

Satu Soivanen

**KATIHA-
TILASTOSOVELLUS**
Väestötieteellisen tietokanta-aineiston
tilastolliset vaatimukset

Opinnäytetyö
Sähköinen asiointi ja arkistointi


Marraskuu 2011




MIKKELIN AMMATTIKORKEAKOULU

Mikkeli University of Applied Sciences

KUVAILULEHTI

 MIKKELIN AMMATTIKORKEAKOULU Mikkeli University of Applied Sciences		Opinnäytetyön päivämäärä 26. marraskuuta 2011
Tekijä(t) Satu Soivanen	Koulutusohjelma ja suuntautuminen Sähköinen asiointi ja arkistointi	
Nimeke KATIHA-tilastosovellus: Väestötieteellisen tietokanta-aineiston tilastolliset vaatimukset		
Tiivistelmä <p>Opinnäytetyö tehtiin Karjala-tietokannan tilastosovelluksen ohjelmoinnin ja testauksen tueksi. Tilastosovelluksen laatimisen taustalla oli tarve tarjota luovutetun alueen väestötietoja tieteellisen tutkimuksen lähdeaineistoksi. Karjala-tietokantaan jo tallennetut 9 miljoonaa henkilömerkintää vuosilta 1680–1950 antoivat perustellun lähtökohdan tilastollisten laskelmien toteuttamiselle.</p> <p>KATIHA-tilastosovelluksen suunnitteluvaiheessa todettiin, että tietokanta-aineisto sisältää rippi- ja lastenkirjojen lisäksi 70 seurakunnan syntyneiden, kuolleiden, vihittyjen ja muuttaneiden luettelot. Samalla tiedostettiin, että kirkollisten luetteloiden tiedoista voidaan demografisia tutkimusmenetelmiä käyttäen tutkia sekä väestön määrää ja rakennetta että väestön muutoksia. Tiedon käsittelyssä tutkijoiden valittavana havaittiin olevan kolme tilastollista menetelmää: tilastollinen tiivistäminen, päättely ja mallintaminen.</p> <p>Opinnäytetyössä sovellettiin tilastolliseen päättelyyn sisältyvää hypoteesin menetelmän loogisen rekonstruktion mallia tietokanta-aineiston kuolinsyiden luokittelussa. Historiallisten kuolinsyiden luokittelun apuna käytettiin mm. digitoituja sanakirjoja vuosilta 1801 ja 1859. Manuaalisen 120 tuntia kestäneen kuolinsyiden koodaamisen lisäksi toteutettiin ohjelmallinen luokitus sekä ammattien että kuolinsyiden osalta. Ammattinimikkeitä oli jo aiemmin luokiteltu kansainvälisen HISCO-luokituksen mukaisesti. Kuolinsyiden luokituksen perustaksi oli 10 vuotta sitten valittu kuolemansyynimistö vuodelta 1935. Luokituksissa käytetyistä koodistoista dokumentoitiin metatiedot JHS Koodistot ja luokitukset –suositusluonnoksen mukaisesti. Lisäksi luokituksille ja koodistoille määriteltiin laatuvaatimukset.</p> <p>Lopuksi todettiin, että tilastosovelluksen tietokanta-aineiston luokittelua täytyy vielä täydentää ja kehittää esim. hierarkkisten koodistojen suuntaan. Ammattien osalta mallia voisi ottaa EU:n tilastoviranomaisten ohjeistuksen mukaan laaditusta JHS 150 Ammattiluokitus –suosituksesta. Kuolinsyiden koodistoa voitaisiin puolestaan kehittää WHO:n julkaiseman ICD-10 -tautinimistön ryhmittelyn pohjalta.</p>		
Asiasanat (avainsanat) väestötiede, väestö, syntyvyys, kuolleisuus, avioituvuus, muuttoliike, kirkonkirjat, seurakunnan luettelot, väestötieteellinen tietokanta, tilasto, luokittelu, ammattiluokitus, kuolinsyyluokitus		
Sivumäärä 72 s. + liitteet 8 s.	Kieli Suomi	URN
Huomautus (huomautukset liitteistä)		
Ohjaavan opettajan nimi Jukka Selin	Opinnäytetyön toimeksiantaja Karjala-tietokantasäätiö	

DESCRIPTION

 <p>MIKKELIN AMMATTIKORKEAKOULU Mikkeli University of Applied Sciences</p>		Date of the master's thesis 26 November 2011	
Author(s) Satu Soivanen		Degree programme and option eServices and Digital Archiving	
Name of the master's thesis KATIHA Statistics Program: Statistical Requirements for the Data of a Demographic Database			
Abstract This master's thesis was drawn up to support the programming and testing of the statistics program of Karelian Database due to the need to offer demographic data of the Karelia for researchers' use. The 9 million already recorded person data records from the years between 1680 and 1950 gave a reason and a starting point for producing the statistical calculations of Karelian Database.			
<p>During the planning of the KATIHA Statistics Program three relevant aspects came up. First, the database material besides catechetical registers and children books also consisted of parish registers on births, deaths, marriages and the migration of 70 parishes. Second, the demographic methods could be used to study parish registers for finding out quantities and structures of the population and its demographic events. Third, three statistical products could be found after processing the data i.e. summaries, conclusions and models.</p> <p>The method used in this master's thesis was the logical reconstruction of hypothesis in which the conclusion had to be drawn up before the process of analyzing the data. This method was applied to practice in the classification of causes of death. Historical causes of death were classified by using, for instance, digitalized dictionaries from the years 1801 and 1859. The manual classification took about 120 hours after which also an automatic classification for causes of death and occupations was planned and tested. The occupations in the database material were coded by using the international HISCO classification. Ten years earlier the disease terminology from the year 1935 was chosen to be the basis for coding the causes of death. The metadata of the classifications of Karelian Database were documented by following the instructions of the JHS Classification and Codes. The quality requirements of classifications were also determined.</p> <p>Finally, when completing this work it appeared that the classification work had to be continued and developed further. One method to progress to this direction is the idea to make the codes more useful by creating new groups within them. The models of hierarchic codes could be found e.g. either in JHS Number 150 which consists of the references of the classification of occupations or in ICD-10 which is the international classification of diseases published by WHO.</p>			
Subject headings, (keywords) demography, population, natality, mortality, nuptiality, migration, church records, parish registers, demographic database, statistics, classification, coding of occupations, coding of causes of death			
Pages 72 pages + appendices 8 pages	Language Finnish	URN	
Remarks, notes on appendices			
Tutor Jukka Selin		Master's thesis assigned by Karelian Database Foundation	

SISÄLTÖ

1	JOHDANTO.....	1
2	DEMOGRAFISET TUTKIMUSMENETELMÄT	3
2.1	Väestötutkimuksen historiaa	3
2.2	Väestön tilaa mittaavat väestötekijät	4
2.2.1	Väkiluku	5
2.2.2	Väestön rakennetekijät	7
2.3	Väestönmuutoksia kuvaavat väestötekijät.....	9
2.3.1	Syntyvyys	10
2.3.2	Avioituvuus	11
2.3.3	Muuttoliike	11
2.3.4	Kuolleisuus.....	12
3	TILASTOT	13
3.1	Tilastollinen tiivistäminen	14
3.2	Tilastollinen päättely.....	14
3.3	Tilastollinen mallintaminen.....	16
3.4	Tilastollinen tutkimus	16
3.4.1	Tutkimusaineisto	17
3.4.2	Tutkimusaineiston ajallinen tarkastelu	17
3.4.3	Tilastolliset mitta-asteikot	18
3.4.4	Tilastolliset tunnusluvut	20
3.5	Tilastojen käytäntesäännöt	21
4	KARJALA-TIETOKANTA.....	23
4.1	Tietokannan lähdeaineisto	24
4.2	KATIHA-tiedonhakuohjelma.....	26
5	KATIHA-TILASTOSOVELLUS	27
5.1	Tietojärjestelmävaatimukset.....	28
5.2	Käyttäjän vaatimukset.....	30
5.3	Tiedonhaun käyttötapaukset.....	31
5.3.1	Syntyneiden luetteloiden tilastollinen laskenta.....	33
5.3.2	Kuolleiden luetteloiden tilastollinen laskenta	34
5.3.3	Vihittyjen luetteloiden tilastollinen laskenta.....	35
5.3.4	Muuttaneiden luetteloiden tilastollinen laskenta	36

6	KATIHA-TILASTOSOVELLUKSEN TIETOKANTA.....	37
6.1	Väestönaineiston tilastoinnin lähtökohdat.....	37
6.2	Ammattien luokittelu	39
6.3	Kuolinsyiden luokittelu	40
6.3.1	Kuolinsyyluokittelun historiallista taustaa	41
6.3.2	Kuolinsyiden luokittelu Suomessa.....	42
6.3.3	Karjala-tietokannassa esiintyvien kuolinsyiden luokittelu	44
6.4	KATIHA-tilastosovelluksen luokiteltu tietokanta-aineisto	49
6.4.1	Luokitellun aineiston tilastokäyttö.....	55
6.4.2	Luokittelussa käytettyjen koodistojen metatiedot.....	57
6.4.3	Luokitusten ja koodistojen laatuvaatimukset	60
7	PÄÄTÄNTÖ	64
	LÄHTEET	67

LIITTEET

- 1 Tilastojen käytännesäännöt
- 2 Ammattinimikkeiden HISCO-luokitus
- 3 Kuolemansyynimistö 1935
- 4 Luokitusohjelma Kuol_amm.prg
- 5 Luokitusohjelma Kuol_syy.prg
- 6 Esimerkki kuolinsyiden tilastoinnista

1 JOHDANTO

Verkkoympäristöstä löytyy jo runsaasti tutkimusaineistoa ja myös väestötieteellistä lähdeaineistoa sähköisessä muodossa. Karjala-tietokanta on yksi tällainen väestöhistoriallisen tiedon lähde. Luovutetun alueen kirkonkirjojen tarjoama lähdeaineisto on kulttuurihistoriallisesti ainutlaatuinen, koska se muodostaa ajallisesti ja alueellisesti yhtenäisen asiakirja-aineiston tiettyine erityispiirteineen.

Opinnäytetyön taustalla on tarve mahdollistaa Karjala-tietokantaan sisältyvän väestötieteellisen tutkimusaineiston tiedonhaku internetissä. Mikkelissä toimivan Karjala-tietokantahankkeen juuret ulottuvat 1980-luvun puoliväliin, jolloin malli tietokannan kokoamiseen saatiin Ruotsista, Demografista Databasen –hankkeen aloittamisesta. Tämän nykyisin Uumajan yliopiston yhteydessä toimivan tietokantayksikön tavoite on tieteellisen tutkimustoiminnan edistäminen, mikä on myös Karjala-tietokannan koamisen peruste ja päätavoite.

Tietosuojalautakunnan myöntämän poikkeusluvan (Nro 48/30.11.1992) nojalla Karjala-tietokantaan on tallennettu jo noin 20 vuoden ajan henkilötietoja luovutetun alueen kirkonkirjoista, joita säilytetään Mikkelin maakunta-arkistossa. Tallennettavien tietojen yhteismääräksi on arvioitu 10 miljoonaa henkilötietuetta eli -merkintää. Henkilömerkintä muodostuu kirkonkirjasivulle kirjatun yhden henkilön tiedoista, jotka yleensä on merkitty sivun yhdelle riville. Tällä hetkellä tallennettujen henkilömerkintöjen määrä on ylittänyt jo 9 miljoonaa. Rippi- ja lastenkirjoista on tallennettu yli 5 400 000 merkintää ja muista kirkonkirjoista yhteensä noin 3 660 000 henkilömerkintää.

Karjala-tietokannan tiedonhakua varten on kehitetty KATIHA-henkilöhakuohjelma, jonka avulla käyttäjä voi helposti vaillinaisinkin hakuehdoin hakea etsimäänsä henkilöä. KATIHA-sovelluksen aktiivisin käyttäjäryhmä muodostuukin asiakkaista, jotka hyödyntävät henkilöhakumahdollisuutta sukututkimuksen teossa. Tieteellisen tutkimuksen tiedontarpeet eivät kuitenkaan niinkään liity yksilöihin vaan tapahtumiin ja tilastollisiin ilmiöihin. Karjala-tietokantaan tallennetuista henkilötiedoista löytyy paljon tietoja, jotka voisivat antaa vastauksia tieteellisen tutkimuksen asettamiin kysymyksiin luovutetun Karjalan väestön osalta. Väestötieteellisten perustietojen hakuun liittyvä tilasto-osio täydentäisi tietoverkossa jo toimivaa KATIHA-henkilöhakuohjelmaa ja lisäisi tietokanta-aineiston hyödyntämistä erilaisiin tutkimuksiin.

Tässä opinnäytetyössä selvitetään Karjala-tietokannan tilasto-osion toteuttamisedellytyksiä ja tilastollisia vaatimuksia. Tutkimusongelmana on määrittää KATIHA-tilastosovellukseen liitettävän tietokannan sisältö ja tietorakenne. Lähtökohtana on, että tilastosovellus yhdistettäisiin KATIHA-henkilöhakuohjelman jo verkossa olevaan tietokantaan, jolloin uusi sovellus toimisi valmistuttuaan Karjala-tietokannan verkkoaineiston toisena käyttöliittymänä.

Opinnäytetyön teoreettisen viitekehyksen aluksi tutustutaan demografisiin tutkimusmenetelmiin, jotka kohdistuvat väestön määrään ja rakenteisiin sekä väestömuutoksiin. Sen jälkeen käsitellään tilastollisten menetelmien hyödyntämistä tutkimuksissa sekä selvitetään tilastojen kokoamiseen ja käyttöön liittyviä säädöksiä ja käytänteitä.

Neljännessä luvussa esitellään Karjala-tietokantahankkeen taustaa ja nykytilannetta painottuen tietokannan tietorakenteen ja KATIHA-henkilöhakuohjelman kuvaukseen. Tämän opinnäytetyön soveltavan osuuden muodostavat tekstin viides ja kuudes luku. Viidennessä luvussa pohditaan sekä tilastosovelluksen tietojärjestelmää että käyttäjää koskevia vaatimuksia. Käyttäjän vaatimuksiin liittyen määritellään myös, millaisia tilastollisia laskelmia kirkonkirja-aineistosta voidaan toteuttaa syntyneiden, kuolleiden, vihittyjen ja muuttaneiden osalta.

Kuudennessa luvussa perustellaan tarve luoda Karjala-tietokannan aineistosta erillinen tietokanta tilastosovellusta varten. Uuden tietokannan muodostamisen yhteydessä todetaan tietokanta-aineistoon liittyvät vaatimukset sekä ammattien että kuolinsyiden luokittelusta. Näiden tilastollisten vaatimusten osalta poraudutaan erityisesti kuolinsyiden luokituksessa esiintyviin haasteisiin. Lisäksi esitellään luokitusten yhdistäminen uuteen tilastosovelluksen tietokantaan sekä testataan kuolleiden tilastointia kuolinsyiden mukaan. Samoin selvitetään luokitusten pohjana olevien koodistojen metatiedot sekä luokitusten ja koodistojen laatuvaatimukset, jotta verkossa toimiva tilastosovellus olisi luotettava tiedonlähde historian tutkijoille, yhteiskunta- ja väestötieteilijöille sekä kaikille muille Karjala-tietokannan tilastotietojen tarvitsijoille.

Opinnäytetyön päätteeksi pohditaan KATIHA-tilastosovelluksen tietokanta-aineiston luokittelun lähtökohtia ja toteuttamista sekä koodistojen mahdollista kehittämistä tutkijoita paremmin palvelevaan muotoon.

2 DEMOGRAFISET TUTKIMUSMENETELMÄT

Monikielinen väestötieteen sanakirja määrittelee väestötieteen seuraavasti: ”Väestötiede eli demografia on tieteenala, jonka tehtävänä on tutkia ihmisten muodostamia populaatioita, väestöjä, erityisesti niiden suuruutta, rakennetta ja kehitystä sekä niiden kvantitatiivisesti ilmaistavissa olevia ominaisuuksia”. (Fougstedt ym. 1962, 21.)

Väestötiede voidaan myös jakaa erilaisiin osa-alueisiin tai tutkimushaaroihin. Kari Pitkänen ja Ari-Pekka Sihvonen esittelevät Helsingin yliopiston julkaisemassa verkkomateriaalissa ”Johdatus väestötieteeseen” (Pitkänen & Sihvonen 2003) väestötieteestä taulukon 1 mukaisen jaon kolmeen eri tutkimussuuntaan.

TAULUKKO 1. Väestötieteen kolme perinteistä tutkimussuuntaa

Deskriptiivinen	Formaalinen eli teoreettinen	Väestöanalyysi
Väestöilmiöiden kuvailu	Väestötekijöiden tarkastelu	Väestöjen tutkiminen
Väestötilastot	Mittausmenetelmät	Väestöselvitykset

Deskriptiivinen väestötiede on pitkälti väestöllisten ilmiöiden kuvailua ja väestötilastojen laatimista. Formaalinen eli teoreettinen väestötiede on väestötekijöiden yhteyksien tarkastelua käyttäen matemaattisia kaavoja ja malleja usein abstraktien väestöjen avulla. Väestöanalyysi on puolestaan kuvailua pidemmälle vietyä todellisten väestöjen ja niiden kehityksen tutkimista. (Pitkänen & Sihvonen 2003.)

2.1 Väestötutkimuksen historiaa

Laatimassaan väestötilaston historiakatsauksessaan (Nieminen 1999, 6) tilastokeskuksen kehittämispäällikkö Mauri Nieminen mainitsee, että demografisten tutkimusmenetelmien ja samalla väestötieteen isäksi voidaan nimetä John Graunt. Tämä lontoolainen kangaskauppias teki 1600-luvun puolessa välissä omia demografisia laskelmia ja laati laskelmistaan mm. elinajantauluja. Syntyneiden ja kuolleiden lukumäärien muutoksista hän laski Lontoon väkiluvuksi 384 000 ja teki havainnon, että poikia syntyy enemmän kuin tyttöjä. John Grauntin tutkimusten perusaineistona oli Lontoon seurakuntien viikoittaiset kuolleiden luettelot, Bills of Mortality (Pressat 1972). Näistä luet-

teloista ilmeni muun muassa kuolinsyy, joka oli arvokas tieto mm. kulkutautien yleisyyden takia. Varakkaat hyödynsivät tietoa kulkutautien määrän kasvusta esim. siten, että muuttivat kaupungin ulkopuolelle epidemian ajaksi ja palasivat sitten myöhemmin takaisin kaupunkiin kulkutautivaaran ollessa ohi. Lääketieteen professori Arno Forsius esittelee Grauntin tutkimuksia ”Tautien nimistöistä ja luokituksista” – artikkelissaan (Forsius 2010b). Forsiuksen selvityksen mukaan Graunt tarkasteli tutkimuksessaan kuolintodistuksia Lontoossa yli 60 vuoden ajalta ja havaitsi, että kuolleisuus oli kaupungeissa suurempi kuin maaseudulla. Graunt totesi myös, että miesten ja naisten lukumäärät väestössä muodostuivat lähes samansuuruisiksi, sillä miesten suurempi kuolleisuus tasoitti lukumäärät, vaikka poikia syntyikin tyttöjä enemmän. Tutkiessaan kuolleisuutta Graunt ryhmitteli kuolleisuuden kuolinsyiden mukaisesti. Kuolinsyiden luokittelu on siten aloitettu jo 1600-luvulla.

