportikseen (Sund 2000).

Tutkimuksen tilastoteoreettinen osuus on pyritty esittämään tiiviillä, mutta perustellulla tavalla. Luettavuuden sekä ymmärrettävyyden parantamiseksi kaavojen päättelyissä on tavallista enemmän selityksiä ja välivaiheita. Lisäksi ”yleisesti tunnettuihin” matemaattisiin totuuksiin nojauduttaessa on käytetty myös omaksumista helpottavia lähdeviitteitä. Kaiken kaikkiaan materiaalin ensisijaisena tarkoituksena onkin toimia kokoavana johdatuksena aihepiiriin ja siihen liittyvään kirjallisuuteen.

Tutkimuksen soveltavassa osassa on esimerkinomaisesti kuvattu, kuinka menetelmät soveltuvat hoitoilmoitusrekisteristä poimitun aineiston mallintamiseen. Analyysit on tehty Survo 98 -ohjelmistolla (Mustonen 1992; 1996), johon implementoitiin tutkimuksen kuluessa lukuisia erikoismenetelmiä ja apuohjelmia. Yleisellä tasolla analyyseista voidaan todeta, että aineiston suuren koon takia yksinkertaisetkin tarkastelut osoittautuivat tietojenkäsittelyllisessä mielessä varsin raskaiksi ja suunnittelua vaativiksi

toimenpiteiksi. Analyyseja suoritettaessa vastaantulleita ongelmia ja huomioita on myös pyritty kirjaamaan ylös, sillä nimenomaan ne täydentävät teoriaosuutta käytännön näkökulmasta ja antavat kattavan kuvan analyysiprosessista.

Tarkastelujen perusteella hoitoilmoitusrekisteristä poimitussa aineistossa on muutamia yleistettävissä olevia käytännön ongelmia. Ensinnäkin aineistossa saattaa olla sisäkkäisiä sekä limittäisiä hoitojaksoja. Toiseksi ei ole myöskään yksiselitteistä tapaa tunnistaa, milloin peräkkäiset hoitojaksot voidaan yhdistää tai monesko hoitojakso on kyseessä. Lisäksi osa havainnoista on poistopäivän osalta puutteellisia ja samaa hoitojaksoa kuvaavat asiakaslaskennasta ja poistoilmoituksesta peräisin olevat tiedot saattavat poiketa päädiagnoosin osalta toisistaan.

Analyysien mukaan laitoksissa läsnäolevien skitsofreenikoiden päivittäiset varannot ovat pienentyneet selvästi vuoden 1996 alun ja vuoden 1998 lopun välisenä aikana; yli vuoden laitoksessa olleita on noin puolet varannosta. Myös hoitoa saaneiden potilaiden ja hoitopäivien määrät ovat vähentyneet, mutta alkaneiden hoitojaksojen määrä on puolestaan kasvanut voimakkaasti. Vuosina 1997 ja 1998 skitsofrenian laitosinsidenssit olivat 0.54 ja 0.46 promillea. Lisäksi uusien sisäänotettavien skitsofreenikoiden päivittäiset määrät noudattivat ikä- ja sukupuoliryhmittäin Poisson-jakaumia. Vuosien 1994-1998 välisenä aikana skitsofreenikoilla oli keskimäärin 3.2 laitoskäyntiä; noin 5 %:lla potilaista oli yli kymmenen hoitokertaa. Eniten sisäänottoja ja poistoja oli maanantaisin sekä perjantaisin; vähiten puolestaan viikonloppuna.

Skitsofreenikoiden hoitojaksojen pituudet ovat lyhentyneet. Laitoksesta poistuminen on nuorilla skitsofreenikoilla muita hitaampaa alle puolen vuoden pituisissa hoitojaksoissa. Vähintään vuoden pituisia hoitojaksoja on vanhuksilla ja nuorilla noin 6 % ja muilla noin 3 % kaikista ikäryhmän hoitojaksoista. Naisten hoitojaksot ovat hiukan miesten hoitojaksoja lyhyempiä. Poistamisen jälkeen laitokseen joutuu uudestaan vanhuspotilaista noin 60 %; muissa ikäryhmissä uusijoita on yli 70 % potilaista. Laitoksesta poistamisen hasardifunktio on skitsofreenikoilla muodoltaan laskeva eli mitä kauemmin laitoksessa on oltu, sitä kauemmin siellä vielä tullaan olemaankin. Vastaavasti hoitokertojen välisen odotusajan kasvaessa todennäköisyys joutua uudestaan laitokseen pienenee.

Muodostettu skitsofreenikoiden potilaspopulaatiota kuvaava semi-Markov-malli ennustaa yksinkertaisuudestaan huolimatta populaation liikkeitä hyvin, kun vertailukriteerinä käytetään potilaiden havaittuja liikkeitä. Jos sisäänotettavien potilaiden määrä pysyy suhteellisen tasaisena, pitkäaikaispotilaiden määrä tulee mallin mukaan nykyisillä hoitokäytännöillä supistumaan noin 60%:iin vuoden 1998 lopun tilanteesta, kun taas lyhytaikaisten määrä pysyy ennusteen mukaan lähes samalla samalla tasolla kuin vuonna 1998.

Periaatteessa suoritetut analyysit ovat varsin teknispainotteisia ja yleisluontoisia, vaikka aineistoksi on valittu nimenomaan skitsofreniapotilaita. Toisin sanoen kiinnostus on kohdistunut pääasiassa menetelmiin eikä niinkään skitsofrenian kannalta mielenkiintoisiin kysymyksenasetteluihin; populaatioiden ominaisuuksia tarkastellaan esimerkiksi puuttumatta sen enempää Suomen psykiatrisen hoitojärjestelmän rakenteellisiin muutoksiin tai muutosten vaikutuksiin potilaiden hoitotilanteissa (ks. esim. Piirtola 1995; Salokangas ym. 2000). Myös semi-Markov-mallia on tässä tutkimuksessa hyödennetty varsin vähän sen mahdollisuuksiin nähden; mallia käyttämällä olisi esimerkiksi mahdollista ennustaa erilaisten hoitokäytäntöjen vaikutusta tuleviin laitosvarantoihin. Nyt minkäänlaisia vaihtoehtoisia ennusteita ei ole estimoitu, sillä niiden lisäarvo mallin teknisen toimivuuden testaamisen kannalta on vähäinen. Tarkasteltaessa suoritettuja analyyseja puhtaasti tekniseltä kannalta voidaankin yhteenvedonomaisesti ja yleistävästi todeta, että esitetyt menetelmät soveltuvat semi-Markov-mallia myöten erittäin hyvin hoitoilmoitusrekisteristä peräisin olevan aineiston analysointiin.

## 5.1 Aineiston hyödyntämisen ennakkoehdoista

Tässä tutkimuksessa on kuvattu, miten ja millä ehdoilla hoitoilmoitusjärjestelmä tuotetaan. Näkökulma on ollut puhtaasti toteavan teknis-hallinnollinen ja kritiikitön tilasto- ja rekisterituotannon ennakkoehtoja kohtaan. Tiedonsosiologisesti painottuneeseen tilastotutkimukseen kohdistuneen kirjallisuuskatsauksen perusteella Wallenius (1980, 18-30) toteaa kokoavasti, että havainnointi ja tietojen keruu ei koskaan ole ”puhdasta”, vaan se tapahtuu ennakkokäsitysten ja hahmotusmallien

puitteissa. Tiedon raportointikaan ei ole ”puhdasta” kuvauksen välittämistä vaan siinä käytetään käsitteistöä, joka joskus on hyvinkin teoriapainotteista. Eri tilanteissa käytetään erilaisia viitekehyksiä, jolloin niin tiedon keruu kuin toimittaminenkin tapahtuu eri mallien mukaan. Keskeistä on myös se, että intressiryhmissä ja instituutioissa tarvitaan tietoa kontrollin ja legitimoinnin välineenä. Siten tiedolla on myös ideologinen komponentti. Tieto kietoutuu yhteen vallan ja intressiristiriitojen kanssa, mutta suhde on kaksisuuntainen: intressit vaikuttavat tiedon hankinta- ja järjestämistapoihin, mutta tieto muuttaa ihmisten ja organisaatioiden toimintaa. (Wallenius 1980.)