Tilastokeskuksen kehittämispäällikkö Mauri Niemisen (1999, 8) mukaan Suomessa vuoden 1686 kirkkolaki oli tärkeä väestötieteellisten tutkimusten kannalta, koska laissa määriteltiin monipuolisesti ne tiedot, joita myöhemmin hyödynnettiin väestötilastoja kehitettäessä. Syntyneiden, kuolleiden ja avioliiton solmineiden tietojen merkitsemisen lisäksi papisto velvoitettiin rekisteröimään myös tietoja muuttoliikkeestä eli ”mistä tullut ja mihin lähtenyt”. Nieminen toteaa, että kirkonkirjat muodostavat arvokkaan demografisen tietolähteen. Se, että papit olivat yhteistyössä henkikirjoitusta toimittavien virkamiesten kanssa, edesauttoi tietojen oikeellisuuden tarkistamisessa. Eri henkilöluetteloissa olevia tietoja vertailtiin ja toisinaan havaittiin puutteita tietojen merkitsemisessä. Kansallisarkiston tutkimusjohtaja Päivi Happonen toteaa kirkonkirjojen ja henkikirjojen tietojen vertailuun painottuvassa väitöskirjassaan, että lähteiden tietosisällön puutteellisuus ja lähteiden luotettavuuteen liittyvät seikat rajoittavat käytännössä tieteellisten menetelmien hyödyntämistä väestöhistoriallisissa tutkimuksissa. (Happonen 2009, 14.)

2.2 Väestön tilaa mittaavat väestötekijät

Lähdetietojen kattavuus ja luotettavuus on tärkeää, sillä väestötieteessä tarkastellaan nimenomaan erilaisia mitattavissa olevia väestötekijöitä (Pitkänen & Sihvonen 2003). Nämä ominaisuudet jaetaan kuvan 1 mukaisesti kahteen ryhmään: 1. väestön tila ja 2. väestön muutokset.

1) väestön tila	2) väestön muutokset
<ul style="list-style-type: none"> ■ väestön määrä ■ väestön alueellinen sijainti ■ väestön rakennetekijät <ul style="list-style-type: none"> + ikä(rakenne) + sukupuoli(rakenne) + siviilisääty(rakenne) 	<ul style="list-style-type: none"> ■ syntymät ■ kuolemat ■ muutot / asuinpaikan muuttumiset ■ siviilisäädyn muuttumiset

KUVA 1. Väestötieteessä tutkittavat väestötekijät (Pitkänen & Sihvonen 2003)

Tilastokeskuksen verkkomateriaalin ”Johdatus väestötieteen perusteisiin” kirjoittaja Mauri Nieminen opettaa, että väestötiede tutkii väestönkehitystä, eli syntyneiden, kuolleiden ja muuttaneiden määrää sekä muita väestönmuutokseen vaikuttavia ilmiöitä, joita ovat väestön avioituvuus ja eronneisuus. Niemisen mukaan väestötieteessä tarkastellaan myös sitä, miten väestössä tapahtuvat muutokset vaikuttavat yhteiskunnan taloudelliseen ja sosiaaliseen kehitykseen. (Nieminen 2011.)

Väestön tilaan kuuluvat erityisesti väestön määrä eli väkiluku, alueellinen jakautuminen sekä väestön ikä-, sukupuoli- ja siviilisäätyrakenne. Teoksessa ”Suomen väestö” mainitaan lisäksi, että väestön rakenteesta voidaan saada monipuolinen kuva, jos se jaetaan esim. kielen, syntymäpaikan, koulutuksen ja ammatin mukaisiin ryhmiin. Väestönmuutostekijöiksi nimitetään puolestaan tekijöitä, joiden vaikutuksista väestön määrä ja rakenne muuttuvat. Väestön määrän ja sen rakenteen kehitystä tutkittaessa on tarpeen havainnoida erilaisia väestötapahtumia, kuten syntymiä ja kuolemia, asuinpaikan vaihdoksia sekä siirtymiä siviilisäädystä toiseen. (Koskinen ym. 2007, 15.)

2.2.1 Väkiluku

Kirkollisten luetteloiden merkitys väkiluvun laskennan ja väestöllisten ilmiöiden kuvaamisen perusaineistona havaittiin jo 1700-luvun alkupuolella (Nieminen 1999, 8). Väestötilaston historiakatsauksessaan Nieminen (1999, 10) mainitsee kauppakollegion asessori Jakob von Hökerstedtin, joka oli vuonna 1738 jättänyt valtakunnanneuvostolle ehdotuksen väkiluvun laskemisesta kirkonkirjojen perusteella. Silloin ehdotusta ei

hyväksytyt raamatullisin perustein. Järjestelmällinen väestötietojen keruu Ruotsi-Suomessa aloitettiin kuitenkin vuonna 1749 perustetun taulustolaitoksen toimesta (Happonen 2009, 24).

Väestötieteen professori Kari Pitkänen lajittelee väestötekijöiden lähteet kolmeen ensisijaiseen lähderyhmään: 1) väestöä koskevat rekisterit, 2) väestötilasto, 3) otoksiin perustuvat kysely- ja haastatteluaineistot sekä muut erillisselvitykset. Suomessa perinteisenä lähdeaineistona on ollut väestötilasto. Väestötilastolla voidaan ymmärtää paitsi tilastoviranomaisten julkaisemia virallisia tilastotaulukoita myös erilaisia arkistotaulukoita, jotka ovat tutkijoiden käytettävissä. Otoksiin perustuvia kyselytutkimuksia tai muita erillisselvityksiä ei teollistuneissa maissa yleensä tehdä väestön määrän, sen rakenteen tai väestönmuutosten määrän selvittämiseksi, sillä tiedot saadaan varsin luotettavasti kansallisista väestötilastoista. Tilanne on tältä osin erilainen kuin kehitysmaissa. (Pitkänen 2003, 1.)

Kun ensimmäiset koko Ruotsi-Suomea koskevat väestötiedot kerättiin vuodesta 1749 alkaen, seurakuntatasolla papit vastasivat tietojen keräämisestä ja niiden taulukoinnista. Tämä vaikutti kirkonkirjatietojen täsmentymiseen, sillä tilastotiedot pohjautuivat rippikirjoihin ja luetteloihin kastetuista, vihityistä, muuttaneista ja kuolleista. Papiston tuli laatia väestönmuutoksia koskevan taulukon lisäksi väkilukutaulukko vuosittain. Väkilukutauluihin väestö luokiteltiin muun muassa iän, sukupuolen, siviilisäädyn ja sosiaalisen aseman mukaan. Seurakuntien tiedoista tehtiin rovastikunnittain, hiippakunnittain ja lääneittäin yhteenvetoja, joiden avulla koottiin koko Ruotsin valtakuntaa koskevat väkilukutiedot vuonna 1756 perustetun Tabellkommissionin toimesta. (Happonen 2009, 24.)

Aluksi väkilukutaulukot laadittiin vuosittain, mutta vuonna 1751 siirryttiin kolmen vuoden jaksotukseen. Vuodesta 1775 väkilukutaulukoita koottiin viiden vuoden välein. Läänikohtaisista yhteenvetoista luovuttiin vuoden 1774 jälkeen ja hiippakunnittaisista vuonna 1792. Väestötietojen keruu säilyi seurakuntatasolla ja Porvoon vuoden 1809 valtiopäivien esityksestä tilastojen tehtävä annettiin senaatin kamaritoimituskunnalle (Nieminen 1999, 16). Vaikka Suomi oli Venäjän autonominen osa, Ruotsin ajan mallia toteutettiin myös vuonna 1812 muun Suomen yhteyteen liitetyn Vanhan Suomen luterilaisissa seurakunnissa. (Happonen 2009, 24.)

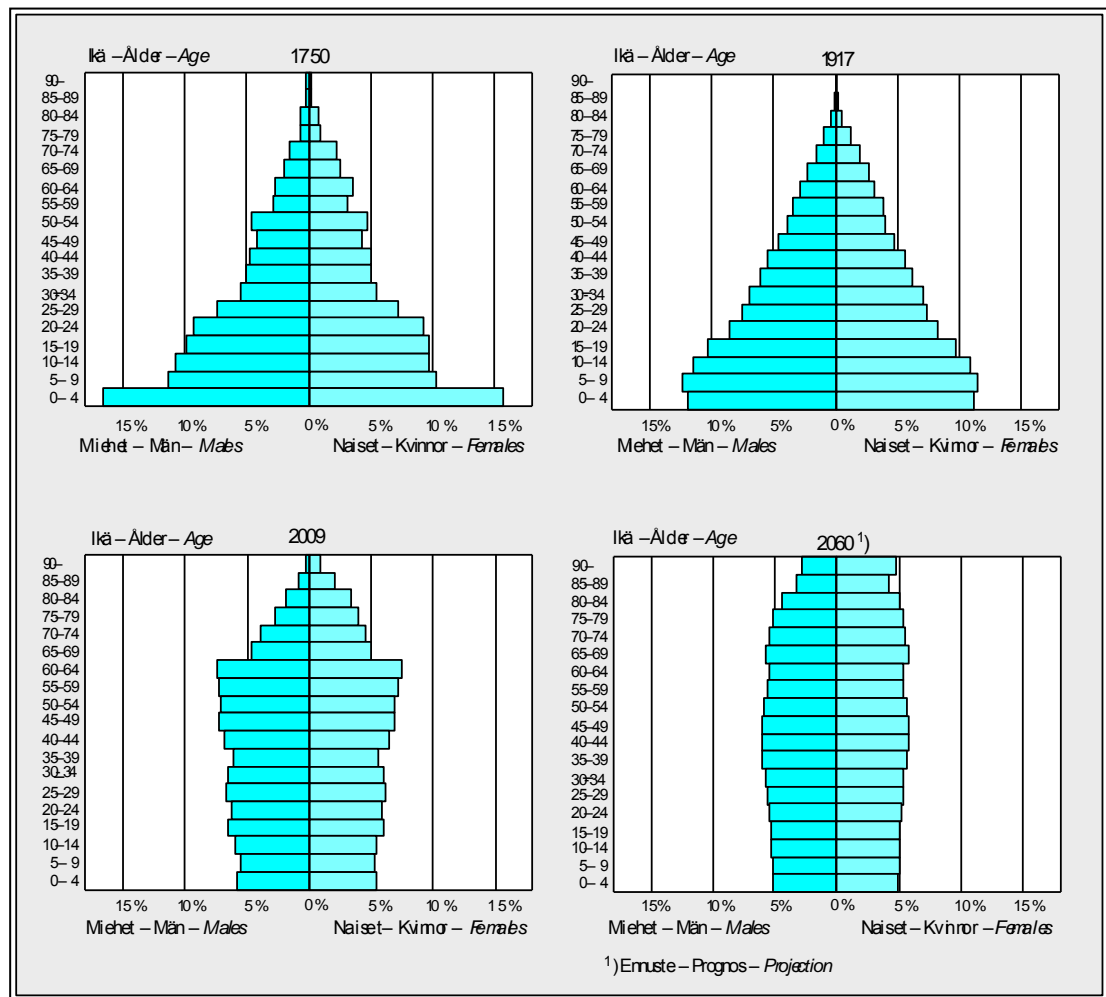
Vuodesta 1749 alkanutta kirkonkirjojen inventointia jatkettiin aina vuoteen 1940 saakka (Haimi 1987, 17). Väestön määrää ja rakennetta koskevien tietojen keruujärjestelmä uudistettiin Suomessa vuonna 1950. Kyseessä oli ensimmäinen koko maan kattanut välitön väestönlaskenta. Tiedot koottiin kyselylomakkeista, jotka oli lähetty kaikkiin kotitalouksiin ja laitoksiin. Lomaketietoihin pohjautuvia laskentoja tehtiin sittemmin vielä vuosina 1960, 1970, 1975, 1980 ja 1985 (Koskinen ym. 2007, 19). Vuoden 1990 väestönlaskenta toteutettiin jo pelkästään rekistereitä hyödyntäen. Henkilötietoja kerättiin noin 30 rekisteristä, joita olivat mm. väestötietojärjestelmä, verohallinnon rekisterit, eläkelaitosten työsuhderekisterit, kansaneläkelaitoksen eläkerekisteri, yritys- ja toimipaikkarekisteri, opiskelijarekisterit, työnhakijarekisteri, tutkintorekisteri ja varusmiesrekisteri (Nieminen 1999, 43).

2.2.2 Väestön rakennetekijät

Suomen väestö –teoksessa todetaan, että väestön ikärakenteen tyypillisin esitystapa on sen kuvaaminen väestöpyramidin avulla (Jalovaara, Koskinen & Martelin 2007, 25). Väestöpyramidissa kunkin ikäryhmän kokoa kuvataan vaakasuoralla palkilla. Miehiä kuvaavat palkit esitetään kuvion keskellä olevan pystyakselin vasemmalla puolella ja naisia kuvaavat palkin oikealla puolella siten, että nuorin ikäryhmä muodostaa pyramidin perustan ja vanhin sen huipun. Ikäryhmän koko voidaan kuvata joko absoluuttisesti tai suhteellisesti.

Absoluuttisten suuruuksien perusteella piirretty väestöpyramidi kuvaa yhtä väestöä tietyllä hetkellä eri ikä- ja sukupuoliryhmien osalta. Tällöin vaaka-akselin asteikko kertoo ihmisten lukumäärän esimerkiksi tuhansina. Suhteellisten osuuksien perusteella laadittavaan väestöpyramidiin väestömäärät lasketaan puolestaan seuraavasti: yksittäisen ikä- ja sukupuoliryhmän väestömäärä jaetaan koko väestön määrällä. Tällöin tietty vaakasuoran palkin pituus vastaa yhtä suurta väestömäärää niin miehiä kuin naisiakin kuvattaessa, jolloin miesten ja naisten suhteellisia määriä voidaan palkkien avulla vertailla. (Jalovaara, Koskinen & Martelin 2007, 25–26.)

Nimitys väestöpyramidi johtuu siitä, että kuvio on perinteisesti muistuttanut pyramidia eli se on ollut ylöspäin kapeneva. Väestöpyramidin muodon perusteella pyritään usein arvioimaan tulevaa väestönkehitystä (Jalovaara, Koskinen & Martelin 2007, 26).



KUVA 2. Väestön ikärakenne 1750–2060 (SVT2010, 101)

Demografisista mittaluvuista kertovassa artikkelissaan Marika Jalovaara, Seppo Koskinen ja Tuija Martelin esittävät väestön kehityksen arvioinnin lähtökohtana sen, että nuorempiin ikäluokkiin kuuluvat ovat tulevaisuuden synnyttäjii (Jalovaara, Koskinen & Martelin 2007, 26). Kirjoittajien mukaan on ilmeistä, että Suomen väestöä kuvaava pyramidi levenee tulevaisuudessa ylöspäin eli vanhemmat ikäluokat ovat suurempia kuin nuoremmat ikäluokat. Tämän suuntaisen kehityksen osoittaa myös kuvassa 2 olevat tilastokeskuksen laatimat väestöpyramidit väestön ikärakenteesta 1750–2060.

Siviilisäätärakenne on yhteydessä väestön ikärakenteeseen. Ikärakenteen vanhetessa naimattomien osuus väestöstä pienenee ja leskien sekä eronneiden osuus suurenee.

Tosin Markku Ryyänen ja Veijo Notkola kirjoituksessaan väestönkehityksestä muistuttavat, ettei siviilisäätyjakauma ei enää olekaan luotettava väestön perherakenteen kuvaaja, sillä nykyisin avoliitot ovat yleistyneet (Ryyänen & Notkola 2007, 284).

Äidinkielellä on oma merkityksensä väestöä kuvaavana tekijänä. Esimerkiksi ruotsinkielisten osuus koko Suomen väestöstä oli noin kahdeksan prosenttia 1950-luvulla. Osuus pieneni kuuteen prosenttiin vuoden 1990 alkuun mennessä ja vuonna 2009 ruotsinkielisiä oli 5,43 prosenttia koko väestöstä (SVT 2010, 99).

Ulkomaalaisten osuus Suomen väestöstä on ollut verraten pieni, mutta osuus kasvaa jatkuvasti. Eniten ulkomaalaisia Suomen väestöön on tähän asti tullut entisestä Neuvostoliitosta tai Venäjältä sekä Ruotsista ja Virosta (Ryyänen & Notkola 2007, 284). Suomalaisen väestön ikääntyessä ulkomaalaisista voi tulla merkittävä Suomen yhteiskuntaan vaikuttava tekijä.

Elinkeinorakenteen muutokset kuvastuvat myös väestörakenteessa. Ryyänen ja Notkolan mukaan vuonna 1960 maa- ja metsätaloudessa työskenteli noin 36 % ammatissa toimivasta väestöstä, mutta 2000-luvun alussa osuus oli enää hieman yli viisi prosenttia (Ryyänen & Notkola 2007, 284). Vastaavasti palveluelinkeinojen osuus on kasvanut voimakkaasti. Suomen tilastollisen vuosikirjan mukaan palvelujen osuus Suomen bruttokansantuotteesta vuonna 2008 oli 65 % (SVT 2010, 664).

2.3 Väestönmuutoksia kuvaavat väestötekijät

Väestön määrää ja rakennetta mittaavien ja ilmaisevien tekijöiden lisäksi väestötieteessä tarkastellaan väestönmuutostapahtumia, joita ovat syntymät, avioliiton solmimiset ja avioerot, muutot sekä kuolemat. Väestötieteen verkkomateriaalissa todetaan, että myös väestönmuutostapahtumilla on oma rakenteensa iän ja sukupuolen mukaan (Nieminen 2011).

Esimerkkinä Nieminen mainitsee kuolleiden määrien tarkastelu iän ja sukupuolen mukaan sekä syntyneiden lasten määrien tutkiminen äidin iän mukaan. Mauri Nieminen toteaa, että väestönmuutostapahtumien yleisyys riippuu hyvin paljon iästä. Lapset synnytetään 25–30-vuotiaina, muuttoliikettä tapahtuu eniten työelämään tulevilla

20–29-vuotiailla ja kuolemat yleistyvät mitä iäkkäimmiksi ihmiset tulevat. Nieminen kuvaa väestönmuutostapahtumia ikäsidonnoiksi. (Nieminen 2011.)

Väestönmuutosten ikä- ja sukupuolirakennetietoja löytyy jo 1750-luvun lähteistä, sillä seurakuntien väestönmuutostauluihin merkittiin kuukausittain syntyneiden, vihittyjen ja kuolleiden määrät ryhmiteltynä iän ja sukupuolen mukaan. Syntyneet jaoteltiin lisäksi aviollisen ja aviottoman syntyperän mukaan ja kuolleet iän mukaan. (Happonen 2009, 24.)