Urposen (1987) mukaan tietojärjestelmiin liittyviä ongelmallisia ennakkoehtoja ovat muun muassa tilastojen perustaminen hallinnollisille kategorioille teoreettisten kategorioiden asemasta, kvantifioinnin ylikorostaminen, laadullisten ilmiöiden kuvaamisvaikeudet, keskittyminen helposti saatavaan tietoon, taloudellisten seikkojen hallitseva asema, tärkeiden yhteiskunnallisten ilmiöiden jääminen tietojärjestelmien ulottumattomiin sekä ajallisen vertailukelpoisuuden säilyttämisen aiheuttama tilastoinnin jäykkyys. Hakkarainen (2000) toteaa, että tilastotuotannossa tarvitaan voimavarojen ohella myös poliittista tahtoa - valtuutusta tehdä tietyn aihealueen tilastoja. Lisäksi tilastojen käyttäjä ei useinkaan ole tietoinen kuvattavan kohdeilmiön mittaamisen mahdollisuuksista; toisin sanoen monilla tilastojen käyttäjillä on erilainen käsitys yhteiskuntatilastojen luonteesta kuin niiden toimittajilla tai suunnittelijoilla (Hakkarainen 2000; Wallenius 1980, 18).Ylipäänsä tilastosuunnittelijat muistavat harvoin, että yhteiskunnassa vallitsee erilaisia tavoitteita: ristikkäisiä, päällekkäisiä ja ristiriitaisiakin. Ei ole ollenkaan mahdotonta, että jonkin yhteiskuntaryhmän tavoitteiden mukaista olisi tuottaa sellaista tietoa, jonka jokin toinen ryhmä haluaisi pitää salaisena (Wallenius 1980, 17). Eri yhteyksissä on tullut esiin myös ”kentän” haluttomuus tuottaa tietoa, jonka käyttötarkoitus on epäselvä (ks. esim. Aaltonen 1996; Nuotio 1982). Osaan näistä ongelmista pystytään vastaamaan tietojärjestelmien kehittämisellä, mistä hoitoilmoitusjärjestelmän systematisoitu uudistaminen onkin hyvä esimerkki.

Pelkkä ”tekninen” tietojärjestelmien tutkimus ja kehittäminen ei kuitenkaan pureudu riittävällä tavalla yhteiskunnallisen tiedontuotannon keskeisiin ongelmiin. Simpura (1989) onkin ehdottanut konstruktiivista lähestymistapaa, jossa on kysymys muun

muassa siitä, miten kullakin hetkellä vallitsevat käsitykset tarkasteltavista asiantiloista muodostuvat. Konstruktiivisesta näkökulmasta tietojärjestelmien tutkimisen kannalta tärkeäksi tutkimuskohteeksi nousevat ne tavat ja ehdot joilla tietojärjestelmän tuotoksia tulkitaan. Simpura erittelee kolme tiedontuotannon yhteiskunnallista tehtävää: tietojärjestelmän voi olettaa tuottavan kuvauksia asiantiloista, toimintakoneiston toiminnasta sekä toimintajärjestelmän resursseista. Perusolemukseltaan useimmat tietojärjestelmät ovat hänen mukaansa hallinnollisen kontrollin, suunnittelun ja toiminnan kehittämiseen liittyvän päätöksenteon välineitä. Tietojärjestelmän perustehtävien välillä on myös mutkikkaita kytkentöjä; järjestelmän kehittyminen muuttaa sitä käsitystä asiantiloista, jonka järjestelmään kirjattavat tiedot tuottavat. Tämän huomion valossa on tärkeää tutkia myös sitä, ketkä vaikuttavat järjestelmään eli tuottavat järjestelmän sisältämää tietoa tai käyttävät sitä. Keskeinen kysymys koskee niitä intressejä, joita eri toimijoilla voi olla sekä tietojärjestelmän käyttäjinä että sen sisältöön ja rakenteeseen vaikuttajina. Simpura toteaa toimijoiden pääintressien olevan ilmeisiä: vaikutusvalta ja aseman säilyttäminen. Edelleen olennaista valtaa merkitsevät hänen mukaansa pääsy tietolähteille ja tiedon legitimoitu tulkintaoikeus. Lisäksi mahdollisuus vaikuttaa järjestelmään esimerkiksi ohjeistuksen kautta on sekin osa valtaa. Toimijoiden erilaiset mahdollisuudet ajaa intressejään ovat taas riippuvaisia heidän hallussaan olevan kulttuurisen, akateemisen ja hallinnollisen pääoman määrästä. (Simpura 1989.)

Esimerkiksi viimeaikaisen palveluketjuideologian hyödyntäminen vaatii aineistoilta entistä parempaa kattavuutta ja luotettavuutta, mikä sinänsä vaikuttaa suoraan myös tuotannon käytäntöihin ja niihin vaikuttaviin intresseihin. Toisin sanoen on todettu, että rekisteriin kerättyjä tietoja on mahdollista käyttää tietyin edellytyksin entisestä poikkeavalla tavalla, mikä on herättänyt intressin vaatia rekisteriltä näitä edellytyksiä.

Edellä esitetyn perusteella erilaiset teoreettiset, ideologiset ja käytännölliset viitekehykset luovat ne puitteet, joissa tilastojen laatijat ja käyttäjät toimivat. Käytännössä tilastotuotannon ennakkoehdot, tutkijoiden maailmankuva ja intressit heijastuvat voimakkaasti myös tutkimustavoitteisiin ja -asetelmiin. Hoitoilmoitusrekisterin tapauksessa käytännöllis-ammatilliset perustelut sekä taloustieteen viitekehys staattisine panos–tuotos-malleineen ovat olleet vaikutuspääomaltaan voimakkaimpia ja näin

ollen käytetyimpiä lähestymistapoja viime vuosiin saakka (vrt. Hakkarainen 2000). Lisäksi hoitoilmoitusrekisterin tietojen hyödyntämiseen on ollut kiinnostusta useissa eri yhteyksissä esimerkiksi epidemiologisten sekä lääke-, sosiaali- ja hallintotieteellisten tutkimusten tietolähteenä, mikä jo sinänsä osoittaa hoitoilmoitusrekisterin suunnitelmallisen kehittämisen kantaneen hedelmää mahdollisista piilevistä valtataisteluista ja intressiristiriidoista huolimatta.