2.3.1 Syntyvyys

Syntyvyyttä mitataan yleisellä syntyvyysluvulla, jossa syntyneiden vuosittainen määrä jaetaan keskiväkiluvulla ja kerrotaan tuhannella (Nieminen 2010). Keskiväkilukuna käytetään yleensä vuoden alun ja lopun väkilukujen keskiarvoa (Pitkänen 2007). Yleinen syntyvyysluku voidaan laskea myös esimerkiksi viiden vuoden ajalta. Tällöin viiden vuoden aikana syntyneiden määristä lasketaan keskiarvo, joka jaetaan ajanjakson keskiväkiluvulla, ja näin saatu luku kerrotaan tuhannella (Nieminen 2010).

Syntyvyyden tutkimuksessa ovat tarkastelun kohteena väestön jälkeläisten syntymiseen suoranaisesti liittyvät laskennalliset tekijät. Olavi Haimi selvittää väestöntutkimuksen periaatteita ja menetelmiä käsittävässä teoksessaan alalla käytettäviä termejä ja tunnuslukuja. Syntyvyysluvulla tarkoitetaan syntymätapausten lukumäärää suhteessa väkilukuun. Se kuvaa siten syntyvyyden suhteellista esiintymistä eli tiheyttä. Käsitettä hedelmällisyys käytetään laskettaessa synnytysten määrää ja ajoitusta hedelmällisyydessä olevien naisten keskuudessa. (Haimi 1897, 129.)

Tarkasteltaessa syntyvyyden yleistä merkitystä väestön kasvulle ja väestörakenteelle kiinnitetään huomiota syntyneiden absoluuttiseen määrään eli vuoden aikana syntyneiden lukumäärään. Toinen vaihtoehto on tutkia syntyneiden suhteellista määrää eli syntyneiden lukumäärää tuhatta asukasta kohti. (Ruokolainen & Notkola 2007, 77.)

2.3.2 Avioituvuus

Avioituvuutta mitataan suhteuttamalla avioliittojen solmimisten määrät keskiväkilukuun. Avioituvuusluku lasketaan siten, että avioliittojen solmimisten määrä jaetaan keskiväkiluvulla ja kerrotaan tuhannella. Mauri Nieminen toteaa, että avioituvuusluvun tulkintaan liittyy samanlaisia ongelmia kuin muidenkin väestöilmiöitä kuvaavien yleisten tunnuslukujen tulkintaan. Laskelmissa käytetty keskiväkiluku sisältää paljon sellaista väestöä, joka ei ole altis kyseiselle väestönmuutostapahtumalle. Avioliiton yleisyys riippuu hyvin paljon iästä, joten Niemisen mukaan avioliiton solmimista kuvaakin parhaiten erilaiset ikäryhmittäin lasketut tunnusluvut. Ikäryhmittäiset tunnusluvut lasketaan yleensä erikseen miehille ja naisille sekä erikseen avioliiton järjestysluvun mukaan. (Nieminen 2011.)

Olavi Haimi huomauttaa, että syntyvyys on ollut suuressa määrin riippuvainen solmitujen avioliittojen määrästä eli avioituvuudesta. Avioituvuuden mittaamisella on ollut tärkeä asema väestöntutkimuksessa, vaikkei se kuulu niihin ilmiöihin, jotka aiheuttavat välittömästi muutoksia väkiluvussa. Väkiluvulla on kuitenkin oma merkityksensä avioitumiseen, sillä avioliiton solmimiseen vaikuttaa sopivan ikäisten miesten ja naisten määrällinen tasapaino. (Haimi 1987, 160.)

Lastensaanti on kautta aikojen keskittynyt avioliittoihin tai niihin rinnastettaviin suhteisiin. Avioliitot ovat siis olleet se yksikkö, jossa niin väestön lisääntyminen on perinteisesti tapahtunut. Avioliiton ulkopuolella syntyneiden osuus 1700-luvun puolivälissä Suomessa oli vain puolitoista prosenttia. Vastaava osuus 1920-luvulla oli kasvanut jo kahdeksaan prosenttiin. (Pitkänen & Jalovaara 2007, 157.)

2.3.3 Muuttoliike

Muuttoliikkeellä tarkoitetaan väestön asuinpaikan vaihdoksia yli valtakunnan rajojen tai näiden rajojen sisäpuolella yli hallinnollisten alueiden rajojen. Haimi kritisoi muuttoliikkeen rajoittamista vain hallinnollisten rajojen yli tapahtuvaan muuttoon, sillä kunnan sisällä tapahtuvat huomattavan pitkätkin muutot jäävät muuttotilastojen ulkopuolelle. (Haimi 1987, 205.)

Muuttaneisuutta mitataan samanlaisilla tunnusluvuilla kuin muitakin väestönmuutoksia. Yleinen muuttaneisuusluku kuvaa muuttaneiden määrää tuhatta keskiväkiluvun asukasta kohti (Nieminen 2011).

Muuttoliikettä tutkineet Jouni Korkiasaari ja Ismo Söderling huomauttavat, että tarkka rajanveto muuttoliikkeen ja tilapäisen liikkuvuuden välillä on vaikeaa ja vaihtelevaa (Korkiasaari & Söderling 2007, 240). Muuttoliike voidaan ymmärtää myös laajempaan käsitteenä liikkuvuus, joka sisältää myös tilapäisen siirtymisen kuten turismin. Muuttoliike voi olla luonteeltaan vapaaehtoista tai pakkoluonteista. Vapaaehtoinen muuttoliike voidaan puolestaan jakaa maan sisäiseen muuttoon ja valtioiden rajat ylittävään muuttoon. Kansainvälistä muuttoliikettä kutsutaan siirtolaisuudeksi (Haimi 1987, 205). Pakkoluonteista muuttamista tapahtuu mm. sotien, poliittisen vainon, nälkähädän ja ympäristötuhojen seurauksena (Korkiasaari & Söderling 2007, 240).

2.3.4 Kuolleisuus

Kuolleisuuden tutkimuksessa tarkastelun kohteena on kuoleman väestöllinen vaikutus. Kuolevuusluvulla tarkoitetaan kuolemantapausten suhteellista määrää väestön keskuudessa (Haimi 1987, 173). Kuolleisuuden tarkastelussa on kaksi laskennallista vaihtoehtoa: vuosittaisen tai pitemmän ajanjakson kuolevuusluvun laskeminen (Nieminen 2011). Ensimmäinen lasketaan siten, että kuolleiden vuosittainen määrä jaetaan keskiväkiluvulla ja kerrotaan tuhannella. Toinen tapa on laskea tietyn ajanjakson aikana kuolleiden lukumääristä keskiarvo ja jakaa se ajanjakson keskiväkiluvulla ja kertoa näin saatu luku tuhannella.

Kuolleisuutta voidaan tarkastella eri näkökulmista. Ensiksikin kuolleisuus on yksi väestönmuutostekijöistä, joka vaikuttaa väestön kokoon ja muovaa sen rakennetta. Koskinen ja Martelin tähdentävät, että kuolleisuuden kehityksen tutkiminen on tärkeää aiemman väestönkehityksen ymmärtämiseksi ja tulevien muutosten ennustamiseksi. Toisaalta kuolleisuutta koskevat tiedot muodostavat kuvan väestön yleisestä terveydentilasta ja luovat perustan myös elinajan odotteen laskemiselle. Kuolleisuustietoja voidaan käyttää laajemmaltikin tieteellisessä tutkimuksessa ilmaisemaan väestön hyvinvointia eli hyvinvoinnin mittarina. (Koskinen & Martelin 2007, 169.)

3 TILASTOT

Tärkein tilastojen tuottaja Suomessa on Tilastokeskus (Grönroos 2004, 1). Sen tehtävä periytyy erilaisten välivaiheiden kautta autonomian ajalla toimineelta senaatin kamaritoimituskunnalta, jonka laatimien tilastollisten yhteenvedojen lyhennelmiä ryhdyttiin 1820-luvulla julkaisemaan *Finlands Allmänna Tidningissä* (Happonen 2009, 24). Vuonna 1865 Suomeen perustettiin tilastollinen päätoimisto, joka alkoi julkaista Suomen viralliseen tilastoon liittyvää sarjaa väestötilastoista. Kyseinen sarja sisältää viiden ja kymmenen vuoden välein julkaistuja väestön rakennetietoja ja vuosittaisia väestönmuutostietoja (Koskinen, Martelin, Notkola, Notkola & Pitkänen 2007). Erillinen kolmiosainen Suomen väestötilaston pääpiirteet vuosilta 1750-1890 julkaistiin 1800-luvun lopulla (Happonen 2009, 25).

Väestötilastojen pohjana oleviin väkilukutauluihin sisältyi paljon myös seurakunnasta ilman muuttokirjaa lähtenyttä eli poissa olevaa väestöä. Väkilukutaulujen luotettavuusongelmien takia tilastoinnin pohjaksi suunniteltiin väestönlaskentaa kirkonkirjatietojen sijaan. Suurimmissa kaupungeissa toteutettiin väestönlaskentoja vuosina 1870-1930, jolloin kaupungeissa asuneet henkilöt tilastoitiin ja tulokset ryhmiteltiin muun muassa sosiaalisin, kielellisin ja uskonnollisin perustein. Väestötilastoissa esitetyt uskonnolliset jakaumat perustuivat kuitenkin väestönlaskentatietojen sijasta kirkonkirjoihin. (Happonen 2009, 25.)

Vuonna 1938 säädettiin laki, jonka mukaan väestörekisteriin perustuvat väestötilastot korvattiin koko maahan ulottuvilla välittömillä ja joka kymmenes vuosi suoritettavilla väestönlaskennoilla (Nieminen 1999, 31). Tosin toinen maailmansota ja siitä toipuminen estivät suunnitelmien toimeenpanon reilun vuosikymmenen ajan. Ensimmäinen koko maata koskeva väestönlaskenta saatiin toteutettua vuoden 1950 lopussa. Vuodesta 1990 väestönlaskennat on tehty sähköisesti eri rekisterien tietoja yhdistelemällä (Happonen 2009, 25). Rekisteripohjainen väestönlaskenta on mahdollistanut sen, että lähes kaikki väestönlaskentatiedot tuotetaan ja julkaistaan nykyisin vuosittain (Nieminen 1999, 43). Suomessa hallinnollisten rekisteriaineistojen käyttö tilastotuo-
tannossa on kansainvälisesti huippuluokkaa (Tilastokeskus 2011).

Neil J. Salkind (2008, 9) tiivistää tilaston tarkoittavan työkalua, joka auttaa meitä ymmärtämään ympäröivää maailmaamme. Tilaston avulla järjestämme keräämäämme tietoa uudelleen ja havainnoimme sitä tietyin perustein (Salkind 2008, 9). Lauri Nummenmaa (2009, 16) jakaa kerätyn tiedon käsittelyn kolmeen erilaiseen tilastolliseen menetelmään: 1. tietoa tiivistävä ja kuvaileva menetelmä, 2. ilmiötä koskevia päätöksiä tehtäessä käytettävä menetelmä ja 3. ilmiötä matemaattisesti mallintava menetelmä.

3.1 Tilastollinen tiivistäminen

Tilastollinen tiivistäminen tarkoittaa menetelmää, jonka tavoitteena on kuvailla suuren lukujoukon ominaisuuksia huomattavasti vähäisemmällä määrällä lukuja (Nummenmaa 2009, 17). Esimerkiksi keskiarvon laskeminen on tilastollista tiivistämistä, sillä keskiarvo sisältää samaa informaatiota kuin alkuperäiset mittaustuloksetkin, mutta osa tiedoista on jätetty yksinkertaistamisen takia pois.

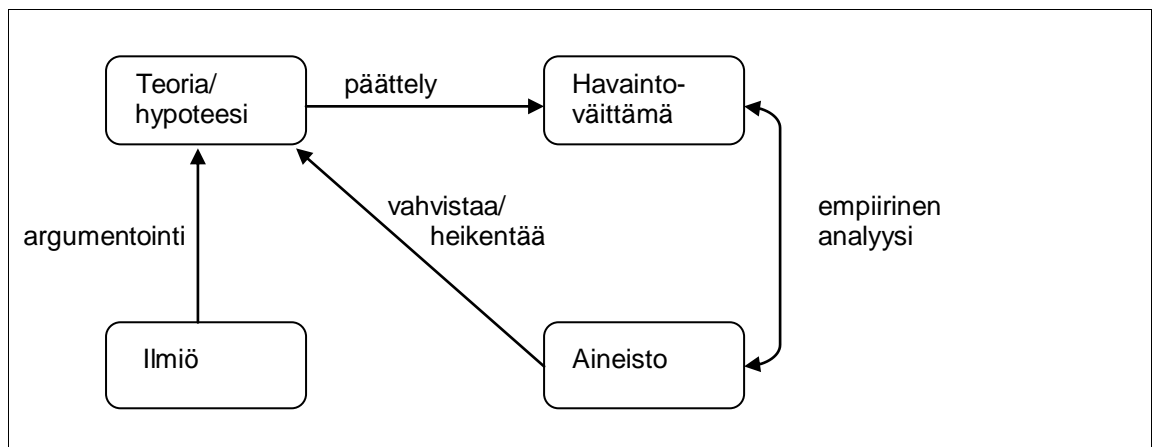
Väestöntutkimuksessa käytetyt tunnusluvut jaetaan yleisiin, ikäryhmittäisiin ja ikävakiotuihin lukuihin. Yleiset tunnusluvut lasketaan suhteuttamalla väestötapahtumien määrä keskiväkilukuun. Ikäryhmittäisiä tunnuslukuja saadaan suhteuttamalla tietynikäisten henkilöiden väestötapahtumien määrät vastaavan ikäisen väestön keskiväkilukuun. Ikävakioidulla puolestaan tarkoitetaan sitä, että väestöryhmien ikärakenteen vaikutus mittalukuun pyritään poistamaan. Ikävakioidusta tunnusluvusta voidaan mainita esimerkkinä kokonaishedelmällisyysluku, joka ilmoittaa, kuinka monta lasta koko hedelmällisen iän eläneet naiset ovat keskimäärin synnyttäneet. (Koskinen, Martelin, Notkola, Notkola & Pitkänen 2007, 21.)

3.2 Tilastollinen päättely

Väestöntutkimuksessa käytettävien haastattelu- tai kyselytutkimusaineistojen osalta on erittäin harvoin mahdollista tutkia kaikkia kiinnostuksen kohteena olevia yksilöitä tai ilmiöitä. Useimmiten tutkija tarkastelee vain tiettyä otosta tutkimuksensa kohteesta ja tekee havaintonsa sen perusteella. Tilastolliset menetelmät auttavat tutkijaa päättämään, kuinka todennäköisesti otoksen eri ominaisuudet toteutuisivat koko tutkimus-

kohteessa. Tilastollinen päättely säästää siten myös aikaa ja työtä, kun ei tarvitse mitata kaikkia mahdollisia tutkimuksen piiriin kuuluvia kohteita. (Nummenmaa 2009, 18.)

Tilastollista päättelyä voidaan käyttää myös hypoteeseja testaavan tutkimuksen osana. Mikko Ketokivi esittää havainnollisesti hypoteesin menetelmän loogisen rekonstruktion kuvana (kuva 3). Ennen päättelyä tutkimuksen kohteena olevasta ilmiöstä argumentoidaan tieteellinen teoria tai hypoteesi. Teoriasta pääteltyä havaintoväittämää tarkastellaan puolestaan empiirisesti eli tutkimuskohdetta havainnoidaan ja mitataan. Lopulta todetaan, vahvistaako vai heikentääkö empiirinen analyysi ja aineisto teoriaa. (Ketokivi 2009, 38–39.)



KUVA 3. Hypoteesin menetelmän looginen rekonstruktio (Ketokivi 2009, 39)

Ketokiven mukaan hypoteesin menetelmä vaatii käyttäjältään paljon, koska tutkijan täytyy pohtia mahdollisia teorioita jo ennen aineiston käsittelyä. Hypoteesin menetelmää käyttävän ei siis pidä analysoida aineistoaan ennen kuin hän on laatinut teorian, millaisia tilastollisia piirteitä aineistosta odotetaan löytyvän. Jos tutkija pystyy sitten myöhemmin osoittamaan aineistonsa tukevan tätä teoriaa, hän on mitä ilmeisimmin luonut uutta tietoa. Hypoteesin menetelmää pidetäänkin monilla tieteenaloilla tärkeänä tiedettä edistävänä menetelmänä. (Ketokivi 2009, 40.)

Hypoteesin menetelmästä on kehitetty myös ns. vaihtoehdoisen hypoteesin menetelmä, jossa ennen testausta pyritään löytämään kaikki ne vaihtoehdoiset teoriat, joita havaintoaineiston avulla testataan ja analysoidaan (Kanji 2006, 2).

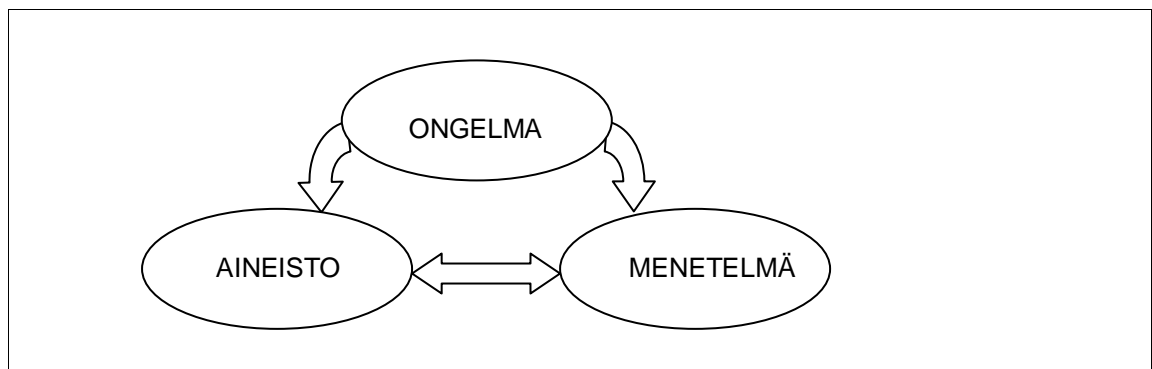
3.3 Tilastollinen mallintaminen

Tilastomenetelmien avulla pyritään myös monesti luomaan havainnoista matemaattinen malli, jolla ilmiötä pyritään selittämään ja ennustamaan. Malli on yksinkertaistettu esitys siitä, miten asiat ja ilmiöt liittyvät toisiinsa (Nummenmaa 2009, 19).

Väestötieteessä tilastollista mallintamista toteutetaan väestöilmiöiden mittaamiseen tarkoitettujen tunnuslukujen avulla. Väestötutkimuksissa tunnuslukujen tarkoituksellisuus on tärkeää. Esimerkiksi jos tarkastellaan syntymien yleisyyttä Suomessa jonkin ajanjakson aikana, voidaan laskea ajanjakson aikana syntyneiden lukumäärä. Mutta tulos ei kerro kuitenkaan, ovatko syntymät Suomessa yleisiä vai harvinaisia. Yleisyyden arvioimiseksi syntyneiden määrä pitäisi suhteuttaa siihen väestöön, jossa syntymät tapahtuvat. (Koskinen ym. 2007, 20.)