Yksilötason rekisteriaineistoa käytettäessä tutkija saa entistä paremman lähikontaktin aineistoon, kun käytettävissä on mahdollisimman paljon alkuperäisiä havaintoja. Niiden perusteella tutkija voi tehdä omia ratkaisujaan aineiston käsittelyssä, eikä hänen tarvitse tyytyä tilastontekijän luokituksiin (Ilmakunnas 1998). Rekisteriaineistojen tietoihin perustuva analyysi on lähes poikkeuksetta luonteeltaan kvantitatiivista, joskin usein kaukana klassisesta, kontrolloituun kokeeseen perustuvasta tiukasta hypoteesintestauksesta (vrt. Lehto 1998). Hedelmällisimmillään analyysi on ymmärtävän eksploratiivista ja paradigmojen rajoja ennakkoluulottomasti rikkovaa. Rekisteriaineistoihin liittyy joka tapauksessa niin paljon ennakko-oletuksia, ettei puhtaan objektiivisista johtopäätöksistä ole edes tarkoituksenmukaista puhua. Tilastollinen päättely tarjoaa siitä huolimatta matemaattisen täsmällisiä metodeja saatujen ideoiden todentamiseksi. Järkevän käytön ehtona on kuitenkin menetelmien, aineiston ja tutkittavan ilmiön pintaa syvemmälle ulottuva tuntemus - menetelmien oletukset on osattava ottaa huomioon ja toisaalta odottamattomien tulosten syyt on pystyttävä jäljittämään. Tutkijan tieteelliset ja yhteisölliset sitoumukset heijastuvat välttämättä tutkimuksen kulkuun, mutta toimijuuden ei silti tarvitse olla luovaa toistoa tietyn menetelmän ja käsitteellisen maailman rajoissa (vrt. Ronkainen 1998). Myös tutkimuksen aikana tehdyt havainnot vaikuttavat varmasti lopullisiin johtopäätöksiin, sillä juuri ne syventävät aineiston ja ilmiön tuntemusta. Parhaimmillaan tuloksena on yksityiskohtainen ja periaatteessa myös toiston mahdollistava vakuuttava kuvailu tutkimusprosessin etenemisestä ja saatujen tulosten järkevyydestä. Toisaalta tilastollisen ja matemaattisen päättelyn osaamisen välttämättömyys voi johtaa tutkittavan ilmiön kannalta täysin epäoleelliseen tekniseen näpertelyyn. Yleinen mielipide asettaa helposti yhtäläisyysmerkit tieteellisyyden ja matemaattisuuden välille ilman perusteluja. Samalla mielikuvien tasolla elää käsitys menetelmätieteiden vaikeudesta, ja tällöin teknistä

esitystä käyttävää tutkijaa pidetään lahjakkaana, koska hän kykenee käyttämään vaikeita menetelmiä. Edelleen kuvitellaan virheellisesti, että tällainen esitys sinänsä takaa sen, että tutkimus on ideologisesti vähemmän sitoutunutta kuin muu tutkimus (Ilmakunnas 1998).

## 5.2 Heterogeenisuuden hallinnasta

Kun tutkitaan eri ryhmien dynaamisten prosessien välisiä eroja, on ensin määriteltävä, mitä ryhmiä halutaan tarkastella. Ryhmien määrittelyn mahdollistamiseksi on oltava tietoa tai teoria siitä, mitkä taustatekijät heijastelevat eroja oleellisella tavalla. Tässä tutkimuksessa tarkoituksena ei ole ollut esittää teoriaa dynaamisia prosesseja selittävistä tekijöistä, vaan kuvata, millaisia prosessit ovat ja miten niitä voidaan mallintaa. Mallien toimivuuden kannalta on kuitenkin oleellista luokitella aineisto sopivalla tavalla mahdollisimman homogeenisiin ryhmiin.

Tässä tutkimuksessa käytetyn pelkän karkean ryhmittelyn sijaan heterogeenisuuden hallintaan on mahdollista soveltaa myös erilaisia kovariaatit huomioivia vakiointi- ja regressiomenetelmiä (ks. esim. Cox 1972; Ugwuowo & McClean 2000). Tällaisista menetelmistä on kiistatta etua monissa tapauksissa, ja niitä kannattaisikin mahdollisuuksien mukaan yrittää hyödyntää. Haittapuolia ovat lähinnä vahvat oletukset ja mallipohjaisuuden luoma monimutkaisuus. Esimerkiksi kovariaattien huomioiminen semi-Markov-mallissa on analyyttisesti vaikeaa.

Mallien kannalta olennaisin jako ryhmiin on itse asiassa tehtävä jo operationalisoidessa tutkittavaa ilmiötä sopivaksi järjestelmäksi. Substanssitiedolla on tässä vaiheessa usein kriittinen osuus. Toisaalta olettamusten paikkansapitävyys on joka tapauksessa todennettava aineistoa käyttäen. Toisin sanoen aineistoa hyödyntävä tietämyksen muodostaminen (data-mining) on erittäin tärkeä osa mallintamisprosessia. Esimerkiksi kysymyksenasettelun kannalta oleellisten tapahtumaketjujen etsiminen ja luokittelu on mahdollista tulkita ryhmittelyanalyysiksi tai hahmontunnistukseksi. Rekisteriaineistojen entistä monipuolisempaa hyödyntämistä ajatellen tällaiset lähestymistavat ovat oleellisessa asemassa.

## 5.3 Lopuksi

Tämä tutkimus on tilastotieteen pro gradu -tutkielma. Määritelmän mukaan pro gradu -tutkielman ensisijaisena tavoitteena on harjaannuttaa opiskelija itsenäiseen tutkimustyöhön, asianomaisen tieteenalan tukimusmenetelmien hallintaan, lähteiden käyttöön sekä tieteelliseen esitystapaan ja kirjalliseen ilmaisuun.

Tilastotieteeksi voidaan tulkita niin puhtaan teoreettis-matemaattiset kuin soveltavan mallintamispainotteisetkin lähestymistavat. Seppo Mustonen on myös todennut, että tilastotiedettä pitäisi oikeastaan kutsua tietojenkäsittelyksi, sillä sitä hyödyllisen tiedon survominen aineistosta käytännössä on. Suomen kielessä tietojenkäsittely ymmärretään kuitenkin laajemmassa mielessä myös ohjelmoitavissa olevaksi automatisoimiseksi, jota tilastotiede ei perusolemukseltaan suinkaan ole. Tilastotieteelle on äärimmäisen vaikea vetää luonnollisia rajoja, joten jonkinlaisen kuvan saamiseksi voisi kuvitella koordinaatiston, jossa yksi akseli on tietojenkäsittelytieteestä matematiikkaan ja toinen soveltavasta teoreettiseen. Tässä koordinaatistossa ”neutraali” tilastotiede olisi origossa ja näin ollen käytännössä kohdattavan tilastotieteen voisi hahmottaa enemmän tai vähemmän satunnaisena vaelteluna origon ympärillä tutkijan kiinnostuksen kohteiden mukaan.

Näiden ennakkoehtojen puitteissa tämä tutkimus on saavuttanyt nykyisen muotonsa. Tilastotieteellinen lähestymistapa on enemmän soveltavan tietojenkäsittelyn kuin puhtaan teoreettisen matematiikan puolella. Tutkimuksen pääasiallisen sovellusalan voidaan tulkita olevan terveydenhuollon operaatioanalyysi. Lisäksi mukana on myös yhteiskuntatieteellisiä näkökulmia.

Tässä tutkimuksessa kuvattuja menetelmiä ei ole juurikaan hyödynnetty terveydenhuollon tutkimuksessa ainakaan Suomessa. Vastaavalla tavalla menetelmien tilastoteoreettisia ominaisuuksia kokoavia esityksiäkään ei saatavilla olevasta kirjallisuudesta tunnu löytyvän; erityisesti varantojen määrittämiseen käytetyt analyyttiset semi-Markov-mallit vaikuttavat harvinaisilta. Muutamaan vuoteen ei ole myöskään tehty

vastaavaa hoitoilmoitusrekisterin ominaisuuksia kuvaavaa katsausta eikä yhtä tuoreita hoitoilmoitusrekisterin tietoja käyttäviä skitsofreniatutkimuksia ole vielä näkynyt.

Tutkimus olisi tuskin koskaan nähnyt päivänvaloa nykyisenkaltaisessa muodossaan ilman sitä apua ja kannustusta, jota olen aina tarvittaessa saanut tuekseni. Haluankin kiittää Anni Hakkaraista motivoivista keskusteluista ja arvokkaista kommenteista. Kiitokset Seppo Mustoselle silmieni avaamisesta tilastolliselle tietojenkäsittelylle sekä luottamuksesta omaehtoiseen työskentelyyni. Kiitos myös Timo Hakuliselle kiinnostuksesta työtäni kohtaan. Olen kiitollinen kannustavasta suhtautumisesta sekä hyödyllisistä kommenteista ja ehdotuksista lisäksi Ilmo Keskimäelle, Mikko Nenoselle sekä Pekka Rissaselle. Miellyttävästä ja joustavasta työympäristöstä sekä uskosta työhöni kiitokset Tilastotieteen laitokselle ja Hannu Niemelle sekä Stakesille; erityisesti Ritva Heinille, Olli Nylanderille ja Hannu Hämäläiselle. Kiitoksia vaivannäöstä ja vastauksista kysymyksiini Sari Kauppiselle, Arja Kituselle, Markku Koiraselle, Oleg Nikiforoville, Simo Pelanterille, Anja Rasimukselle, Timo Tuorille sekä Kimmo Vehkalahdelle. Erityiskiitokset Annarita Koiraselle kaikesta avusta ja kannustuksesta. Kiitoksia myös kaikille niille, jotka tavalla tai toisella ovat virallisissa tai epävirallisissa tilaisuuksissa osoittaneet kiinnostusta työtäni kohtaan.

Loppujen lopuksi tämä tutkimus on ainoastaan raportti omasta oppimisprosessistani ja kapea häivähdys siitä innostuksesta, joka liittyy haluuni ymmärtää asioita. Toivottavasti tuotoksestani on jotain hyötyä myös muille.

# Kirjallisuus

Aaltonen, H. (1996): Koko elämä kansien väliin. Hoitoilmoitusta kritisoidaan tarpeettomana ja epäkäytännöllisenä. Tukiviesti 26: 2, 31.

Akritas, Michael G. (1986): Bootstrapping the Kaplan-Meier Estimator. Journal of the American Statistical Association 81: 1032-1038.

Alho, Juha (toim.) (1999): Statistics, Registries and Science - Experiences from Finland. Tilastokeskuksen julkaisuja. Otava. Keuruu.

Allison, Paul D. (1984): Event History Analysis : Regression for Longitudinal Event Data. Quantitative Applications in the Social Sciences Series 46. Sage Publications. Beverly Hills.

Andersen, Per Kragh & Borgan, Ørnulf & Gill, Richard D. & Keiding, Niels (1993): Statistical Models Based on Counting Processes. Springer-Verlag. New York.

Anderson, T. W. & Goodman, Leo A. (1957): Statistical Inference about Markov Chains. Annals of Mathematical Statistics 28: 1, 89-110.

Ansell, J. I. & Phillips, M. J. (1989): Practical Problems in the Statistical Analysis of Reliability Data. Applied Statistics 38: 2, 205-247.

Armitage, Peter & Colton, Theodore (toim.) (1998): Encyclopedia of Biostatistics. John Wiley & Sons. Chichester.

Aro, Seppo & Koskinen, Riitta & Keskimäki, Ilmo (1990): Sairaalastapoistorekisterin diagnoosi-, toimenpide- ja tapaturmatietojen luotettavuus. Duodecim 106, 1443-1450.

Aromaa, Arpo (1980): Tilastot ja rekisterit lääketieteellisen tutkimuksen perustana. Sosiaalivakuutus 18: 7-8, 202-209.

Bartholomew, David J. (1971): The Statistical Approach to Manpower Planning. Statistician 20: 1, 3-26.

Bartholomew, David J. (1975): Errors of Prediction for Markov Chain Models. Journal of the Royal Statistical Society, Series B (Methodological) 37: 3, 444-456.

Bartholomew, David J. (1981): Mathematical Methods in Social Science. Handbook of Applicable Mathematics: Guidebook 1. Wiley-Interscience. Chichester.

Bartholomew, David J. (1982): Stochastic Models for Social Processes. Third Edition. John Wiley & Sons. Chichester.

Bartholomew, David J. & Forbes, Andrew F. & McClean, Sally I. (1991): Statistical Techniques for Manpower Planning. Second Edition. John Wiley & Sons. Chichester.

Barton, Thomas R. & Pillai, Vijayan K. (1995): Using an Event History Analysis to Evaluate Social Welfare Programs. Research on Social Work Practice 5: 2, 176-194.

Billingsley, Patrick (1961): Statistical Methods in Markov Chains. Annals of Mathematical Statistics 32: 1, 12-40.

Blossfeld, Hans-Peter & Rohwer, Götz (1995): Techniques of Event History Modeling. Lawrence Erlbaum Associates. New Jersey.

Box-Steffenmaister, Janet M. & Jones, Bradford S. (1997): Time Is of The Essence: Event History Models in Political Science. American Journal of Political Science 41: 4, 1414-1461.

Breslow, Norman (1970): A Generalized Kruskal-Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship. Biometrika 57: 3, 579-594.

Brookmeyer, Ron & Crowley, John (1982): A Confidence Interval for the Median Survival Time. Biometrics 38, 29-41.

Brown, G. W. (1960): Length of Hospital Stay and Schizophrenia: A Review of Statistical Studies. Acta Psychiatrica et Neurologia Scandinavica 33, 414-430.

Butler, Ronald W. & Huzurbazar, Aparna V. (1997): Stochastic Network Models for Survival Analysis. Journal of the American Statistical Association 92, 246-257.

Cannon, Tyrone & Kaprio, Jaakko & Lönnqvist, Jouko & Huttunen, Matti & Koskenvuo, Markku (1998): The Genetic Epidemiology of schizophrenia in a Finnish Twin Cohort: A Population-Based Modeling Study. Archives of General Psychiatry 55: 1, 67-74.

Chatfield, Christopher (1973): Statistical Inference Regarding Markov Chain Models. Applied Statistics 22: 1, 7-20.

Clayton, D.G. (1988): The analysis of event history data: a review of progress and outstanding problems. Statistics in medicine 7, 819-841.

Clogg, C. C.; toim (1989): Sociological Methodology 1989. American Sociological Association. Blackwell. Oxford.

Cox, David R. (1972): Regression Models and Life-Tables (with discussion). Journal of the Royal Statistical Society, Series B (Methodological) 34: 2, 187-220.

Cox, David R. & Oakes, D. (1984): Analysis of Survival Data. Chapman and Hall. London.

Courgeau, Daniel & Lelièvre, Éva (1997): Changing Paradigm in Demography. Population: An English Selection 9, 1-10.

Davies, Ruth & Johnson, David & Farrow, Stephen (1975): Planning Patient Care With a Markov Model. Operational Research Quarterly 26: 3, 599-607.

De Stavola, Bianca L. (1988): Testing Departures from Time Homogeneity in Multistate Markov Processes. Applied Statistics 37: 2, 242-250.