3.4 Tilastollinen tutkimus

Tilastollisen tutkimuksen lähtökohdana on ongelma, jonka ratkaisemiseksi tarvitaan tutkimusaineisto ja -menetelmä. Leila Karjalainen (2010, 10) jakaa tutkimuksen lähtökohdat siten kolmeen elementtiin kuvan 4 mukaisesti.



KUVA 4. Tilastollisen tutkimuksen lähtökohdat (Karjalainen 2010, 10)

Käytettävä aineisto vaikuttaa ratkaisevasti tutkimustulosten luotettavuuteen, minkä takia aineiston hankinta on suoritettava suunnitelmallisesti. Käytettävät laskentamenetelmät ja -mallit on puolestaan osattava valita oikein. Tietokone ei tee mekaanisia laskuvirheitä, joten mahdollisten virheellisten tulosten syynä on joko aineiston keruussa tapahtunut virhe tai väärin valittu menetelmä. (Karjalainen 2010, 14.)

3.4.1 Tutkimusaineisto

Aineistoa voidaan kerätä kyselylomakkeilla, mittaamalla, havainnoimalla tai käyttämällä valmiita tilastoja, esim. väestötilastoja. Tutkimusaineiston kuvaaminen käsittää aineiston tarkastelun, luokittelun ja graafisen esittämisen. Sama lähdemateriaali voidaan luokitella tai esittää graafisesti monellakin eri tavalla. Aineistoa voidaan myös analysoida monella eri menetelmällä. Analyysin valintaan vaikuttaa ratkaisevasti se, millainen tutkimus on kyseessä. Tilastomenetelmiä voidaan kuitenkin käyttää sekä kvalitatiiviseen että kvantitatiiviseen tutkimukseen. (Holopainen & Pulkkinen 2008, 19.)

Väestötieteellisen tutkimuksen aineistona käytetään yleensä väestötilastoja, joiden tiedot kattavat koko väestön ja kaikki tietynä ajanjaksona tapahtuneet väestönmuutokset. Otantatutkimukset ovat nimensä mukaisesti vain osan väestöä kattavia kyselyjä. Väestötieteessä otantatutkimuksia käytetään kehitysmaiden väestöjen tutkimukseen, koska menetelmä ei ole taloudellisesti ja työvoimamäärältään läheskään yhtä vaativa kuin koko väestön laskenta ilman luotettavia väestörekisterijärjestelmiä. (Pitkänen & Sihvonen 2003.)

3.4.2 Tutkimusaineiston ajallinen tarkastelu

Väestötieteessä on tärkeää ymmärtää perioditarkastelun ja kohorttitarkastelun ero. Periodilla tarkoitetaan useimmiten vuoden tai muutaman vuoden pituista poikkileikkausajankohtaa. Periaatteessa periodi voi olla kuinka lyhyt ajanjakso tahansa, esimerkiksi kuukausi tai päiväkin. Kohortti puolestaan tarkoittaa jonakin tietynä ajankohtana jonkin samanlaisen väestötapahtuman kokeneiden henkilöiden muodostamaa joukkoa, jota tutkitaan myöhemmin uudestaan. Esimerkiksi tietynä vuonna syntyneet muodostavat syntymäkohortin ja tietynä vuonna avioituneet avioliittokohortin. Vastaavasti voidaan myös puhua periodi- tai kohorttianalyysistä. Periodianalyysissä mitalluvut lasketaan jollekin lyhyelle poikkileikkausjaksolle eli periodille. Näin laskettuja lukuja kutsutaan periodikohtaisiksi tunnusluvuiksi. Mittalukuja kohorttikohtaisesti laskettaessa tutkitaan jotakin tiettyä kohorttia ja seurataan, mitä tälle kyseiselle kohortille ajan kuluessa tapahtuu. Tällä tavoin laskettuja lukuja kutsutaan kohorttikohtaisiksi tunnusluvuiksi. (Pitkänen & Sihvonen 2003.)

Historiallisten tutkimusten mielenkiinnon kohteina on tutkia, miten muuttujan arvot muuttuvat ajan mukaan. Tällöin kysymyksessä on ns. aikasarjat (Karjalainen 2010, 145). Aikasarjan havaintoarvoihin vaikuttavat monet eri tekijät. Tavallisimpia näistä seikoista ovat trendi, suhdannevaihtelu, kausivaihtelu ja satunnaisvaihtelu. Trendi on pitkän aikavälin kehityssuunta, kun taas kausivaihtelu liittyy yleensä vuodenaikoihin. Suhdannevaihtelua aikasarjoihin aiheuttavat taloudelliset ja yhteiskunnalliset muutokset. Satunnaisvaihtelua eli epäsäännöllistä vaihtelua aikasarjoihin voivat tehdä esimerkiksi lakot, luonnonmullistukset ja sodat. (Karjalainen 2010, 148–149.)

Aikasarjalla tarkoitetaan joukkoa eri aikoina samasta ilmiöstä tai muuttujasta tehtyjä havaintoja. CSC – Tieteellinen laskenta Oy:n asiantuntija Hannu Karttunen täsmentää, että tilastotieteessä aikasarja tarkoittaa yleensä tasavälistä sarjaa, jossa peräkkäisten havaintojen aikaväli on aina sama (Karttunen 2001, 113).

3.4.3 Tilastolliset mitta-asteikot

Moniin analyysimenetelmiin liittyy erityisiä toimenpiteitä, joilla alkuperäisestä lähdeaineistosta saadaan varsinainen tutkimusaineisto. Kun lähdeaineisto on koottu, joudutaan sitä useimmiten siis vielä jalostamaan eli muuntamaan tiedoksi, jota on helpompi ymmärtää tai käyttää tutkimusaineistona (Wright & London 2009, 33). Lisäksi tutkitavilla muuttujilla voi olla erityyppisiä arvoja, mikä vaikuttaa siihen, mitä tilastollisia laskemia muuttujilla voidaan tehdä (Karttunen 2001, 17).

Muuttujien mitta-asteikon määrittely on tarpeen, koska suoritettavat tilastolliset toimenpiteet määräytyvät juuri mitta-asteikon mukaan. Tilastotieteen perusteet – oppikirjan mukaan (Karjalainen 2010, 20) tilastollisia mitta-asteikkoja voidaan nimetä neljä: luokitteluasteikko, järjestysasteikko, välimatka-asteikko ja suhdelukuasteikko.

Luokitteluasteikko eli laatueroasteikko on kysymyksessä silloin, kun muuttujan arvot voidaan erotella toisistaan laadullisesti ilmoittamalla vain ryhmä tai luokka, johon havainto kuuluu. Muuttujien arvoilla ei ole mitään yksiselitteistä järjestystä, vaikka luokkien nimet voidaankin korvata numerokoodilla. Esim. sukupuoli voidaan merkitä mies = 1, nainen = 2. Luokkien tai koodien järjestyksen vaihtaminen ei kuitenkaan

vaikuta aineiston käsittelyyn tai saataviin tuloksiin. Lukumäärien ja niistä laskettujen prosentiosuuksien laskeminen ovat ainoat sallitut aritmeettiset toimenpiteet.

Järjestysasteikon eli ordinaaliasteikon muuttujan arvot voidaan panna mitatun ominaisuuden mukaan järjestykseen. Esimerkiksi sanallinen arvosana on järjestysasteikon muuttuja, jonka arvoja ovat mm. tyydyttävä, hyvä ja kiitettävä. Välimatka-asteikon eli intervalliasteikon muuttuja mittaa määrää, jolloin muuttujien suuruusjärjestyksen lisäksi voidaan laskea muuttujan arvojen erotus. Suhdelukuasteikon eli suhdeasteikon muuttujille on välimatkojen laskemisen lisäksi mahdollista laskea muuttujien arvojen suhteet. Esimerkiksi kuntien väkilukuaineistosta voidaan laskea, kuinka moninkertainen Espoon väkiluku on Mikkelin väkilukuun verrattuna. (Karjalainen 2010, 21.)

TAULUKKO 2. Muuttujien mitta-asteikot (Warner 2008, 7)

Mitta-asteikko	Sallitut operaatiot	Tilastokäyttö	Arvojen käsittely
Luokitteluasteikko	=, ≠	Ei-Laskennallinen	Kvalitatiivinen
Järjestysasteikko	=, ≠, <, >	Ei-Laskennallinen	Kvantitatiivinen
Välimatka-asteikko	=, ≠, <, >, +, -	Laskennallinen	Kvantitatiivinen
Suhdelukuasteikko	=, ≠, <, >, +, -, ×, /	Laskennallinen	Kvantitatiivinen

Rebecca Warner erotelee (taulukko 2) mitta-asteikot sallittujen operaatioiden lisäksi vielä tilastokäytön ja arvojen käsittelyn mukaan (Warner 2008, 7). Mitta-asteikon määrittämisen yhteydessä tulisi pohtia, voidaanko mitta-asteikon muuttujia käyttää laskennallisiin toimenpiteisiin ja käsitelläkö arvoja laadullisesti vai määrällisesti.

Välimatka-asteikon ja suhdelukuasteikon kohdalla ei kuitenkaan ole välttämätöntä erotella, kumpi asteikko on kyseessä, sillä kyseisten asteikkojen tunnistaminen voi olla vaikeaa. Näiden asteikkojen toisistaan erottamisessa voi tosin hyödyntää klassista esimerkkiä lämpötilasta, joka on välimatka-asteikon, muttei suhdelukuasteikon muuttuja (Karjalainen 2010, 21). Lämpötilaerot voidaan siis laskea, mutta esimerkiksi lämpötilojen +3 C ja -5 C suhde ei ole mielekäs. Leila Karjalaisen mukaan useimmiten riittää jaottelu vain kolmeen mitta-asteikkoon, jotka ilmaisevat siis joko muuttujien laatua, järjestystä tai määrää.

3.4.4 Tilastolliset tunnusluvut

Tutkimusaineiston analysoinnissa käytetään apuna tilastollisia tunnuslukuja, jotka kuvaavat aineiston yksittäisiä piirteitä. Vaikka tunnuslukuja käytettäessä osa informaatiosta häviää, niin toisaalta laajojenkin aineistojen tieto saadaan tiiviiseen muotoon (Heikkilä 2010, 82). Mitta-asteikon määrittämisen jälkeen huolehditaan siitä, että aineistosta tehdään vain kyseiselle muuttujatyypille sallittuja toimenpiteitä (Karjalainen 2010, 21).

TAULUKKO 3. Eri mitta-asteikoille soveltuvat tunnusluvut (Heikkilä 2010, 90)

Mitta-asteikko	Sijaintiluvut	Hajontaluvut	Muut tunnusluvut
Luokitteluasteikko	Moodi		
Järjestysasteikko	Moodi Mediaani Fraktiilit	Vaihteluväli Kvartiiliväli	
Välimatka-asteikko	Moodi Mediaani Keskiarvo Fraktiilit	Vaihteluväli Vaihteluvälin pituus Kvartiiliväli Kvartiilivälin pituus Keskipoikkeama Keskihajonta Varianssi	Vinous Huipukkuus
Suhdelukuasteikko	Moodi Mediaani Keskiarvo Fraktiilit	Vaihteluväli Vaihteluvälin pituus Kvartiiliväli Kvartiilivälin pituus Keskipoikkeama Keskihajonta Varianssi Variaatiokerroin	Vinous Huipukkuus

Tilastolliset tunnusluvut voidaan jakaa taulukon 3 mukaisesti kolmeen ryhmään: sijaintiluvut, hajontaluvut ja muut tunnusluvut (Heikkilä 2010, 90). Sijaintiluvut ilmaisevat havaintoarvojen sijaintia tai suuruusluokkaa. Sijaintilukuja ovat moodi, mediaani, keskiarvo sekä fraktiilit eli prosenttipisteet (Karjalainen 2010, 87). Sijaintiluvut kuvaavat tilastoaineiston keskimääräistä sijaintia, minkä perusteella lukuja kutsutaan tilastotieteessä myös keskiluvuiksi (Byckling & Salonen 2011). Hajontaluvuilla puolestaan kuvataan havaintojen keskinäistä sijaintia ja niiden jakautumista. Hajontaluvuista käytetyimmät ovat vaihteluväli, kvartiiliväli, keskihajonta ja variaatiokerroin (Karjalainen 2010, 95). Muita tunnettuja tunnuslukuja (Karjalainen 2010, 101) ovat vinous, joka ilmaisee havaintojen jakautuminen keskiarvon eripuolille, ja huipukkuus, joka vertaa jakauman muotoa normaalijakaumaan.

3.5 Tilastojen käytäntesäännöt

Suomen lainsäädännössä henkilötietojen käsittelyn määräyksiä historiallista tai tieteellistä tutkimusta varten tarkennetaan Henkilötietolain 14. pykälässä (Henkilötietolaki 1999). Lain mukaan tutkimustarkoituksessa voi käsitellä henkilörekisteritietoja, jos se on tutkimukselle välttämätöntä ja henkilörekisterin käyttö perustuu asianmukaiseen tutkimussuunnitelmaan. Tietojen käsittelyssä täytyy luonnollisesti noudattaa yksityisyyden suojaa sekä tietojen käytössä ja luovuttamisessa että henkilörekisterin hävittämisessä tai arkistoinnissa. Henkilötietojen käyttöä tilastotarkoituksia varten säätelee puolestaan Henkilötietolain 15. pykälä. Henkilötietojen käsittely on sallittua, jos tilaston tarkoituksena olevaa tiedontarvetta ei voida toteuttaa ilman kyseisten tietojen käsittelyä ja jos tilaston tuottaminen kuuluu rekisterinpitäjän toimialaan sekä jos tilastorekisteriä käytetään vain tilastollisiin tarkoituksiin tietosuojamääräyksiä noudattaen. (Henkilötietolaki 1999.)

Yhdistyneiden Kansakuntien Euroopan talouskomission tilastokonferenssin julkaisussa ”Tilastotietojen luottamuksellisuus & yksikötason tietojen käyttö - Hyviä käytäntöjä koskevat periaatteet ja suuntaviivat” (YK 2007) mainitaan neljä johtavaa periaatetta tilastollisen tutkimusaineiston käytössä. Niistä ensimmäinen liittyy tietojen luottamuksellisuuteen: virallisiin tilastotarkoituksiin kerättyjen yksikötason tietojen luottamuksellisuus on turvattava. Toinen periaate sallii pääsyn yksikötason tietoihin ainoastaan silloin, kun tietoja käytetään tilastotarkoituksiin. Kolmas periaate vaatii yksikötason tietojen käyttöoikeuksilta yhtenäistä linjaa lainsäädännön ja muiden säännösten kanssa. Neljäs periaate edellyttää tutkijoiden käyttö lupien myöntämisessä ja yksikötason tietojen käyttötarkoituksista avointa ja julkista tiedottamista.

EU:n tilastoviranomaisten laatimissa Euroopan tilastojen käytäntesäännöissä mainitaan ne periaatteet, joiden noudattamiseen Euroopan tilastovirasto Eurostat ja kansalliset tilastoviranomaiset ovat sitoutuneet. Periaatteet kattavat tilastojärjestelmän toimintaedellytykset, tilastolliset prosessit ja tuotteiden laadun (Tilastokeskus 2008). Tilastokeskuksen julkaisemat Eurostatin periaatteet eli Euroopan tilastojen käytäntesäännöt voidaan esittää myös taulukkona (liite 1).

Liitteessä 1 esitetyt tilastojen käytäntesäännöt perustuvat yhteensä 15 periaatteeseen. Toimintaedellytyksiin liittyviä periaatteita on lueteltu 6: ammatillinen riippumattomuus, tiedonkeruuvaltuudet, resurssien riittävyys, laatuun sitoutuminen, tilastosalaisuus sekä puolueettomuus ja objektiivisuus. Tilastollisiin prosesseihin liittyviä periaatteita mainitaan puolestaan 4: pätevät menetelmät, asianmukaiset tilastolliset menettelyt, kohtuullinen vastaustaakka ja kustannustehokkuus. Tuotteiden laatuun liittyviä periaatteita esitellään 5: relevanssi, tarkkuus ja luotettavuus, ajantasaisuus ja oikea-aikaisuus, yhtenäisyys ja vertailukelpoisuus sekä saatavuus ja selkeys.

Yhteenvedona käytäntesäännöistä voisi kiteyttää, että EU:n tilastollisten periaatteiden mukaan tilastojen tarkoitus on kuvata luotettavasti sitä todellisuutta, jota ne on suunniteltu edustamaan. Myös Grönroos esittää tilastotieteen oppikirjassaan EU:n tilastoperiaatteista ensimmäisenä juuri luotettavuuden periaatteen, minkä takia lähteet ja menetelmät olisi valittava käyttäen tieteellisiä metodeja ja havainnot mahdollisista virheistä olisi ilmoitettava viivytyksettä. Siten luotettavuus liittyy lähteiden ja menetelmien valinnan lisäksi myös tilastojen tuottamiseen ja julkaisemiseen. Viranomaisten tilastot olisi tuotettava puolueettomasti ja julkaistava niin nopeasti kuin se luotettavuuden kannalta on mahdollista. Tilastot olisi laadittava siten, että ne palvelisivat käyttäjien tarpeita mahdollisimman hyvin huomioiden samalla kuitenkin viranomaisten määrittelemät vaatimukset, jotka perustuvat asetettuihin tavoitteisiin. Vaatimukseen sisältyy mm. käsitteiden ja luokitusten käyttö, minkä tulisi olla yhdenmukaista, jotta eri maiden tilastot olisivat vertailukelpoisia. (Grönroos 2004, 1–2.)

Virallisiin tilastoihin liittyvien vaatimusten toteuttamista ja tuloksia olisi EU:n tilastoviranomaisten mukaan verrattava tilastojen tuottamisen työmäärään ja kustannuksiin. Lisäksi tilastoaineiston osalta saatujen tietojen käyttö olisi estettävä muihin kuin tilastollisiin tarkoituksiin sekä tiedot olisi suojattava laittomalta paljastumiselta. Asianomaisille olisi kuitenkin myönnettävä oikeus saada tietoja käytetyistä suojausmenetelmistä ja tarkoituksista, joita varten tiedot on kerätty. (Grönroos 2004, 2.)

Käytäntesäännöt koskevat EU:n neuvoston päätöksen No 322/97 mukaan Euroopan tilastojärjestelmään kuuluvia tilastoja, joita EU ja kansalliset tilastoviranomaiset tuottavat. Kansallisiin tilastoviranomaisiin luetaan Suomessa Tilastokeskuksen lisäksi myös muita tilastotuottajia, joiden tilastotietoja käytetään eurooppalaisten tilastojen kokoamisessa. (Tilastokeskus 2008.)