Delasie, Luc (1998): Bridging the gap between clinicians and health managers. European Journal of Operational Research 105, 248-256.

Delasie, Luc & Kastelein, Allard & van Merode, Frits & Vissers, Jan (toim.) (1998): Feature Issue: Managing Health Care under Resource Constraints. European Journal of Operational Research 105, 247-370.

Denny, J. L. & Yakowitz, S. J. (1978): Admissible Run-Contingency Type Tests for Independence and Markov Dependence. Journal of the American Statistical Association 73, 177-181.

Dillon, Andrew (1994): What the General Manager Needs. Teoksessa: Millard & McClean 1994.

Eaton, William W. & Whitmore, G. A. (1977): Length of Stay as a Stochastic Process: A General Approach and Application to Hospitalization for Schizophrenia. Journal of Mathematical Sociology 5, 273-292.

Efron, Bradley (1981): Censored Data and the Bootstrap. Journal of the American Statistical Association 76, 312-319.

Elandt-Johnson, Regina C. & Johnson, Norman L. (1980): Survival models and data analysis. Wiley series in probability and mathematical statistics. John Wiley & Sons. New York.

Feller, William (1968): An Introduction to Probability Theory and Its Applications. Third Edition. John Wiley & Sons. New York.

Gehan, Edmund A. (1965a): A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. Biometrika 52: 1/2, 203-223.

Gehan, Edmund A. (1965b): A Generalized Two-Sample Wilcoxon Test for Doubly Censored Data. Biometrika 52: 3/4, 650-653.

Gissler, Mika (1999a): Routinely Collected Registers in Finnish Health Research. Teoksessa: Alho 1999.

Gissler, Mika (1999b): Administrative Registers in Health Research - A Cohort Study of Finnish Children Born in 1987. STAKES Research Report 97. Helsinki.

González-Manteiga, Wenceslao & Cao, Ricardo & Marron, James S. (1996): Bootstrap Selection of the Smoothing Parameter in Nonparametric Hazard Rate Estimation. Journal of the American Statistical Association 91: 1130-1140.

Graybill, Franklin A. (1983): Matrices with Applications in Statistics. Second Edition. Wadsworth International Group. Belmont.

Haimi, Olavi (1987): Väestöntutkimus - Periaatteet ja menetelmät. Väestöntutkimuslaitoksen julkaisusarja D 19/1987. Väestöliitto. Helsinki.

Hakkarainen, Anni (1980): Pitkäaikaissairaat psykiatrisissa sairaaloissa - Havaintoja ja ennusteita. Lääkintöhallituksen tutkimuksia 21. Helsinki.

Hakkarainen, Anni (1982): Mielisairaalaan otettujen potilaiden hoitokerrat ja hoitoajat 1970-luvulla. Sosiaalilääketieteellinen aikakauslehti 19, 94-106.

Hakkarainen, Anni (1985): Psykiatrisen potilaspopulaation muutostapahtumat väestössä. Sosiaalilääketieteellinen aikakauslehti 22, 210-223.

Hakkarainen, Anni (1989): Psykiatrisen potilasväestön muutokset: teoreettinen malli ja sen sovellutukset Suomessa. Lääkintöhallituksen tutkimuksia 48. Helsinki.

Hakkarainen, Anni (1998): Valtakunnalliset rekisterit auttavat paikallisen terveydenhuollon suunnittelua. Kansanterveys 5/1998 (Kansanterveyslaitoksen tiedotuslehti), 4 & 9.

Hakkarainen, Anni (2000): Tilastotuotannon mahdollisuudet. Suomen Tilastoseuran vuosikirja 2000 (painossa).

Hakkarainen, Anni & Aitta, Timo & Sund, Reijo & Nylander, Olli (1999): Life Table Model and the Length of Stay of Surgical Patients. Bulletin of the International Statistical Institute, 52nd Session, Contributed Papers, Tome LVII, Book 2, 5-6.

Harris, Theodore E. (1963): The Theory of Branching Processes. Springer-Verlag. Berlin.

Heckman, James J. & Singer, Burton (1984): Econometric Duration Analysis. Journal of Econometrics 24, 63-132.

Heliövaara, M. & Reunanen, A. & Aromaa, A. & Knekt, P. & Aho, K. & Suhonen, O. (1984): Validity of hospital discharge data in a prospective epidemiological study on stroke and myocardial infarction. Acta Medica Scandinavica 216, 309-315.

Henry, N. W. & McGinnis, R. & Tegtmeyer, H. W. (1971): A finite model of mobility. Journal of Mathematical Sociology 1, 107-118.

Hoem, Jan M. (1976): The Statistical Theory of Demographic Rates - A Review of Current Developments. Scandinavian Journal of Statistics 3: 169-185.

Huzurbazar, Aparna V. (1999): Flowgraph Models for Generalized Phase Type Distributions Having Non-Exponential Waiting Times. Scandinavian Journal of Statistics 26, 145-157.

Häkkinen, Unto & Linna, Miika & Salonen, Markku (1994): Korvausmenettelyn ja kuntakoon vaikutus erikoissairaanhoidon taloudelliseen riskiin. Suomen Lääkärilehti 49: 25, 2454-2458.

Hämäläinen, Hannu (1998): Stakesin rekisterit ja niiden tutkimuskäyttö. Teoksessa: Valkonen ym. 1998a.

Härdle, Wolfgang (1991): Smoothing techniques : with implementations in S. Springer Series in Statistics. Springer-Verlag. New York.

Härö, Sakari (1980): Valtakunnallisista rekistereistä ja informaatiopolitiikastamme. Sosiaalivakuutus 18: 7-8, 232-234.

Ilmakunnas, Seija (1998): Kansantaloustiede ja tilastot - Mersulla lehmipolkua pitkin? Teoksessa: Paananen ym. 1998.

Isohanni, M. & Mäkikyrö, T. & Moring J. & Räsänen, P. & Hakko, H. & Partanen, U. & Koiranen, M. & Jones, P. (1997): A comparison of clinical and research DSM-III-R diagnoses of schizophrenia in a Finnish national birth cohort. Social Psychiatry and Psychiatric Epidemiology 32: 303-308.

Janssen, Jaques & Limnios, Nikolaos (toim.) (1999): Semi-Markov Models and Applications. Kluwer Academic Publishers. Dordrecht.

Kalbfleisch, John D. & Prentice, Ross L. (1980): The Statistical Analysis of Failure Time Data. John Wiley & Sons. New York.

Kannisto, Väinö (2000): Relationships between population studies and social medicine. Sosiaalilääketieteellinen Aikakauslehti 37: 1, 76-81.

Kao, Edward P. C. (1972): A Semi-Markov Model to Predict Recovery Progress of Coronary Patients. Health Services Research 7, 191-208.

Kao, Edward P. C. (1974): Modeling the Movement of Coronary Patients within a Hospital by Semi-Markov Processes. Operations Research 22: 683-699.

Kaplan, E. L. & Meier, Paul (1958): Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53, 457-481.

Kauppinen, Sari (1999): Sosiaalihuollon laitoshoito ja asumispalvelut 1998. Stakesin tilastoraportti 31/1999. Stakes.

Kauppinen, Sari & Nikiforov, Oleg & Viramo, Petteri & Karjalainen, Timo & Frey, Harry & Nenonen, Mikko & Nylander, Olli (1999): Dementia Suomessa 1995. Stakesin tilastoraportti 6/1999. Helsinki.

Kay, Richard (1986): A Markov Model for Analysing Cancer Markers and Disease States in Survival Studies. Biometrics 42, 855-865.

Keiding, Niels (1991): Age-specific Incidence and Prevalence: a Statistical Perspective (with discussion). Journal of the Royal Statistical Society, Series A 154: 3, 371-412.

Kelton, W. David & Kelton, Christina M. L. (1984): Hypothesis Tests for Markov Process Models Estimated From Aggregate Frequency Data. Journal of the American Statistical Association 79, 922-928.

Keskimäki, Ilmo (1997): Social equity in the use of hospital inpatient care in Finland. STAKES Research Reports 84. Helsinki.

Keskimäki, Ilmo & Salinto, M. & Koskinen S. (1997): Viranomaisrekisterit terveyden ja terveyspalvelujen käytön tutkimusaineistoina. Suomen Lääkärilehti 52: 6, 585-591.

Keskimäki, Ilmo & Aro, Seppo (1991): Accuracy of data on diagnoses, procedures and accidents in the Finnish Hospital Discharge Register. International Journal of Health Sciences 2: 1, 15-21.

Keskimäki, Ilmo & Seitsalo, Seppo & Österman, Heikki & Rissanen, Pekka (2000): Reoperations After Lumbar Disc Surgery - A Population-Based Study of Regional and Interspecialty Variations. Spine 25: 12, 1500-1508.

Kiefer, N. M. (1988): Economic Duration Data and Hazard Functions. Journal of Economic Literature 26, 646-679.

Kleemola, Maija (1998a): Tietosuojavaatimukset henkilörekistereihin perustuvissa tieteellisissä tutkimuksissa. Teoksessa: Valkonen ym. 1998a.

Kleemola, Maija (1998b): Terveystutkimusten tietosuoja EU-Suomessa. Kansanterveys 5/1998 (Kansanterveyslaitoksen tiedotuslehti), 3.

Klein, John P. (1998): Survival Distributions and Their Charasteristics. Teoksessa: Armitage & Colton 1998.

Knowles, Trevor (1994): What the Director of Social Services Needs. Teoksessa: Millard & McClean 1994.

Korkeila, J.A. & Lehtinen, Ville & Tuori, Timo & Helenius, H. (1998): Patterns of psychiatric hospital service use in Finland: a national register study of hospital discharges in the early 1990s. Social Psychiatry and Psychiatric Epidemiology 33, 218-223.

Koskenvuo, M. & Kaprio, J. & Langinvainio, H. & Romo, M. & Pulkkinen, P. (1985): Changes in incidence and prognosis of ischaemic heart disease in Finland: a record linkage study of data on death certificates and hospital records for 1972 and 1981. British Medical Journal 290: 1773-1775.

Koskinen, Seppo & Martelin, Tuija & Valkonen, Tapani (1998): Rekisterit avaavat ainutlaatuiset tutkimusmahdollisuudet. Kansanterveys 5/1998 (Kansanterveyslaitoksen tiedotuslehti), 9-10.

Lagakos, S. W. & Sommer, C. J. & Zelen, M. (1978): Semi-Markov Models for Partially Censored Data. Biometrika 65:2, 311-317.

Lagergren, Mårten (1998): What is the role and contribution of models to managements and research in the health services? A view from Europe. European Journal of Operational Research 105, 257-266.

Lancaster, Tony (1992): The Econometric Analysis of Transition Data. Cambridge University Press. New York.

Le, Chap T. (1997): Applied Survival Analysis. Wiley Series in Probability and Statistics. Wiley-Interscience. New York.

Lehto, Anna-Maija (1998): Laatua surveytutkimukseen. Teoksessa: Paananen ym. 1998.

Li, L. X. & Benton, W. C. (1996): Performance measurement criteria in health care organizations: Review and future research directions. European Journal of Operational Research 93, 449-468.

Linna, Miika (1999): Measuring Hospital Performance: the Productivity, Efficiency and Costs of Teaching and Research in Finnish Hospitals, STAKES research report 98. Helsinki.

Little, Roderick J. A. (1992): Incomplete Data in Event History Analysis. Teoksessa: Trussel ym. 1992.

Lounamaa, Anne & Tuori, Timo & Pelanteri, Simo (1997): Psykiatria 1996. Stakesin tilastotiedote 26/1997. Helsinki.

Lääkintöhallitus (1986). Tautiluokitus 1987, Systemaattinen osa. Lääkintöhallitus ja sairaalaliitto. Valtion painatuskeskus. Helsinki.

McClean, Sally I. (1978): Continuous-time Stochastic Models of a Multigrade Population. Journal of Applied Probability 15, 26-37.

McClean, Sally I. (1994): Modelling and Simulation for Health Applications. Teoksessa: Millard & McClean 1994.

McClean, Sally & Devine, Colum (1995): A nonparametric maximum likelihood estimator for incomplete renewal data. Biometrika 82:4, 791-803.

McClean, Sally I. & Gribbin, J. Owen (1987): Estimation for Incomplete Manpower Data. Applied Stochastic Models and Data Analysis 3, 13-25.

McClean, Sally I. & Gribbin, J. Owen (1991): A Non-Parametric Competing Risk Model for Manpower Planning. Applied Stochastic Models and Data Analysis 7, 327-341.

McClean, Sally & Millard, Peter (1993): Patterns of stay after admission in geriatric medicine: an event history approach. The Statistician 42, 263-274.

McClean, Sally & Montgomery, Erin (1999): Estimation for Semi-Markov Manpower Models in a Stochastic Environment. Teoksessa: Janssen & Limnios 1999.

McClean, Sally I. & Montgomery, Erin & Ugwuowo, Fidelis (1998): Non-homogenous Continuous-time Markov and Semi-Markov Manpower Models. Applied Stochastic Models and Data Analysis 13, 191-198.

McFarland, David D. (1970): Intragenerational Social Mobility as a Markov Process: Including a Time-Stationary Markovin Model that Explains Observed Declines in Mobility Rates Over Time. American Sociological Review 35: 3, 463-476.

McGinnis, Robert (1968): A Stochastic Model of Social Mobility. American Sociological Review 33: 5, 712-722.

Mehlmann, Alexander (1977): A Note on the Limiting Behaviour of Discrete-Time Markovian Manpower Models with Inhomogenous Independent Poisson input. Journal of Applied Probability 14, 611-613.

Mehlmann, Alexander (1979): Semi-Markovian Manpower Models in Continuos Time. Journal of Applied Probability 16, 416-422.

Miikki, Ritva & Orre, Soili (1998): Pääjohtaja Vappu Taipale: Rekisterinpito tarvitsee kansalaisten tuen. Dialogi 8: 8, 18-19.

Millard, Peter (1994): What the Clinician Needs. Teoksessa: Millard & McClean 1994.

Millard, Peter H. & McClean, Sally I. (toim.) (1994): Modelling Hospital Resource Use - A different approach to the planning and control of health care systems. Royal Society of Medicine Press. London.

Millard, Peter H. & McClean, Sally I. (toim.) (1996): Go with the flow - A systems approach to healthcare planning. Royal Society of Medicine Press. London.

Moolgavkar, Suresh H. & Prentice, Ross L. (1986): Modern Statistical Methods in Chronic Disease Epidemiology (1986). Wiley-Interscience. New York.

Morrison, Donald G. & Massy, William F. & Silverman, Fred N. (1971): The effect of Nonhomogeneous Populations on Markov Steady-State Probabilities. Journal of the American Statistical Association 66, 268-274.