4 KARJALA-TIETOKANTA

Karjala-tietokantahankkeen taustalla on Karjala-tietokantasäätiö, jonka tavoitteena on tutkimusrekisterin muodostaminen luovutetun Karjalan kirkonkirja-aineistosta sekä aineiston tarjoaminen tutkimuksen käyttöön. Tavoitteeseensa säätiö pyrkii edistämällä luovutetun Karjalan väestöä koskevan väestötietomateriaalin saamista tutkimusta paremmin palvelemaan käyttöön, helpottamalla Karjalan väestöä koskevan tutkimuksen tekoa sekä tukemalla arkistohallinnon toimintaa. (Karjala-tietokantasäätiö 2011.)

Tietokantahanke aloitettiin Mikkelissä 1980-luvulla maakunta-arkiston johtaja Raimo Viikin toimesta. Hänen aloitteestaan Mikkelin kaupunki asetti vuonna 1985 toimikunnan, joka selvitti Ruotsissa 1970-luvulla aloitetun Demografiska Databasen -hankkeen mukaisen tietokannan toteuttamismahdollisuudet Mikkelissä. Uumajan yliopiston yhteydessä toimiva tietokantayksikkö (DDB 2011) on tarjonnut tallentamaansa kirkonkirja-aineistoa tutkimuskäyttöön jo usean vuosikymmenen ajan.

Vuonna 1988 Karjala-tietokantahankkeeseen palkattiin työntekijöiksi Pentti Hiltunen ja Paula Siitonen, jotka työstivät ensimmäiset tallennusohjeet ja -ohjelmat kirkonkirjatietojen atk-tallennusta varten. Tallennusmenetelmiä testattiin seitsemän tallentajan toimesta vuonna 1989 ja hyvien tulosten myötä tietokantatoimikunta ehdotti säätiön perustamista, jotta hankkeen resurssi- yms. edellytykset olisi turvattu jatkossa. Vuonna 1990 Karjala-tietokantasäätiön perustivat Kuopion, Joensuun ja Jyväskylän yliopistot, Karjalan Liitto, Kirkkohallitus, Mikkelin kaupunki, Mikkelin maakunta-arkisto, Suomen Sukututkimusseura, Väestörekisterikeskus sekä yksitoista karjalaisyhteisöä. Säätiön perustamisen yhteydessä esillä olleen tietokannan suunnitteluvaiheen loppuraportin tiivistelmässä mainittiin, että tavoitteena on luoda atk-pohjainen väestötieteellinen tietokanta, jota erityyppinen tutkimus voi helposti ja monipuolisesti käyttää atk:n avulla hyväksi (Karjala-tietokanta 1990).

Tietojen tallennuksen edistämiseksi vuoden 1998 lopulla perustettiin erillinen etätyöprojekti Etelä-Savon TE-keskuksen projektituen avulla (Ropponen 2008). Etätyöprojektin yli 10 toimintavuoden aikana on tehty noin 400 palkkatukisopimusta ja kirkonkirjatietoja on sen ansiosta tallennettu vuosittain 500 000 – 600 000 henkilötietuetta.

Tallentajat työskentelevät eri puolilla Etelä-Savoa kotonaan tai lähellä kotiaan säätien käyttöön varatuissa työtiloissa. Vuosittain tallennustyöhön perehdytetään yli 30 henkilöä.

4.1 Tietokannan lähdeaineisto

Tietokantaan tallennetaan henkilötietoja luovutetun alueen luterilaisten ja ortodoksisien seurakuntien kirjoista, joita ovat syntyneiden, vihittyjen, kuolleiden ja muuttaneiden luettelot sekä rippikirjat/pääkirjat ja lastenkirjat. Vanhin kirkonkirjamateriaali on 1600-luvun lopulta ja nuorin aineisto vuodelta 1949. Karjala-tietokannan koko on 2000-luvulla kasvanut yli 9 miljoonaan henkilömerkintään, jotka sisältävät noin 1,2 miljoonan eri henkilön tiedot, sillä sama henkilö esiintyy kirkonkirjoissa yleensä monta kertaa. (Karjala-tietokantasäätiö 2011.)

Monissa Euroopan maissa oli jo keskiajalla käytössä erilaisia kirkollisia luetteloita muun muassa kastetuista ja haudatuista. Tällaiset luettelot tulivat pakollisiksi jo 1500-luvun alkupuolella Englannissa ja katolinen kirkko antoi Tridentin kirkolliskokouksessa vuonna 1563 määräyksen luetteloiden pitämisestä (Nieminen 1999, 6). Katsauksessaan väestötilaston historiaan Mauri Nieminen mainitsee, että Suomen väestökirjanpidon juuret ulottuvat 1500-luvun alkupuolelle saakka, jolloin ensimmäiset luettelot väestöstä laadittiin Kustaa Vaasan toimesta. Henkikirjoituksen tarkoituksena oli selvittää valtakunnasta kerättävien verojen määrä. Vuonna 1540 määrättiin, että maakirjoihin oli luetteloitava talonpojat ja heille maksuun laitettut verot. Maakirjat ja muut satunnaisesti verotusta varten kerätyt tiedot ovat ainoa tietolähde 1500-luvun suomalaisesta väestöstä. 1600-luvulla alettiin pitää luetteloita asekuntoisista miehistä. Varsinainen henkikirjoitus alkoi 1600-luvun alkupuolella. Vuonna 1634 alettiin kerätä ns. henkirahaa verona. Henkikirjoihin kirjattiin aluksi vain 12 vuotta täyttäneet henkilöt. Henkikirjoitus säilyi eri muodossaan aina vuoteen 1989 saakka, jolloin henkikirjojen vuosittaisesta tarkistamisesta luovuttiin. (Nieminen 1999, 7–8.)

Varsinaisen pohjan väestötilastojen laadinnalle loivat kuitenkin seurakuntien toimesta laadittavat väestöluettelot. Ruotsin seurakuntien esimerkin mukaisesti Turun piispa Rothovius määräsi vuonna 1628 suomalaisia kirkkoherroja pitämään kirjaa vihityistä, kastetuista ja haudatuista. Varsinaisena väestörekisterinä oleva seurakunnan pääkirja kehittyi kuitenkin ns. rippikirjasta, jonne seurakuntien pappien piti merkitä seurakun-

talaisten kristinopin tuntemus ja lukutaito. Vuoden 1686 kirkkolaisissa velvoitettiin pitämään kaikissa seurakunnissa perhekuntakohtaista luetteloa. Tämä laki loi osaltaan pohjan Suomen nykyiselle jatkuvalle väestörekisterinpidolle. (Nieminen 1999, 8.)

Karjala-tietokantaan tallennetaan lähes kaikki luovutetun Karjalan kirkonkirjojen sisältämät tiedot. Lähdetietokannassa kunkin seurakunnan tiedot ovat omassa hakemistossaan.

TAULUKKO 4. Karjala-tietokannan tietorakenne Antrean seurakunnan osalta

Hakemiston sisältö	Tiedosto	Selite
SYKA-tiedostot	SY013 SK013 SH013	syntyneiden henkilötiedot syntyneiden kummit syntyneiden huomautustiedot
MULU-tiedostot	MU013 MH013	muuttaneiden henkilötiedot muuttaneiden huomautustiedot
VILU-tiedostot	VI013 VT013 VH013	vihittyjen henkilötiedot vihittyjen todistajat vihittyjen huomautustiedot
KUHA-tiedostot	KU013 KO013 KH013	kuolleiden henkilötiedot kuolleiden omaisten tiedot kuolleiden huomautustiedot
LAKI-tiedostot	LA013 LV013 LM013 LK013 LT013 LH013 LL013	lasten henkilötiedot vanhempien tiedot kinkerikerrat kinkerivuodet taitomerkinnot huomautustiedot vanhempien huomautustiedot
RIKI-tiedostot	RI013 RK013 RT013 RV013 RP013 RS013 RH013	henkilötiedot kinkerivuodet taitomerkinnot avioliittomerkinnot puolisot suhteet huomautustiedot

Taulukossa 4 esitetään Antrean seurakunnan hakemisto ja sen tiedostot. Hakemistossa olevat tiedostot on nimetty siten, että ensiksi on tiedoston tietosisällön tunnus ja lopuksi seurakuntakoodin tunnus. Samalla tavoin on jaoteltu kaikkien tähän mennessä tallennettujen 70 seurakunnan tiedot hakemistoihin ja kirkonkirjatyypeittäin eri tiedostoihin. Kaiken kaikkiaan tallennettavia luovutetun alueen seurakuntia on yhteensä 82.

4.2 KATIHA-tiedonhakuohjelma

Windows-pohjainen käyttöliittymä Karjala-tietokantaan valmistui vuonna 2004 ja se avattiin yleisölle Mikkelin maakunta-arkiston 70-vuotisjuhlien yhteydessä. Sen jälkeen useat sukututkijat ovat hyödyntäneet arkiston tutkijasalissa käytössä olevaa KATIHA-ohjelmaa ja sen tietokantaa, joka sisältää henkilötietoja aina vuoteen 1950 saakka (Ropponen 2008, 22).

KATIHA-henkilöhaun verkkoversio puolestaan julkaistiin kesäkuussa 2008. Tietosuoja-säädökset rajoittavat tietoverkossa saatavilla olevan aineiston noin 5 miljoonaa henkilömerkintään, sillä 100 vuotta nuorempia tietoja ei voida julkaista verkossa. Poikkeuksena tähän tietosuojasääntöön ovat kuolleiden luetteloiden tiedot, joita verkkoversiossakin on vuoteen 1950 asti (Ropponen 2008, 23). Verkkoversio, jonka kautta tietokannasta on julkaistu 60 %, on saanut sukututkijoista innokkaan käyttäjäjoukon. Sovelluksen verkkosivuilla kävijöiden lukumäärä on noin 600 kävijää päivittäin.

Sekä Mikkelin maakunta-arkiston tutkijasalissa oleva KATIHA-tiedonhakuohjelma että sen verkkoversio on laadittu lähinnä sukututkijoita varten (Keckman-Koivuniemi 2010). Hakuja on mahdollista tehdä seuraaviin kenttiin: asuinpaikka, ammatti, etunimet, patronyymini, sukunimi, sukupuoli, syntymävuosi, vihkivuosi, kuolinvuosi ja muuttovuosi (tulo- tai lähtöajankohta).

Verkko-osoitteen <http://www.karjalatk.fi/katiha> kautta käytettävään KATIHA-ohjelmaan on sen kolmen käyttövuoden aikana kohdistunut monia kehittämistoiveita. Heti palvelun käyttöönoton jälkeen tuli käyttäjiltä kritiikkiä palvelun hitaudesta ja ohjelman käytön monimutkaisuudesta. Palvelimen vaihdon jälkeen ohjelman toiminta nopeutui.

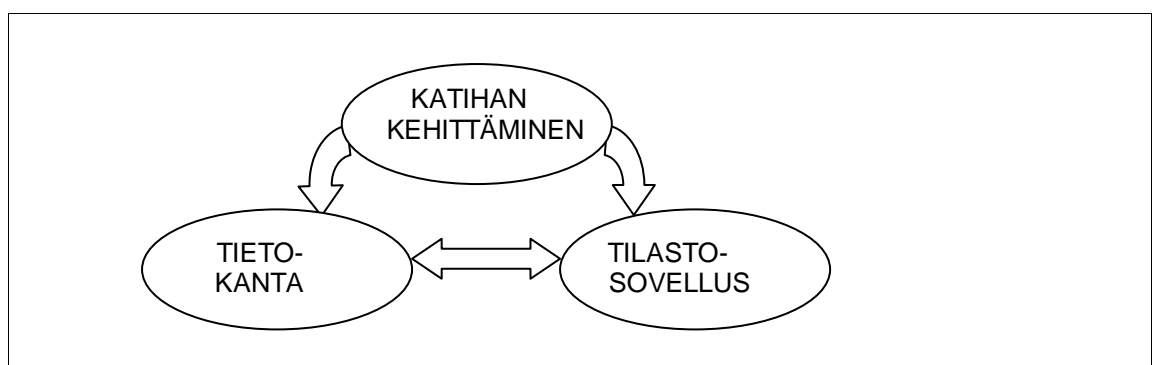
Puutteistaan huolimatta KATIHA-ohjelma helpottaa merkittävästi luovutetun alueen seurakuntien kirkonkirjojen käyttöä. Sukututkimuksen lisäksi tutkimusrekisteriä voitaisiin hyödyntää monitieteisissä tutkimuksissa, mutta se edellyttäisi tietokannan käyttöliittymän kehittämistä tutkimusta palvelemaan suuntaan. Tietokannan tietoja voitaisiin hyödyntää myös Kansallisarkiston vuoden 2009 digitointihankkeessa (Kescskeméti 2009) digitoitujen kirkonkirjasivujen indeksointiin, joka osaltaan parantaisi luovutetun alueen seurakuntien digitoidun aineiston käytettävyyttä.

5 KATIHA-TILASTOSOVELLUS

Suurien tietomäärien käsittelyssä ja analysoinnissa on valtavia haasteita. Tutkimusaineiston keruu ja sen luotettava arkistointi vaativat runsaasti työtä, joten tietotulvassa liikkuva tutkija tarvitsee laadukkaita ja tehokkaita työkaluja hallitakseen aineistojaan koko elinkaaren (CSC 2011). Laadukas tiedonhallinta edesauttaa tutkimusdatan käsittelyä ja aineiston tarjoamista mahdollisesti muihinkin tutkimuksiin.

Tutkimusdatan avoimuutta ja saatavuutta on helpottanut tutkimusympäristöjen muuttuminen sähköisiksi (Borg & Kuula 2007, 79). Yhteiskunnallisen tietoarkiston julkaisemassa OECD-datasuosituksen selvityksessä painotetaan, että kyse on perustavaa laatua olevasta uudelleenrakentumisesta, joka on jo muuttanut keskeisesti tieteellisen tutkimuksen käytäntöjä. Datasuosituksessa tavoiteltu avoimuuden periaate on toteutettavissa pitkälti verkkopalvelujen avulla. Siksi myös tutkimusten aineistotietojen pitää suosituksen mukaan olla yhä helpommin ja täsmällisemmin saatavilla internetissä.

Karjala-tietokannan sisältämää kirkonkirja-aineistoa on verkossa, mutta käyttöliittymää ei ole rakennettu tilastotietojen hakua varten vaan sukututkimusta palvelevaa henkilöhakua varten. Jyväskylän yliopiston historian ja etnologian laitoksen edustajien kanssa käydyissä keskusteluissa on herännyt tarve kehittää Karjala-tietokannan KATIHA-ohjelmaa tieteellisen tutkimuksen tiedontarpeiden näkökulmasta.



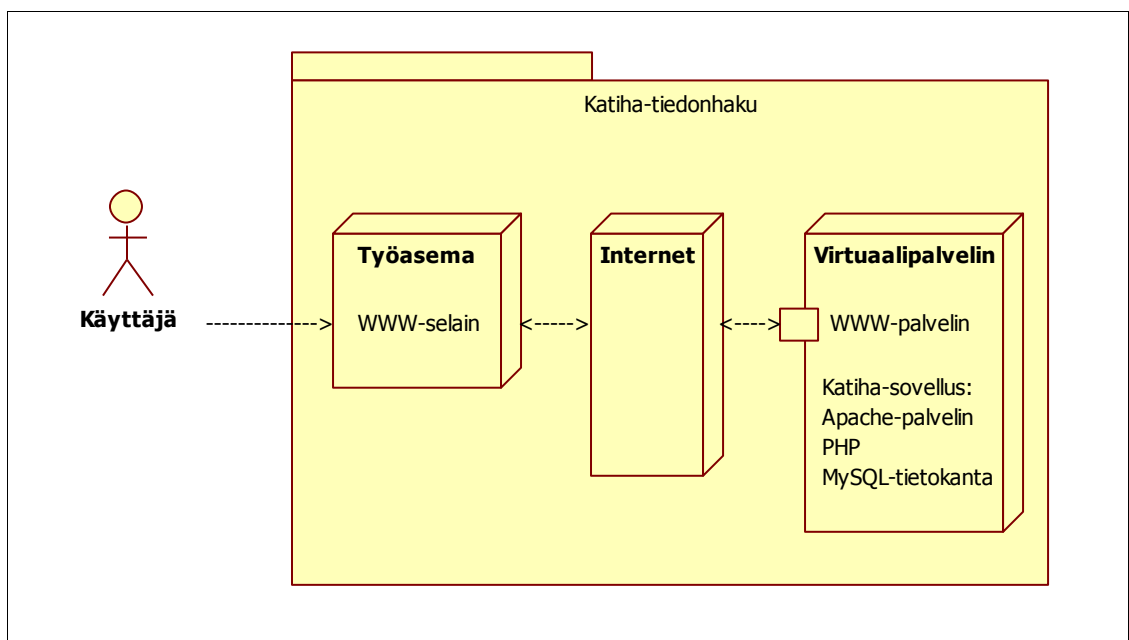
KUVA 5. KATIHA:n tilastokäytön kehittämisen lähtökohdat (vert. kuva 4)

Luvussa 3 (kuva 4 s. 16) esitetyt tutkimuksen lähtökohdat voidaan ohjelman kehittämisen suhteen esittää kuvan 5 mukaan seuraavasti: 1. Ongelma = KATIHA-ohjelman kehittäminen, 2. Aineisto = Karjala-tietokanta ja 3. Menetelmä = Tilastosovellus.

Karjala-tietokantaan on jo tallennettu 9 000 000 henkilömerkintää, mutta suurin osa eli yli 5 400 000 merkintää on poimittu rippi- ja lastenkirjojen sivuilta. Tilastosovelluksen tietokantaan tulisi tallennetusta aineistosta mukaan runsas kolmannes eli noin 3 660 000 henkilömerkintää, jotka on koottu syntyneiden, kuolleiden, vihittyjen ja muuttaneiden luetteloista. Näiden luetteloihin sisältyvistä kirkonkirjatiedoista voitaisiin laskea mm. luovutetun alueen väkilukutietoja 1680-luvulta aina vuoteen 1950 asti.

5.1 Tietojärjestelmävaatimukset

Nykyinen KATIHA-henkilöhaun verkkoversio on toiminut riittävän nopeasti verkon kautta siitä lähtien kun aineisto siirrettiin Web-hotellipalveluja tarjoavan yrityksen tehokkaalle palvelimelle elokuussa 2009. Myös KATIHA-tilastosovellus on sen valmistuttua tarkoitus julkaista saman nettihotellin palvelimella.



KUVA 6. KATIHA-tiedonhaun tietojärjestelmäarkkitehtuuri

a) Tietojärjestelmäarkkitehtuuri

Tilastosovelluksen tietojärjestelmän lähtökohtana on selainpohjaisuus ja toimivuus sekä MS Internet Explorer että Firefox-selaimilla. Järjestelmän kuormittavuudessa on huomioitava se, että samanaikaisesti on kyettävä palvelemaan jopa 500 palvelun käyttäjää siten, että eri toimintojen vasteajat jäävät alle 15 sekuntiin. Järjestelmän on lisäksi oltava helposti laajennettava ja ylläpidettävä.

Tilasto-osion ohjelmointia varten Karjala-tietokantasäätiö palkkasi vuoden 2011 alussa määräaikaiseksi ohjelmoijaksi tradenomi Teppo Hytösen, jonka vastuulle annettiin tietojärjestelmän yksityiskohtainen tekninen suunnittelu ja toteutus. Ohjelmointityön alussa päätettiin, että käyttöliittymä toteutettaisiin PHP-kielellä, jonka ohjelmakoodin avulla tietokantahaut kohdistettaisiin MySQL-mukaiseen tietokantaformaattiin. Tekniseen toteuttamiseen sisällytettiin myös Visual dBASE – tietokantaohjelmaan pohjautuvien tallennustiedostojen siirtäminen MySQL-muotoon. Käyttöliittymä päätettiin laatia Linux-palvelimen kautta toimivaksi verkkosovellukseksi, samalla periaatteella kuin vuonna 2008 toteutettiin KATIHA-henkilöhakuohjelman verkkoversio. Ohjelmoijan kanssa sovittiin myös, että verkkopalvelinohjelmana toimisi edelleen Apache ja PHP-kielellä ohjelmoitavasta käyttöliittymästä toteutettaisiin ainakin testiversio.

b) Tietojärjestelmän toimijat

KATIHA-verkkosovelluksen toimijat voidaan jakaa ryhmiin tuottaja, asiakas, palveluntarjoaja ja operaattori sekä teknologian tarjoaja.

Tuottaja: Karjala-tietokannan sisällöntuottajina ovat tietokannan tallentajat, joita on ollut yli 400 eri henkilöä vuosina 1990–2011. Tosin tallentajat eivät itse lisää aineistoa suoraan tietokantaan, vaan säätiön toimihenkilöt siirtävät tallennetut tiedot tallentajien käyttämiltä varmistuslevykkeiltä tai muistitikuilta. Tämän perusteella Karjala-tietokantasäätiö tuottaa oman verkkopalvelunsa sisällön itse.

Asiakas: Suurin osa henkilöhakuohjelman verkkoversion käyttäjistä on sukututkijoita, mikä on ilmennyt palvelusta saaduista palautteista.

Palveluntarjoaja ja operaattori: KATIHA-sovellus sijaitsee Nordic Web Hotel Oy:n palvelimella, mutta tietoliikenneyhteydet säätiölle tarjoaa TeliaSonera.

Teknologian tarjoaja: Karjala-tietokantasäätiö tarjoaa sovelluksen teknologian, sillä verkkopalvelua ohjelmoi ja ylläpitää säätiön palkkaama määräaikainen ohjelmoija.

KATIHA-tiedonhaun ylläpito on siis kolmella eri taholla. Verkkopalvelun palvelimen toimivuudesta vastaa Nordic Web Hotel Oy ja tietoliikenteen toimivuudesta TeliaSonera. Säätiö vastaa itse MySQL-tietokannan ylläpidosta ja verkkosovelluksen ohjelman toimivuudesta.

Verkkopalvelun kehittämisen tarkoituksena on tarjota tutkijoille Karjala-tietokannasta laskettavia väestötieteellisiä tilastotietoja. KATIHA-ohjelman täydentäminen tilastoosiolla mahdollistaisi tietokannan aineiston käytön väestöntutkimuksen lisäksi myös esim. yhteiskuntatieteissä, sosiologiassa ja lääketieteen historian tutkimuksessa.

5.2 Käyttäjän vaatimukset

Karjala-tietokantasäätöön verkkopalvelujen kehittämisessä asiakkaiden vaatimukset ovat avainasemassa, sillä tietokanta on tarkoitettu käytettäväksi eikä vain säilytettäväksi. Perusvaatimus on, että tietojärjestelmän tulee toimia käyttäjän kannalta riittävän nopeasti ja luotettavasti. Eri asiakasryhmillä voi kuitenkin olla muita erilaisia odotuksia, kuten Kamenskyn (2008, 149) esimerkin mukaan laadittu asiakastuntemusanalyysi osoittaa KATIHA-sovelluksen käyttäjistä taulukossa 5.

TAULUKKO 5. KATIHA-sovelluksen asiakastuntemusanalyysi

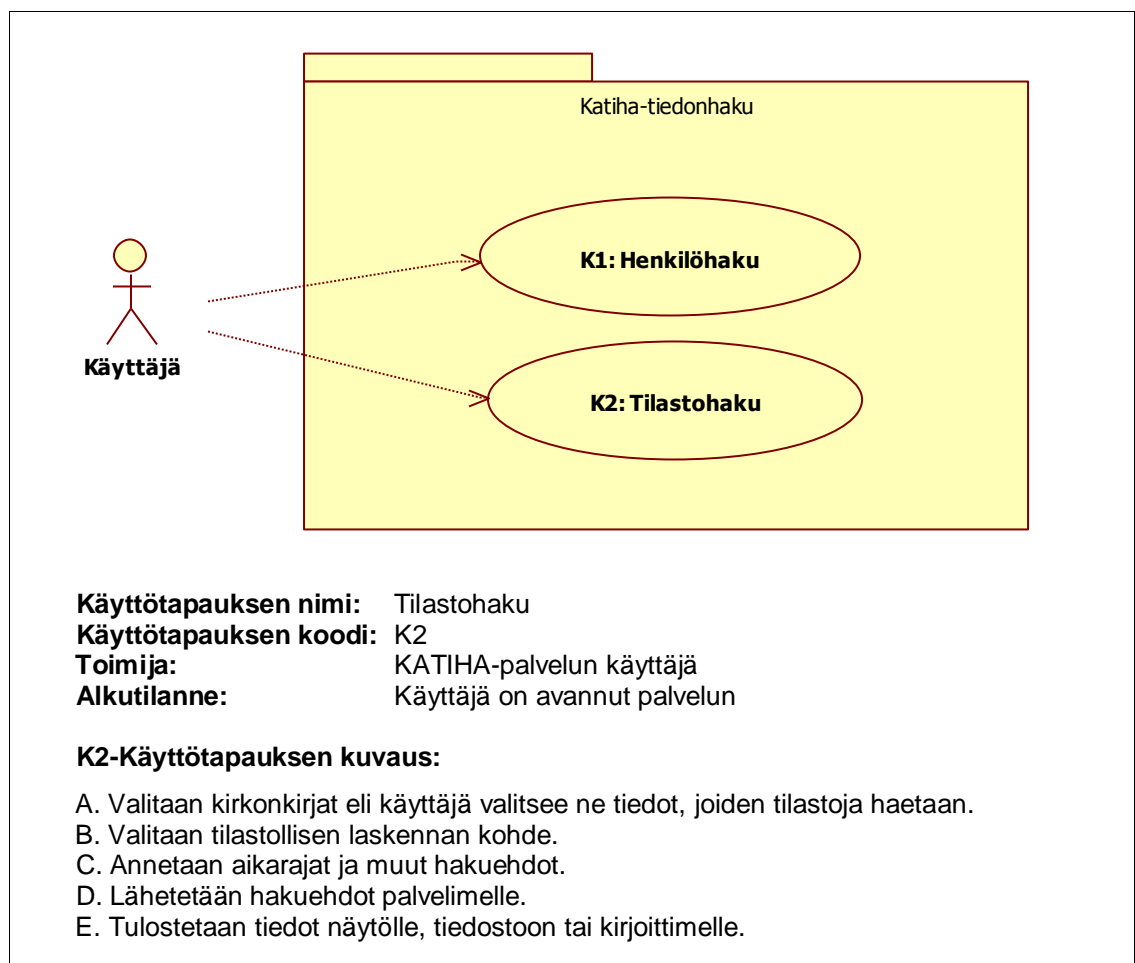
Asiakkaat	Katihan käyttö-intressit	Asiakkaan asema	Asiakasta koskevat kehitysnäkymät	Asiakkaan odotukset meiltä
Ensisijaiset käyttäjät	Sukututkimus	Harrastaja	Sukututkimuksen suosio jatkuu	Katihan päivitykset tiheämpään
Toissijaiset käyttäjät	Tiedon vertailu	Tallentaja	Tietokannan tietosisältö kasvaa	Tietojen haun nopeutuminen
Kolmannet osapuolet	Viralliset selvitykset	Virkailija	Selvityspyyntöjä edelleen	Tallennustyön valmistuminen
Potentiaaliset käyttäjät	Väestötiede, tilastotiedot	Opiskelija, tutkija	Tutkimusaineistoja siirretään verkkoon	Tilastotietojen hakumahdollisuudet

Nykyiset ensisijaiset KATIHA-ohjelman käyttäjät ovat henkilöitä, jotka tekevät sukututkimusta. Osa tarkistaa ja täydentää jo aiemmin selvittämiään sukutietoja, mutta osa on vasta-alkajia harrastuksen parissa. Toissijaisia käyttäjiä ovat Karjala-tietokantasäätöön tallentajat. He käyttävät palvelua tallennuksen apuna lähinnä henkilöiden nimien ja asuinpaikkojen yms. selvittelyä varten. Kolmansina osapuolina voidaan mainita Mikkelin maakunta-arkiston toimistovirkailijat, jotka tekevät sukuselvityksiä esim. perunkirjoituksia varten ja hyödyntävät KATIHA-henkilöhakuohjelmaa selvitystyössään. Viralliset sukuselvitykset perustuvat kuitenkin alkuperäisten kirkonkirjojen tietoihin. Potentiaalisiksi käyttäjiksi voidaan nimetä esim. väestöhistorian tutkijat, joille sovelluksen täydentämisestä tilastoosiolla olisi merkittävää hyötyä.

Väestöhistorian tutkimuksissa esim. kastettujen tai haudattujen luetteloista tilastoituja tietoja voitaisiin käyttää hyödyksi tutkittaessa demografisia muuttujia, kuten syntyneisyyttä, kuolleisuutta sekä ikä- ja sukupuolirakennetta. Vaikka kirkonkirja-aineistot on laadittu seurakunnittain, ne kertovat paitsi yhteisöistä myös niiden välisistä yhteyksistä tai niiden puutteista. Esim. vihittyjen luettelot eivät kerro yksinomaan avioliittojen yleisyydestä vaan valottavat myös menneisyyden yhteisöjen keskinäistä vuorovaikutusta. Aviopuolisoiden kotipaikkakunnista voidaan tehdä päätelmiä esimerkiksi väestön liikkuvuudesta ja kylien tai jopa tilojen välisistä yhteyksistä. (Lamberg 2008.)

5.3 Tiedonhaun käyttötapaukset

KATIHA-tiedonhaun verkkokäyttöliittymiä eli Karjala-tietokannan käyttömahdollisuuksia internetissä on tilasto-osion liittäminen jälkeen kaksi (kuva 7): 1) henkilöhaku Karjala-tietokannasta ja 2) tilastohaku eli tilastotietojen haku tutkimuksen käyttöön.



KUVA 7. KATIHA-tiedonhaun käyttöliittymät ja K2-käyttötapausten kuvaus

KATIHA-tilastosovelluksen käyttötapauksessa haetaan tilastotietoja, joiden lähdeaineistona ovat aikanaan Neuvostoliitolle luovutetun Karjalan alueen seurakuntien kirkonkirjoista syntyneiden, kuolleiden, vihittyjen ja muuttaneiden luettelot. Sovelluksen käynnistyttyä käyttäjälle annetaan mahdollisuus valita luettelotyypeistä se, minkä sisältämiä tietoja tutkitaan. Valinnan myötä tarjotaan vaihtoehdot erilaisiin laskelmiin, joiden suorittamista varten käyttäjältä pyydetään tietoa hakuehdoista. Hakuehtoihin sisältyy mm. seurakuntavalinta ja aikarajat. Hakuehtojen antamisen jälkeen ohjelma välittää tiedot palvelimelle, jossa sijaitsee tilastosovelluksen tietokanta. Lopuksi tilastolaskennan tulokset listataan käyttäjän nähtäväksi ja käsiteltäväksi.

TAULUKKO 6. Tilastollisen laskennan kohteet käyttötapauksessa K2

Luettelotyyppi	Tilastolaskelma
Syntyneet	Syntyneiden lasten lukumäärät Synnyttäneen äidin keskimääräinen ikä Lasten määrät syntymäkuukauden mukaan Syntymä- ja kastepäivän välinen keskimääräinen aika Lasten lukumäärät vanhempien ammattiryhmän mukaan
Kuolleet	Kuolleiden henkilöiden lukumäärät Alle vuoden ikäisinä kuolleiden määrä ikäkuukausien mukaan Kuolleiden määrät kuolinsyyn mukaan Kuolleiden määrät ammattiryhmän mukaan Kuolin- ja hautauspäivän välinen keskimääräinen aika
Vihityt	Solmittujen avioliittojen lukumäärät Sulhasen ja morsiamen ammattiryhmät Sulhasen ja morsiamen keskimääräinen ikä tai ikäero Sulhasen ja morsiamen keskimääräinen ikä ammattiryhmittäin Uudelleen avioituminen
Muuttaneet	Muuttaneiden lukumäärät Mistä tai mihin muutettiin Muuttaneiden ikäjakauma Muuttaneet ammattiryhmittäin Muuttaneet sukupuolittain

Taulukossa 6 luetellaan KATIHA-tilastosovellukseen ohjelmoitavat tilastolliset laskelmat. Väestöhistoriallisesta tietokannasta voisi tilastoida monia muitakin mielenkiintoisia tietoja esim. henkilöiden sukulaisuussuhde-, sisälukutaito-, koulutus- tai huomautusmerkinnöistä. Mutta tilastosovelluksen verkkoversion ensimmäiseen versioon rajataan mukaan syntyneiden, kuolleiden, vihittyjen ja muuttaneiden lukumäärähakujen lisäksi vain sosiaalhistorian tutkimusta hyödyntävä ammattien tilastointi sekä lääketiedettä sivuva kuolinsyiden tilastointi. Väestötapahtumien lukumäärien selvittely on välttämätöntä syntyvyyden, kuolleisuuden, avioituvuuden ja muuttoliikkeen tarkastelulle.

5.3.1 Syntyneiden luetteloiden tilastollinen laskenta

Syntyneiden lasten lukumäärät, syntyneiden poikien ja tyttöjen lukumäärät ja prosentiosuudet lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Lisäksi ilmoitetaan, kuinka monen syntyneen kohdalta puuttuu sukupuolietieto laskenta-ajanjaksolta. Monisyntyisten yhteislukumäärä, sekä yhteislukumäärään sisältyvien kaksos-, kolmos- ja nelossyntyisten lukumäärä lasketaan samoin käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta.

Synnyttäneen äidin keskimääräinen ikä lasketaan tietokannasta saatavan äidin ikätiedon perusteella käyttäjän antamalta ajanjaksolta aikaväleittäin. Samalla ilmoitetaan myös, kuinka suuri prosentiosuus kaikista laskenta-aikavälillä synnyttäneistä äideistä on mukana keskimääräisen iän laskennassa. Monisyntytykset käsitellään siten, että esim. kaksoset tai kolmoset synnyttänyt äiti on laskennassa vain yhteen kertaan.

Syntyneiden lasten määrät syntymäkuukauden mukaan lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Samalla ilmoitetaan myös syntyneiden kuukausittaiset määrät prosentteina koko ajanjakson syntyneiden määrään verrattuna.

Lapsen syntymä- ja kastepäivien välinen keskimääräinen aika päivinä lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin. Keskimääräisen aikaeron laskennassa ovat mukana sellaiset syntyneet, joiden kastepäivä on samana tai syntymäaikaa seuraavana vuonna, ei kuitenkaan sellaiset tietokannassa olevat lapset, joiden kasteajankohdaksi on merkitty syntymäaikaa aikaisempi ajankohta. Myös lapset, jotka on kastettu syntymäpäivänään, ovat mukana keskimääräisen aikaeron laskennassa. Myöhemmin kuin syntymää seuraavana vuonna kastettuja lapsia ei huomioida millään tavalla. Laskentaan ei myöskään oteta mukaan syntyneitä, joiden syntymä- ja kastepäivämäärät ovat muuten puutteellisia.

Lisäksi mahdollistetaan syntyneiden lasten lukumäärien listaus lasten vanhempien ammattiryhmien mukaan. Syntyneiden määrät lasketaan tällöin sekä isien että äitien ammattiryhmien mukaan aikaväleittäin ja yhteensä koko ajanjaksolta.

5.3.2 Kuolleiden luetteloiden tilastollinen laskenta

Kuolleiden henkilöiden lukumäärät lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Kuolleiden henkilöiden lukumäärät kuoliniän mukaan lasketaan samoin annetulta ajalta aikaväleittäin ja yhteensä koko ajalta. Kuolleet ryhmitellään seuraavasti: alle 1v, 1–4v, 5–14v ja yli 15v. Kuolleiden määrä ilmoitetaan ikäryhmittäin lukumäärinä ja prosentteina. Lisäksi ilmoitetaan, kuinka monen laskenta-ajanjakson aikana kuolleen kohdalta puuttuu ikämerkintä tai tieto siviilisäädystä.

Alle vuoden ikäisinä kuolleiden lasten lukumäärät ja prosenttiosuudet lasketaan ikäryhmittäin kuolinikäkuukausien mukaan käyttäjän antamalta ajanjaksolta aikaväleittäin. Lukumäärä- ja prosenttiosuustiedot lasketaan erikseen myös niiden kuolleiden lasten osalta, joiden kuolinikä tiedot puuttuvat.

Kuolleiden henkilöiden lukumäärät kuolinsyyn mukaan lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Kuolinsyy on 'ei mainittu', jos henkilön kuolinsyy ei ole tiedossa.

Kuolleiden henkilöiden lukumäärät ammattiryhmän mukaan lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Ammattiryhmä on 'ei mainittu', jos henkilölle ole merkitty ammattia. Jos ammattia ei ole luokiteltu, ilmoitetaan ammattiryhmäksi 'ei luokiteltu'.

Kuolin- ja hautauspäivien välinen keskimääräinen aika päivinä lasketaan tietokannasta saatavien henkilön kuolin- ja hautausaikatietojen perusteella käyttäjän antamalta ajanjaksolta aikaväleittäin. Keskimääräisen aikaeron laskennassa ovat mukana kuolleet, joiden hautauspäivä on samana tai kuolinaikaa seuraavana vuonna, ei kuitenkaan sellaiset tietokannassa olevat henkilöt, joiden hautausajankohdaksi on merkitty kuolinaikaa aikaisempi ajankohta. Laskentaan ei myöskään oteta mukaan kuolleita, joiden kuolin- ja hautauspäivämäärät ovat muuten puutteellisia.

5.3.3 Vihittyjen luetteloiden tilastollinen laskenta

Solmittujen avioliittojen lukumäärät lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Vihittyjen luetteloista voidaan myös selvittää aikaväleittäin puolison ammattiryhmä, kun toisen puolison ammattiryhmä on annettu. Jos henkilön ammattia ei ole merkitty tietokantaan, ammattiryhmän kohdalla on maininta ”ei mainittu”. Puolisojen ammattiryhmien laskennan yhteydessä ilmoitetaan laskentaan kuuluvien vihkimisten yhteismäärä ja aikaväli.