Mustonen, Seppo (1992): Survo : An Integrated Environment for Statistical Computing and Related Areas. Survo Systems. Helsinki.

Mustonen, Seppo (1996): Survo ja minä. Survo Systems. Helsinki.

Mähönen, Markku (1993): Poistoilmoitusrekisteri sepelvaltimotaudin epidemiologisen tutkimuksen tietolähteenä. Stakesin tutkimuksia 28. Helsinki.

Mäkikyrö, T. & Isohanni, M. & Moring, J. & Hakko, H. & Hovatta, I. & Lönnqvist, J. (1998): Accuracy of register-based schizophrenia diagnoses in a genetic study. European Psychiatry 13, 57-62.

Nenonen, Mikko (1997): Hoitoilmoitusjärjestelmä relaatiotietokantana. Dialogi 7: 5, 40-41.

Nenonen, Mikko (1999): Terveydenhuollon perusaikasarjat eivät tyystin vailla ongelmia. Dialogi 9: 3, 32-33.

Nenonen, Mikko & Pelanteri, Simo & Rasilainen, Jouni (2000): Mielenterveyden häiriöiden hoito Suomessa 1978-1998. Stakesin tilastoraportti 3/2000. Helsinki.

Nikiforov, Oleg (1984): Yleissairaalahoito Suomessa 1960- ja 1970-luvuilla. Lääkintöhallituksen tutkimuksia 32. Helsinki.

Nouko-Juvonen, Susanna (1997): Kehitysvammapalvelut 1995. Stakesin tilastotiedote 10/1997. Helsinki.

Nouko-Juvonen, Susanna (1999): Kuka asuu suomalaisessa keskuslaitoksessa 1990-luvulla? -Laitoshoidossa asuvien kehitysvammaisten henkilöiden tarkastelu. Teoksessa: Vesala 1999.

Nouko-Juvonen, Susanna (2000): Mielenterveyden keskusliiton turva-asumisprojektin arviointitutkimuksen väliraportti. FinSoc Työpapereita 3/2000. Stakes. Helsinki.

Nouko-Juvonen, Susanna & Sund, Reijo (1999): Keskuslaitosten ja asumispalvelujen käyttö palvelujen tarjonnan ja tarpeen vaikutus. Sosiaalilääketieteellinen aikakauslehti 36: 4, 357-364.

Nuotio, Jorma (1982): Tietoa kontrollin vuoksi. Sosiaaliturva 70: 2, 1089-1090.

Nylander, Olli & Nenonen, Mikko (2000): Julkinen sosiaali- ja terveydenhuolto verkottuu. Teoksessa: Uusitalo ym. 2000.

Nylander, Olli & Pelanteri, Simo & Linna, Miika & Nenonen, Mikko (1998): Sosiaali- ja terveydenhuollon asiakasvirrat ja palveluketjut 1995. Stakesin tilastoraportti 19/1998. Helsinki.

Ojala, Toini (1980): Potilasrekisterien käyttökelpoisuudesta. Sosiaalivakuutus 18: 7-8, 235-236.

Orre, Soili (1998): Terveydenhuollon erillisrekistereistä jo toinen työkokous. Dialogi 8: 2, 28.

Ortega, James M. (1987): Matrix Theory: A Second Course. Plenum Press. New York.

Ouhbi, B. & Limnios, N. (1996): Non-Parametric Estimation for Semi-Markov Kernels with Application to Reliability Analysis. Applied Stochastic Models and Data Analysis 12, 209-220.

Paananen, Seppo & Juntto, Anneli & Sauli, Hannele; toim. (1998): Faktajuttu - tilastollisen sosiaalitutkimuksen käytännöt. Vastapaino. Tampere.

Papadopoulou, Aleka A. & Vassiliou, Panagiotis C. G. (1999): Continuous Time Non Homogenous Semi-Markov Systems. Teoksessa: Janssen & Limnios 1999.

Patovaara, Timo (1995): Vektori- ja matriisilaskenta. Tilastotieteen laitoksen opetusmonisteita. Helsingin yliopisto.

Pérez-Ocón, Rafael & Ruiz-Castro, Juan Eloy & Gámiz-Pérez, M. Luz (1999): Semi-Markov Models for Lifetime Data Analysis. Teoksessa: Janssen & Limnios 1999.

Piirtola, Olli (1995): An evaluative description of changes in a comprehensive psychiatric treatment system. STAKES Research Reports 57. Helsinki.

Poikolainen, Kari (1980): Sairaalapoistorekisterin kehittämisestä. Sosiaalivakuutus 18: 7-8, 235.

Poikolainen, Kari (1982): Miten luotettava on sairaalastapoistorekisteri: viisi alkoholin käyttöön liittyvää diagnoosia. Duodecim 98, 461-467.

Poikolainen, Kari (1983): Accuracy of hospital discharge data: five alcohol-related diseases. Drug and Alcohol Dependence 12, 315-322.

Pollard, J. H. (1966): On the Use of the Direct Matrix Product in Analysing Certain Stochastic Population Models. Biometrika 53: 3/4, 397-415.

Pollard, J. H. (1967): Hierarchial Population Models with Poisson Recruitment. Journal of Applied Probability 4, 209-213.

Pollard, J.H. (1973): Mathematical models for the growth of human populations. Cambridge University Press. London.

Prais, S. J. (1955): Measuring Social Mobility. Journal of the Royal Statistical Society. Series A (General) 118: 1, 56-66.

Pulkkinen, Pekka & Kaprio, Jaakko (1980): Terveydenhuollon atk-rekisterien luotettavuudesta ja käytettävyydestä. Sosiaalivakuutus 18: 7-8, 241-242.

Pyke, Ronald (1961a): Markov Renewal Processes: Definitions and Preliminary Properties. Annals of Mathematical Statistics 32: 4, 1231-1242.

Pyke, Ronald (1961b): Markov Renewal Processes with Finitely Many States. Annals of Mathematical Statistics 32: 4, 1243-1259.

Reid, Nancy (1981): Estimating the median survival time. Biometrika 68: 601-608.

Reuter, James. A. (1981): When to evaluate: Evidence from the hazard function. Teoksessa: Tilquin 1981.

Reuter, James & Von Korff, Michael (1981): Chronic Patients in Mental Health Centers. Administration in Mental Health 7: 3, 187-199.

Rissanen, Pekka (1996): Effectiveness, costs and cost-effectiveness of hip and knee replacements. STAKES Research reports 64. Helsinki.

Rissanen, Pekka & Noro, Anja (toim.) (1999): Ikääntyvien potilaiden hoito- ja kotiuttamiskäytännöt - Rekisteripohjainen analyysi aivohalvaus- ja lonkkamurtumapotilaista. Aiheita 44/1999. Helsinki.

Ronkainen, Suvi (1998): Kaikuva empiirisyys - surveyn epistemologiset mahdollisuudet. Teoksessa: Paananen ym. 1998.

Ross, Sheldon M. (1970): Applied probability models with optimization applications. Dover Publications 1992. New York.

Royston, Geoff (1998): Shifting the balance of health care into the 21st century. European Journal of Operational Research 105, 267-276.

Ruotsalainen, Pekka (1991): Menetelmä alueellisen sairaalahoidon tarjonnan määrittämiseksi. Sosiaali- ja terveyshallituksen tutkimuksia 13/1991. Helsinki.

Sairaalaliitto (1983). Toimenpidenimikkeistö 1983.

Sales, Pauline (1971): The Validity of the Markov Chain Model for a Class of the Civil Service. Statistician 20: 1, 85-110.

Salmela, Reijo & Koistinen, Veli (1987): Yleissairaaloiden poistoilmoitusrekisterin kattavuus ja luotettavuus. Sairaala 49, 480-482.

Salo, Mikko (päätoim.) (1989): Sosiaalipolitiikka 1988 - 13. sosiaalipoliittisen yhdistyksen vuosikirja. Sosiaalipoliittinen yhdistys. Helsinki.

Salokangas, Raimo K. R. & Stengård, Eija & Honkonen, Teija & Koivisto, Anna-Maija & Saarinen, Soile (2000): Sairaalasta yhteiskuntaan - Seurantatutkimus sairaalasta kotiuttamisen vaikutuksista skitsofreniapotilaan elämään ja hoitotilanteeseen. Stakesin raportteja 248. Helsinki.

Schinnar, Arie P. & Stewman, Shelby (1978): A Class of Markov Models of Social Mobility with Duration Memory Patterns. Journal of Mathematical Sociology 6, 61-86.

Shorrocks, A. F. (1978): The Measurement of Mobility. Econometrica 46: 5, 1013-1024.

Simpura, Jussi (1989): Tietojärjestelmät sosiaalipoliittisen tutkimuksen kohteena. Teoksessa: Salo 1989.

Singpurwalla, Nozer D. & Wong, Man-Yuen (1983): Estimation of the Failure Rate - A survey of Nonparametric Methods. Part I: Non-Bayesian Methods. Communications in Statistics A, Theory and Methods 12: 5, 559-588.

Sosiaalivakuutus (1980). Tietorekisterit terveyden tutkimuksessa. Sosiaalivakuutus 18: 7-8, 201-253.

Suvisaari, Jaana (1999): Incidence and risk factors of schizophrenia in Finland. Publications of the National Public Health Institute, KTL A16/1999. Helsinki.

Stakes (1995): Tautiluokitus ICD-10. Ohjeita ja luokituksia 1995:3. Stakes. Helsinki.

Stakes (1996a). Nomeskon leikkausluokituksen suomalainen versio. Ohjeita ja luokituksia 1996:3. Stakes. Helsinki.

Stakes (1996b). Hilmo - Sosiaalihuollon ja terveydenhuollon laskenta 1996 ja hoitoilmoitus 1997. Ohjeita ja luokituksia 1996:6. Stakes. Helsinki.

Stakes (1999a). Tautiluokitus ICD-10 2. uudistettu painos. Ohjeita ja luokituksia 1999:2. Stakes. Helsinki.

Stakes (1999b). Hilmo - Sosiaalihuollon ja terveydenhuollon hoitoilmoitus 1999. Ohjeita ja luokituksia 1999:3. Stakes. Helsinki.

Schwartz, Jacob T. (1961): Introduction to Matrices and Vectors. McGraw-Hill Book Company. New York.

Sund, Reijo T. (2000): Simulointimenetelmiä dynaamisten potilaspopulaatioiden mallintamiseen. Julkaisematon käsikirjoitus.

Tarone, Robert E. & Ware, James (1977): On Distribution-Free Tests for Equality of Survival Distributions. Biometrika 64: 1, 156-160.

Taylor, Gordon J. & McClean, Sally I. & Millard, Peter H. (1998a): Continuous-time Markov Models for Geriatric Patient Behaviour. Applied Stochastic Models and Data Analysis 13, 315-323.

Taylor, Gordon J. & McClean, Sally I. & Millard, Peter H. (1998b): Using a Continuous-time Markov Model with Poisson Arrivals to Describe the Movements of Geriatric Patients. Applied Stochastic Models and Data Analysis 14, 165-174.

Teperi, Juha (1994): Use of Registry Data in the Analysis of Medical Practises - Cesarean birth in the Finnish Medical Birth Registry. STAKES Research Reports 47. Stakes. Helsinki.

Thomas, Marlin U. & Barr, Donald R. (1977): An Approximate Test of Markov Chain Lumpability. Journal of the American Statistical Association 72, 175-179.

Tilquin, C. (toim) (1981): Systems Science in Health Care. Pergamon Press. Ontario.

Trussel, James & Hankinson, Richard & Tilton, Judith (toim.) (1992): Demographic Applications of Event History Analysis. Oxford University Press. New York.

Tuma, Nancy Brandon & Hannan, Michael T. & Groeneveld, Lyle P. (1979): Dynamic Analysis of Event Histories. American Journal of Sociology 84: 4, 820-854.

Ugwuowo, Fidelis I. & McClean, Sally I. (2000): Modelling heterogeneity in a manpower system: a review. Applied Stochastic Models in Business and Industry 16, 99-110.

Urponen, Kyösti (1987): Tietojärjestelmät keskustelun kohteena. Sosiaaliturva 75: 10, 528-532.

Uusitalo, Hannu & Parpo, Antti & Hakkarainen, Anni (toim.) (2000): Sosiaali- ja terveydenhuollon palvelukatsaus 2000. Stakesin raportteja 250. Helsinki.

Valkonen, Tapani & Koskinen, Seppo & Martelin, Tuija (toim.) (1998a): Rekisteriaineistot yhteiskunta- ja terveystutkimuksessa. Gaudeamus. Helsinki.

Valkonen, Tapani & Koskinen, Seppo & Martelin, Tuija (1998b): Hallinnolliset ja tilastolliset tietorekisterit tutkimusaineistona. Teoksessa: Valkonen ym. 1998a.

Vardi, Y. (1982): Nonparametric Estimation in Renewal Processes. Annals of Statistics 10: 3, 772-785.

Vassiliou, P.-C. G. (1982): On the Limiting Behaviour of a Non-homogenous Markovian Manpower Model with Independent Poisson Input. Journal of Applied Probability 19, 433-438.

Vassiliou, P.-C. G. (1998): The Evolution of the Theory of Non-Homogeneous Markov Systems. Applied Stochastic Models and Data Analysis 13, 159-176.

Vassiliou, P.-C. G. & Gerontidis, I. (1985): Variances and Covariances of the Grade Sizes in Manpower Systems. Journal of Applied Probability 22, 583-597.

Vesala, Hannu (toim.) (1999): Kehitysvammatutkimusta 1997. Artikkeleita 3. Kehitysvammatutkimuksen konferenssista. Valtakunnallisen tutkimus- ja kokeiluyksikön julkaisuja 78. Kehitysvammaliitto. Helsinki.

Von Korff, Michael (1978): The Dynamics of the Public Mental Hospital Population of Dutchess County New York 1964-1970 (Thesis). The Johns Hopkins University. Maryland.

Wallenius, Riitta (1980): Päihdekäytön seurausten tilastointi - Tutkimus tilaston ja yhteiskunnan suhteista. Alkoholipoliittisen tutkimuslaitoksen tutkimusseloste 140. Helsinki.

Weiss, George H. & Zelen, Marvin (1965): A Semi-Markovian Model for Clinical Trials. Journal of Applied Probability 2, 269-285.

Widder, David Verdon (1945): What is the Laplace Transform? American Mathematical Monthly 52:8, 419-425.

Wu, Lawrence L. (1989): Issues in Smoothing Empirical Hazard Rates. Teoksessa: Clogg 1989.

Yakowitz, Sidney J. (1976): Small-Sample Hypothesis Tests of Markov Order, with Application to Simulated and Hydrologic Chains. Journal of the American Statistical Association 71, 132-136.

Yamaguchi, Kazuo (1991): Event History Analysis. Applied Social Research Methods Series 28. Sage Publications. Newbury Park.