Sulhasen ja morsiamen keskimääräinen ikä lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin. Keskimääräisen iän kysely voidaan jakaa kahteen osaan, sulhasen ja morsiamien tietojen käsittelyyn. Sulhasen keskimääräinen ikä lasketaan kaikista ko. ajanjaksoon kuuluvien sulhasien ikätiedoista. Morsiamien keskimääräinen ikä lasketaan vastaavasti morsiamien tiedoista. Laskelmissa ovat mukana vain ne sulhaset ja morsiamet, joilla on ikätiedot. Käyttäjälle tulee ilmoitus, jos ikätiedot puuttuvat kyseessä olevalta ajanjaksolta.

Puolisojen keskimääräinen ikäero voidaan laskea kaikista ajanjaksoon kuuluvista vihkipareista laskemalla ensin vihkiparien ikäerot. Laskennassa mukana ovat vain ne vihityt parit, joilla on ikätiedot. Kysely antaa ilmoituksen, kuinka monta vihittyä paria on ollut mukana laskennassa. Käyttäjälle tulee ilmoitus, jos laskenta-ajalta ei ole käytävissä ikätietoja.

Sulhasen ja morsiamen keskimääräinen ikä ammattiryhmittäin lasketaan muutoin samalla periaatteella kuin keskimääräinen ikäkin, mutta keskiarvolaskelmat tehdään ammattiryhmittäin. Tässäkin tapauksessa laskelmissa ovat mukana vain ne sulhaset ja morsiamet, joilla on ikätiedot. Jos ikätiedot puuttuvat, tulee käyttäjälle siitä ilmoitus, kuten muissakin ikäkyselyissä.

Uudelleen avioituneitten sulhasen ja morsianten lukumäärät lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Uudelleen avioituneitten lukumäärien lisäksi ilmoitetaan myös kaikkien vihkitapausten lukumäärä koko ajanjaksolta.

5.3.4 Muuttaneiden luetteloiden tilastollinen laskenta

Muuttaneiden lukumäärät lasketaan käyttäjän antamalta ajanjaksolta huomioiden seurakuntaan muuttaneiden ja seurakunnasta pois muuttaneiden lukumäärät sekä netto-muutto aikaväleittäin ja yhteensä koko ajanjaksolta.

Muuttaneiden lukumäärät muuttoseurakunnittain lasketaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Laskennassa on mukana henkilöt, joiden muuttoseurakunta on ilmoitettu.

Muuttaneiden ikäjakauma lasketaan muuttaneiden lukumääristä eri ikäryhmittäin (alle 15-vuotiaat, 15-vuotiaat – 60-vuotiaat ja yli 60-vuotiaat) käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Laskennassa on mukana henkilöt, joiden syntymävuosi ja muuttovuosi on ilmoitettu. Lukumäärien lisäksi ilmoitetaan myös suhteelliset osuudet prosentteina. Ikäjakauma lasketaan erikseen seurakuntaan muuttaneille ja seurakunnasta muualle muuttaneille.

Muuttaneet ammattiryhmittäin lasketaan samoin muuttotyyppin mukaan erikseen seurakuntaan sisään muuttaneille tai seurakunnasta ulos muuttaneille. Laskenta suoritetaan käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Ammattiryhmä on 'ei mainittu', jos henkilöllä ei ole ammattia.

Muuttaneiden lukumäärät sukupuolittain lasketaan edellä olevien laskentaohjeiden mukaisesti käyttäjän antamalta ajanjaksolta aikaväleittäin ja yhteensä koko ajanjaksolta. Laskenta suoritetaan erikseen seurakuntaan tulleiden ja seurakunnasta lähteneiden henkilöiden osalta.

6 KATIHA-TILASTOSOVELLUKSEN TIETOKANTA

Vuosien varrella Karjala-tietokannan aineistoa on yritetty tarjota tiedeyhteisöjen käyttöön. Joensuun yliopiston historian laitoksella tietokantaa on käytetty yhden väitöskirjan ja lisensiaattitutkimuksen sekä muutaman pro gradu –tutkielman lähdeaineistona. Lisäksi tietokantaa on hyödynnetty nimistöä koskevassa väitöskirjatutkimuksessa.

Tietokannasta löytyy hedelmällistä materiaalia myös luonnontieteelle. Englantilaisen Sheffieldin yliopiston tutkija Virpi Lummaa tutkii ihmisten lisääntymisstrategioita ja eri tekijöiden vaikutuksia lapsilukuihin (Forssell 2010). Aineistoa Lummaa on koonnut suomalaisista kirkonkirjoista, joista evoluutiotutkimuksen käyttöön on kerätty jo 60 000 henkilön tiedot. Syyskuun 2010 alussa Karjala-tietokantasäätiö aloitti puolen vuoden tietojenkeruuprojektin Sheffieldin yliopiston rahoittamana. Projektin aikana Jaakkiman ja Raudun 1750-luvulla syntyneistä äideistä ja heidän jälkeläisistään koottiin yli 10 000 henkilön tiedot lähdeaineistoksi Lummaan tutkimushankkeeseen.

Viime vuosien aikana säätiön edustajat ovat keskustelleet historioitsijoiden ja yhteiskuntatieteilijöiden kanssa tietokannan hakuominaisuuksista. Keskustelujen tuloksena on todettu, että tilastotietojen tarjoaminen tietokannasta lisäisi tutkijoiden kiinnostusta tietokantaa kohtaan.

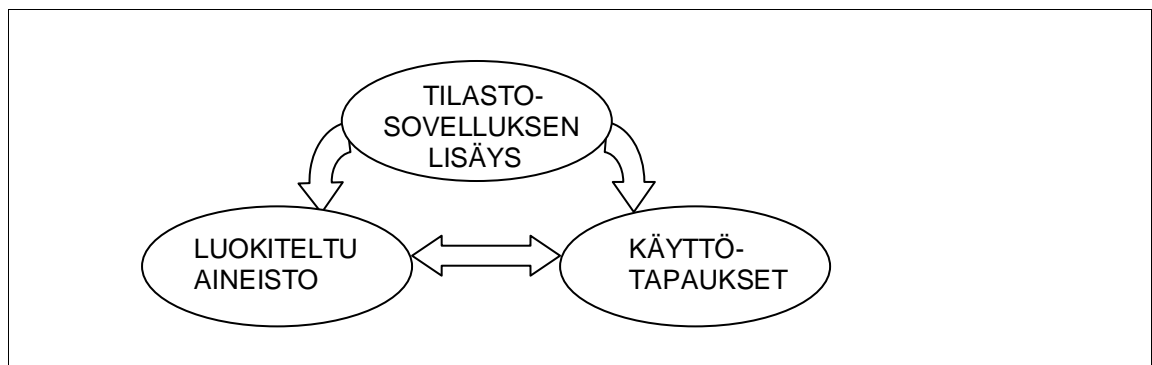
6.1 Väestönaineiston tilastoinnin lähtökohdat

KATIHA-ohjelman täydentäminen tilastosovelluksella edesauttaa tietokannan laajan kvantitatiivisen aineiston tutkimista ja aineistojen kokoamisen jopa poikkitieteellisiin tutkimushankkeisiin, kuten Lummaan projektissa. Kirkonkirjamerkintöihin pohjautuvien tutkimusten luotettavuuden arvioinnissa tulisi kuitenkin muistaa se tosiasia, ettei välttämättä kaikista väestötapahtumista ole olemassa kirjauksia. Esim. mahdollisesti jokaista seurakunnassa syntynyttä lasta ei ole merkitty syntyneiden ja kastettujen luetteluihin. Useimmiten kuolleena syntyneet ja kastamattomina kuolleet lapset jätettiin merkitsemättä, sillä he eivät olleet missään vaiheessa seurakunnan jäseniä. On arvioitu (Happonen 2004, 37), että 1800-luvun puoliväliin saakka syntyneiden ja kastettujen luetteloista puuttuu 10–20 % lapsista.

Karjala-tietokannan aineiston hyödyntäminen vuosittaista seurakunnan aikuisväestön määrää laskettaessa ei ole mahdollista, sillä rippikirjan sivulla henkilöistä voi olla tietoja kymmenenkin vuoden ajalta. Päivi Happonen (2009, 22) esittää kuitenkin mahdollisuuden hyödyntää rippikirjoja väestömäärien laskentaan. Hänen mukaansa laskennan apuna voisi olla vuosittaiset ehtoollismerkinnät, mutta niitä ei ole tallennettu Karjala-tietokantaan niiden paljouden ja päivämäärien tulkintavaikkeuksien takia.

Vaikka Karjala-tietokannan tietosisällössä havaitaankin puutteita ja yksittäisiä tallennusvirheitä, voidaan tietokantaa tarjota tieteellisiin tutkimuksiin. Marko Lambergin mukaan (Lamberg 2008) tutkija joutuu aina ratkaisemaan, kannattaako käydä alkupe- räisaineisto itse lävitse vai onko järkevämpää hyödyntää tarjolla olevia tutkimusrekistereitä. Lamberg huomauttaa, että ainakin suuria väestömääriä käsiteltäessä mahdolliset yksittäiset tutkimusrekisterin sisältämät virheet hukkuvat tilastojen virhemarginaaleihin.

Tilastosovelluksen lisäämisen lähtökohdat voidaan tarkentaa luvussa 5 mainituista (kuva 5 s. 27) lähtökohdista ja ilmaista ne kuvan 8 mukaisesti kolmen tutkimuselementin avulla, jolloin 1. KATIHAN kehittäminen = tilastosovelluksen lisäys, 2. Tietokanta = luokiteltu aineisto ja 3. Tilastosovellus = käyttötapaukset.



KUVA 8. KATIHA-tilastosovelluksen lisäämisen lähtökohdat (vert. kuva 5)

Tilastosovelluksen toteuttaminen voidaan jakaa kahteen yhtä tärkeään osaan: aineiston käsittelyyn ja käyttötapauksen ohjelmointiin. Edellisessä luvussa esitettyjen syntyneiden, kuolleiden, vihittyjen ja muuttaneiden tietoihin kohdistuvien käyttötapauksen ohjelmointiprosessia ei tässä opinnäytetyössä selvitetä erikseen, sillä tilastosovelluksen ohjelmointi annettiin vuoden alkupuolella Teppo Hytösen tehtäväksi. Sen sijaan opinnäytetyön soveltava osuus kohdistuu tietokanta-aineiston luokitteluun.

6.2 Ammattien luokittelu

Väestötietoaineiston sisältöä täytyy luokitella, jotta esim. syntyneiden aineistosta voitaisiin tilastoida syntyneiden lasten lukumäärät vanhempien ammattiryhmän mukaan tai esim. kuolleiden aineistosta hakea kuolleiden lukumäärät kuolinsyyn mukaan. Koska Karjala-tietokantaan on tallennettu jo noin 3 660 000 henkilömerkintää luovutetun alueen seurakuntien syntyneiden, kuolleiden, vihittyjen ja muuttaneiden luettelosta, löytyy tilastosovelluksen tietokanta-aineistosta monta erilaista tapaa merkitä ammattinimikkeet ja kuolinsyyt.

Vuosina 2001–2005 Karjala-tietokannasta on poimittu yhteensä 29 000 erilaista ammattinimikkeen kirjoitusasua. Nimikkeiden suuri määrä johtuu siitä, että sama ammattinimike on kirjoitettu monella eri tavalla. Esimerkiksi leskestä löytyy monta eri muotoa: leski, lki, l.ki, l:ki, änka, enka, enklng. Ammattien lisäksi ammattinimike sisältää myös perhesuhteita ilmaisevat merkinnät; vaimo, poika, tytär, vävy, miniä jne.

Karjala-tietokantasäätiössä ammattinimikkeiden luokittelu aloitettiin 2000-luvun alussa silloisen tutkimuspäällikkö Päivi Repo-Lehikoisen aloitteesta. Karjala-tietokannassa esiintyvistä ammattinimikkeistä toimitettiin tietoja hollantilaisen sosiaalishistorian instituutin käynnistämään HISCO-projektiin, jonka tarkoituksena on edistää kansainvälistä työn historian tutkimusta alkaen 1800-luvulta ja ulottuen 1900-luvun alkupuolelle saakka (HISCO 2011). Kansainvälisen HISCO-ohjeistuksen mukaan ammattinimikkeistä muodostettiin suomalainen 466 koodin HISCO-luokitus (liite 2).

Miksi kansainväliseen kooditukseen on yleensä ryhdytty? Tarve vertailukelpoiseen kansainväliseen historialliseen ammattiluokitukseen (HISCO=Historical International Standard Classification of Occupations) syntyi kolmen European Social Science History Conferences –kokoontumisen pohjalta. Konferenssien yhteydessä vuosina 1958, 1968 ja 1988 syntyi kolme versiota kansainvälisestä ammattiluokituksesta: ISCO58, ISCO68 ja ISCO88 (HISCO 2002, 10). Ammattien historialliseen ISCO-kooditukseen eli HISCOon tuli Euroopan ulkopuolelta mukaan Kanada. Muut HISCOssa vuosittain vaihteessa mukana olevat tutkijat olivat Belgiasta, Iso-Britanniasta, Ranskasta, Saksasta, Hollannista, Norjasta ja Ruotsista (HISCO 2002, 12).

Väestöhistoriallisten aineistojen luokittelussa voidaan toki käyttää muitakin luokittelumalleja. Esimerkiksi Sortavalan väestöaineistoa vuosilta 1800–1940 käsittelevässä väitöskirjassaan Päivi Happonen (ent. Repo-Lehikoinen) on käyttänyt kirkonkirjaväestön ammatillisessa ja sosiaalisessa luokittelussa Heikki Wariksen ryhmittelyä (Happonen 2009, 167):

- I Julkinen virka tai vapaa ammatti (esim. kirkkoherra, kruununvouti, opettaja)
- II Toimenhaltijat ja virkailijat (esimerkiksi palovartija, tullivirkailija)
- III Liike-elämä (esimerkiksi kauppias, tehtailija, käsityöläinen)
- IV Ruumiillinen työ (esimerkiksi työmies, renki, piika)
- V Muut (esimerkiksi neiti, vuokralainen)

Päivi Happonen (2009, 167) mainitsee lisäksi, että väestö olisi ryhmiteltävissä ammatillisesti ja sosiaalisesti myös nykyisissä väestötilastoissa käytettävän elinkeinoluokituksen mukaan.

6.3 Kuolinsyiden luokittelu

Vuonna 1936 Suomessa astui voimaan asetus kuolleisuustilastoa varten toimitettavista tiedoista. Asetus edellytti pääsääntöisesti lääkärin antamaa kuolintodistusta kaikista vainajista ja asetuksen myötä otettiin käyttöön lääkintöhallituksen vahvistama kuolemansyynimistö (Valtioneuvosto 1935). Nimistön taustalla oli vuonna 1929 hyväksytty kansainvälinen tautiluokitus, jossa oli yhteensä 192 kuolinsyytä 18 ryhmässä (Forsius 2007b). Tämä kuolemansyynimistö oli sittemmin pohjana koko maan käsittävälle kuolinsyiden yksityiskohtaiselle erittelylle (SVT 1940).

Vuoden 1936 asetuksen mukaista luokittelua ryhdyttiin käyttämään Karjala-tietokannasta löytyvien kuolinsyiden luokitteluun noin 10 vuotta sitten eli vuosituhannen vaihteessa. Tällöin luokittelutarve tuli ajankohtaiseksi tutkimuspäällikkö Päivi Repo-Lehikoisen Sortavalan asukkaisiin liittyvän liseniaattityön yhteydessä. Koemielessä luokiteltiin ensin pienen Lavansaaren seurakunnan kuolleiden luetteloista poimitut kuolinsyyt, minkä jälkeen siirryttiin Sortavalan kaupunkiseurakunnan ja maaseurakunnan kuolinsyihin. Karjala-tietokannassa olevien kuolinsyiden luokittelua on sittemmin jatkettu vuoden 1935 tautinimistöön perustuvan koodiston (liite 3) avulla.

6.3.1 Kuolinsyyluokittelun historiallista taustaa

Väestöhistoriassa kuolinsyiden luokittelulla on pitkät perinteet. Arno Forsiuksen historiallisen katsauksen (Forsius 2010b) mukaan tautinimistöistä edettiin myös sairauksien tilastointiin. Kulkutautiepidemioiden aiheuttamiin kuolemantapauksiin liittyviä tilastoja on olemassa Venetsiasta jo 1500-luvulta. Vuonna 1610 sveitsiläinen lääkäri Felix Platter laati jopa tarkistettuja tilastoja kuolinsyiden osalta. Hän kulki Baselissa ruton jälkeen talosta taloon, laski ruttoon kuolleet sekä vertasi lukuja kirkonkirjoissa ruttoon kuolleiksi merkittyjen määriin.

Terveystieteeseen liittyvä tilastointi alkoi kuitenkin kehittyä vasta 1600-luvun loppupuolella. Myöhemmin väestötieteen isäksikin kutsutun lontoolaisen John Grauntin teos "Natural and Political Observations --- made upon the Bills of Mortality" ilmestyi vuonna 1662. Tämä kuolinsyihin liittyvä tutkimus (vert. opinnäytetyön s. 4) teki lääkärin ammattikuntaan niin suuren vaikutuksen, että Graunt hyväksyttiin ilman lääketieteen koulutusta Royal Society'n jäseneksi. Saksassa Johann Peter Süssmilch julkaisi vuonna 1741 väestökirjanpidosta teoksen, joka sisälsi tutkimuksen Brandenburgin 1056 eri seurakunnan väestöstä. Vuonna 1756 hän laati ensimmäisen kuolleisuustaulukon koko Preussin väestöstä. Süssmilch korosti väestökehityksen tutkimisessa suurten ja pitkäaikaisten tilastojen merkitystä. Häntä voidaan pitää yhteiskuntaa koskevan tilastollisen ja tilastomatematiikan tutkimuksen uranuurtajana. Siinä hän edelsi belgialaista tilastotieteilijää Lambert Adolphe Jacques Quételetä, joka loi tilastollisen käsitteen "keskiarvoihminen". (Forsius 2010b.)

Englantilainen Thomas Robert Malthus tuli kuuluisaksi puolestaan julkaisemalla "Essay on the Principle of Population" -kirjoituksen vuonna 1798. Malthus esitti ristiriidan väestön kasvumahdollisuuksien ja toimeentulomahdollisuuksien välillä. Malthuksen mukaan väestö kasvaa geometrisesti kaksinkertaistuen aina 25 vuodessa, mutta ravinnontuotanto kasvaa kuitenkin 25 vuoden aikana vain vakiomäärällä. Malthus korosti sitä riippuvuutta, mikä väestön kasvun osalta on merkittävä, nimittäin riippuvuutta ravinnonsaannista. (Pressat 1972, 17.)

Tautinimistöjä koskevan systematiikan kehittäjänä voidaan pitää ranskalaista lääkäriä nimeltään François Boissier de Sauvages, joka vuonna 1763 jakoi sairaudet kymmeen luokkaan, 295 sukuun ja 2400 lajiin. 1800-luvulla oli kuitenkin suosiossa brittiläisen William Cullenin vuonna 1785 julkaisema tautiluokittelu "Synopsis nosologiae

methodicae". Cullen jakoi sairaudet neljään pääryhmään, jotka olivat kuumetaudit, neuroosit, näivetystaudit ja paikalliset taudit. Forsius huomauttaa, että nykyaikaiseen lääketieteeseen verrattuna tautien luokitusten perusteet olivat vielä puutteellisia, sillä esim. mikrobien, aineenvaihdunnan häiriöiden ja perinnöllisyyden aiheuttamien sairauksien syitä ei vielä tunnettu. (Forsius 2010b.)

Lääketieteen nopean kansainvälistymisen seurauksena 1800-luvulla ilmaantui tarvetta luoda yhdenmukaisia ja mahdollisimman luotettavia tautinimistöjä ja -luokituksia eri maiden tilastolliseen vertailuun. (Forsius 2007a.)

6.3.2 Kuolinsyiden luokittelu Suomessa

Ensimmäinen kuolinsyiden luokitus Ruotsissa ja Suomessa otettiin käyttöön vuonna 1749 seurakuntien pitämässä väkilukutauluissa. 1750-luvun alussa aloitettiin valtakunnallinen väestötapahtumien tilastointi, aikaisemmin kuin missään muualla maailmassa. Tilastointi käsitti vihityt, syntyneet ja kuolleet. Kuolleiden kohdalla tilastoihin kirjattiin kuolinikä sekä kuolinsyy mahdollisimman tarkasti. Useat tilastoidut kuolinsyyt olivat kuitenkin pelkästään oireenmukaisia ja niiden aiheuttajana saattoi olla monenlaisia sairauksia. (Forsius 2010a.)

Kuolinsyiden luokituksissa tapahtui muutoksia 1774, 1802 ja 1812. Kuolinsyiden merkitsemisen ohjeena oli vuotuisen väestömuutostaulun tautiluettelo, jonka mukaisesti kuolinsyyt tuli eritellä ja merkitä. Kuolinsyiden tilastojen tarkkuuden lisäämiseksi seurakuntia kehoitettiin hankkimaan papiston käytettäväksi kansantajuisia lääkärikirjoja. Lisäksi pappeja kehoitettiin huolellisuuteen kuolinsyiden merkitsemisessä. Siitä huolimatta kuolinsyytilastot olivat epätarkkoja ja puutteellisia. Vuonna 1844 todettiin, että kuolinsyytilastot olivat jääneet jälkeen lääketieteen kehityksestä. Samalla kiinnitettiin myös huomiota väkilukutaulujen kuolleisuustilastojen mahdollisiin virheisiin ja vääristymiin. Tilastoista ilmeni, että ilmoittamaton tauti esiintyi kuolinsyynä liian usein. Sen vuoksi vuonna 1859 annettiin ohje, että epäselvässä kuolintapauksessa on vaadittava kaupunginlääkäriltä kirjallinen lausunto oikeasta tai ainakin todennäköisestä taudin nimestä, joka tulee myös ilmoittaa kuolinsyynä kuolleiden luetteloon. (Forsius 2007b.)

Kuolinsyylluettelo julkaistiin ensimmäisen kerran suomeksi vuonna 1878, jolloin kuolinsyiden nimikkeistö väheni muiden kuin tartunta- ja kulkutautien sekä tapaturmaisten kuolinsyiden kohdalta. Suomenkielisten taudinimien käyttö lisäsi kuitenkin kuolinsyiden kirjavuutta, sillä papisto alkoi käyttää taudinimistä myös niiden kansankielen mukaisia vastineita. Tästä aiheutui myös lisävaivaa, sillä nämäkin nimet oli sovittava väkilukutaulukon nimikkeiden mukaisiin sarakkeisiin. Vuonna 1880 astui voimaan terveydenhoitojärjestys, joka velvoitti kaupunginlääkärit antamaan kirjallinen lausunto kuolinsyystä, jos vainaja oli eläessään ollut lääkärinhoidossa. Lääkintöhallitus puolestaan velvoitettiin laatimaan tauti- ja kuolleisuustilastojen keräämistä varten lomakkeet, joilla kaupungeissa toimivat lääkärit voivat tehdä tarvittavat ilmoitukset. Maalaiskunnissa käytäntö jatkui kuitenkin entiseen tapaan. (Forsius 2007b.)

Lääkintöhallitus määritteli vuonna 1923 sairaaloille ensimmäisen laajemman tautiluettelon, jota käytettiin sairaaloissa myös kuolleiden kuolinsyiden ilmoittamiseen. Arno Forsiuksen (2007b) mukaan tämä yhteensä 348 taudinimeä ja 21 ryhmää sisältävä luettelo vastasi hyvin aikansa lääketiedettä.

Tautinimistöjen tilastoinnin syynä oli aluksi sairauksien yleisyyden ja esiintyvyyden tutkimus, joka palveli erityisesti hallinnon tavoitteita. Myöhemmin tilastoinnista saatavien tulosten avulla kehitettiin myös hoitojärjestelmiä. Lisäksi terveydenhuoltoon liittyvä tilastointi loi pohjan virallisten väestötilastojen kehittämiseen. (Forsius 2007a.)

Kuolinsyytilastojen paikkansapitävyys on ollut ja on edelleen sidoksissa kuolinsyiden määrittelyihin ja tunnistamisiin. Joidenkin kuolintapausten kohdalla voi vielä ruumiinavauksen jälkeenkin olla vaikeaa määrittellä vainajan varsinainen kuolinsyy. Kuolinsyiden tilastoinnin hyödyllisyys riippuu lisäksi siitä, miten luotettavasti kuolinsyyt ja muut yksilöidyt muuttajat pystytään sijoittamaan tarkoituksensa mukaisiin ryhmiin. Arno Forsius arvioi, että yleisesti ottaen Suomen kuolleisuustilastot vastaavat nykyään hyvinkin korkeita vaatimuksia. (Forsius 2007b.)

6.3.3 Karjala-tietokannassa esiintyvien kuolinsyiden luokittelu

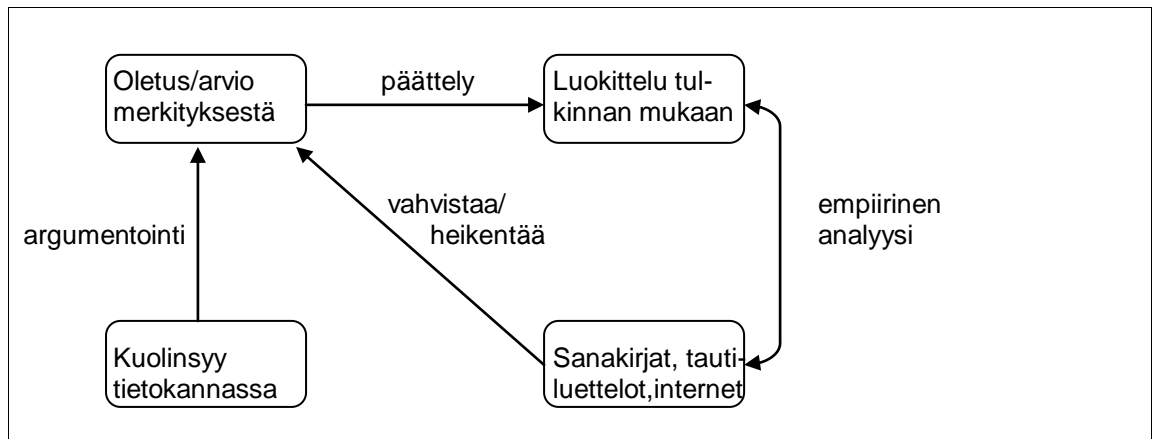
Karjala-tietokannan aineistoon sisältyvien kuolinsyiden luokittelussa on tiedostettava kuolinsyymerkintöjen historiallinen tausta. Vaikka Suomen väestötilastoinnilla on jo vuodesta 1749 alkavat perinteet, ennen 1900-lukua oleviin kuolinsyylastoihin on Arno Forsiuksen (2010a) mukaan kuitenkin suhtauduttava varauksella. Hän kritisoi, että tilastot ovat epäluotettavia ja siten vain suuntaa antavia. Forsius myöntää kuitenkin, että tietojen vertailu muihin lähteisiin voi merkittävästi lisätä kuolinsyiden oikeellisuutta.

Lisäksi Forsius muistuttaa, etteivät nykyiset sanakirjat useinkaan sovellu vuosien taikaisten taudinimien ja kuolinsyiden tulkitsemiseen. Kymmenien ja satojen vuosien aikana tiedot tautien aiheuttajista sekä käsitykset sairauksien olemuksesta ja taudinimien merkityksistä ovat muuttuneet lääketieteen kehittyessä. Historiallisissa lähteissä olevien taudinimien tulkitsemisen vaikeutta kuvaa Forsiuksen mukaan hyvin se, ettei edes lääketieteen asiantuntija ei pysty aina selvittämään, mitä nykyajan sairautta aikoinaan käytetty termi tarkoittaa. (Forsius 2010a.)

Esimerkkinä taudinmäärittelyn muuttumisesta Forsius mainitsee kurkkumädän, jonka taudinkuva tuli tunnetuksi vasta 1800-luvun puolivälissä. Sitä ennen ja sen jälkeenkin kyseinen tauti määriteltiin usein samantapaisin oirein esiintyneen angiinan tai muiden kurkkukipujen mukaan. (Forsius 2007b.)

Kurkkumädän tautiluokittelun haasteellisuus tuli ilmi myös Karjala-tietokannan kuolinsyiden luokittelun yhteydessä. Elo-syyskuussa 2011 tämän opinnäytetyön tekijä luokitteli Karjala-tietokannasta noin 50 000 eri kuolinsyymerkintää. Monet tuhannet merkinnät olivat selviä taudinimiä tai lyhenteitä nimistä, jotka oli melko sujuvaa luokitella Karjala-tietokantasäätiön käyttämän 218 koodia sisältävän koodiston (liite 3) mukaan. Mutta esimerkiksi juuri kurkkumädän luokittelu aiheutti päänvaivaa. Taudin nimi oli kirjoitettu ruotsiksi ”Hängsjuka”. Mutta kyseistä termiä ei kuitenkaan löytenyt vuoden 1936 asetuksen mukaisesta tautinimistöstä, joka sisältää suomenkielisen tautiluokituksen lisäksi myös tautien ruotsinkieliset ja latinankieliset nimet (Valtioneuvosto 1935). ”Hängsjuka”-taudin kohdalla voisi hyvinkin kouluruotsin perusteella päätellä, että kyseessä olisi hirttäytyminen. Luotettavan kuolinsyykoodauksen ei kuitenkaan tule perustua luokittelijan omaan arvioon tai tulkintaan vaan faktatietoon.

Kuolinsyiden luokituksessa voidaan soveltaa (kuva 9) Mikko Ketokiven esitystä hypoteesin menetelmän loogisesta rekonstruktiosta (Ketokivi 2009, 39), jonka teoreettinen kuvaus on sivulla 15.



KUVA 9. Hypoteesin menetelmän soveltaminen kuolinsyiden luokituksessa

Karjala-tietokantaan tallennetuista kuolinsyistä löytyy siis termi ”Hängsjuka”. Vuoden 1935 virallisesta tautinimistöstä sitä ei löydy, mutta Google-haun avulla termistä löytyy muutamia tulkintoja. Yhdysvaltalainen sukututkija June Pelo on julkaissut sivustollaan listauksen kirkonkirjoissa esiintyvistä ruotsinkielisistä kuolinsyistä. Hänen mukaansa ”Hängsjuka = weakness; craziness” eli heikkous tai mielisairaus. (Pelo 2007). Pelon tulkinta juontanee ilmeisesti juurensa siitä, että monesti kirkonkirjoissa hirttäytyminen on kirjattu sanoin ”hirtti itsensä mielenhäiriössä” tai ”hirttäytynyt mielenheikkoudessa”. Mutta tarkoittaako ”Hängsjuka” sittenkään hirttäytymistä? Tietoverkosta löytyy nimittäin vuodelta 1801 peräisin oleva digitoitu lääketieteellinen *Lexicon Nosologicum Polygloton* –sanakirja, josta ”Hängsjuka”-sanalle löytyy synonyymi ”Cardialgia” eli sydänsairaus (Nemnich 1801).

Digitoidun sanakirjan mukaan voisi päätellä, että ”Hängsjuka” on suomeksi sydänsairaus. Mutta tarkempi tutkimus osoittaa, että kyseessä onkin kuristustauti eli kurkkumätä. Vastaus löytyy ”Likhet mellan språken” -sivulta (Suonttavaara Lappby 2010), jossa on suora lainaus meänkielisestä sananlaskusta ja sen selitys suluissa: ”Hirttotauti se kurkun paisutta ja hirttämällä se lopetta (kurkkumätä, tukehduuttaen).” Huolimatta siitä, ettei selitys löytynyt lääketieteellisestä sanakirjasta tai kuolinsyyluettelosta, voidaan kyseistä tulkintaa pitää oikeaan osuvana, sillä meänkieli on Tornionlaaksossa puhuttua suomen kieltä, jossa on vaikutuksia ruotsin kielestä.

Toisena esimerkkinä Karjala-tietokannan kuolinsyiden tulkintavaikeuksista voidaan esittää puolestaan saksankielinen kuolinsyö ”Schwindsucht”. Lauri Hirvensalon vuonna 1966 ilmestyneen saksankielisen sanakirjan mukaan (Hirvensalo 1966, 1147) ”Schwindsucht = surkastus-, keuhkotauti, selkäydinkato”. Näille kolmelle eri taudinimelle on kaikille eri kuolinsyökoodi, joten mikä kyseisistä suomennoksista pitäisi valita? Internetistä löytyy asiaan lisävalaistusta esim. historiallisia saksankielisiä lääketieteellisiä termejä sisältävältä ”Archaic Medical Terms German Home” –sivustolta (Schmidt 2011). Ko. sivuston mukaan ”Schwindsucht = Consumption, but usually applied to Phthisis Pulmonalis”. Termi ”Schwindsucht” tarkoittaa siis surkastumista, mutta on yleisesti käytetty ilmaisu keuhkotuberkuloosista. Karjala-tietokannan kuolinsyyluokittelussa ko. saksankielinen taudin nimi on luokiteltu selkäydinkadon koodilla, joten ”Schwindsucht”-syyn luokitus on korjattava keuhkotuberkuloosin mukaiseksi.

Oman mielenkiintonsa kuolinsyiden luokittamiseen tuovat kuolinsyöt, joita varten ei ole omaa koodia. Pintapuolisesti toimien ne voitaisiin luokitella ryhmään 9050 = muu (liite 3). Esim. kuolinsyyn ”Andtäppä” kohdalla näin voitaisiinkin perustellusti toimia, sillä Historismi-sivuston mukaan ”Andtäppä = hengenahdistus, sydän-, keuhko- tai muu sairaus (Historismi 2006). Sukututkijoiden VaskiVakka-sivusto (VaskiVakka 2009) puolestaan yksinkertaistaa tulkintaa: ”Andtäppä= hengenahdistus”, mutta samasta luettelosta löytyy toinenkin hengenahdistusta tarkoittava sana, joka on ”Astma”. Terveystiedon verkkosanakirjasta (Poliklinikka 2011) hengenahdistukselle löytyy merkitys ”vaikeutunut hengitys”. Verkkosanakirjan mukaan astma on äkillinen, ajoittaisina kohtauksina ilmenevä hengenahdistus, kun taas keuhkoastma on pitkäaikainen tila, jolle ominaisia ovat hengenahdistuskohtaukset. Lisäksi verkkosanakirja määrittelee sydänastman ahdistuskohtaukseksi, jonka syynä on sydämen vajaatoiminta ja nesteen kertyminen keuhkoihin eli keuhkopöhö.

Hengenahdistusta ja/tai astmaa tarkoittavat kuolinsyöt onkin luokiteltu Karjala-tietokannan kuolinsyissä kirjavasti, esim. kuolinsyöt: ”influenssa, astma” = 1125, ”keuhkotauti, keuhkoastma” = 3570, ”sydänlihasten taudit, astma” = 3030, ”keuhkoastma” = 3570 ja ”hengenahdistus” = 3585. Toisaalta tällainen toisistaan poikkeava luokittelu johtuu siitä, että joissakin tapauksissa kuolinsyynä mainitaan kaksi eri syytä, joista jompikumpi on valittava luokittelun perusteeksi.

Kuolinsyiden luokittelussa joutuu tekemään valintoja myös eri merkitysten välillä. Esim. ”Moderpassion”-sairaudelle löytyy kaksi merkitystä: hysteria ja kohdunlaskeuma. Tämän voi todeta Ruotsissa toimivan Demografisk Databas Södra Sverige – yksikön ”Dödsorsaker och sjukdomsnamn” –listan kuolinsyistä (DDSS 2011). Myös Karjala-tietokantasäätiön vuonna 2004 julkaiseman sukututkimuksen asiakirjaoppaan sisältämästä luettelosta löytyy tulkinta ”Moderpassion = hysteria, kohtutauti” (Happonen 2004, 230). Karjala-tietokannan kuolinsyiden luokittelussa taudin kohdalle on merkitty koodi 2820 eli ”Muut hermoston ja aistimien taudit”.

Monet Karjala-tietokannan aineistossa esiintyvät taudinnimet ovat aikaansa sidottuja eikä niitä enää nykyisin käytetä, esim. ”Kräveta”. Taudin merkitys ilmenee vanhasta sähköiseen muotoon digitoidusta vuoden 1859 sanakirjasta ”Samlingar och anteckningar till en beskrifning öfver Ydre Härad i Östergöthland”, jonka sivulta 48 voidaan todeta, että ”Kräveta = sjukdomen kräftan” eli kyseessä on syöpä (Räaf 1859, 48). Joidenkin tautien osalta myös taudin nimeämisessä on ollut vaihtelevuutta eri aikoina. Arno Forsius toteaaakin, että mm. lavantaudin, pilkkukuumeen ja toisintokuumeen erottaminen toisistaan on aiemmin ollut melko mahdotonta (Forsius 2007b).

Karjala-tietokannasta löytyvät kuolinsyyt on luokiteltu vuoden 1936 asetuksen mukaan, mutta luokitustapoja on olemassa muitakin. Esimerkiksi Tilastokeskuksen ja Kuntoutussäätiön yhteisessä tutkimuksessa ”Ammatit ja kuolleisuus - Työllisten ja työttömien ammattiryhmittäinen kuolleisuus 1996–2000” kuolinsyyt on ryhmitelty Maailman Terveysjärjestö WHO:n laatiman kansainvälisen tautiluokituksen 10. uudistuksen (ICD-10 1999) pohjalta kuuteen luokkaan: 1) Kasvaimet, 2) Verenkiertoelinten sairaudet, 3) Hengityselinten taudit, 4) Muut taudit sis. myös ”erittelemättömät taudit”, 5) Itsemurhat ja 6) Muut tapaturmat ja väkivalta. Lisäksi kuolinsyyt on luokiteltu kahteen ryhmään edellä olevista ryhmistä riippumattomasti: a) Alkoholisytyt sis. alkoholiperäiset taudit sekä tapaturmaiset alkoholimyrkytykset) ja b) Muut kuin alkoholisytyt. (Pensola & Ahonen & Notkola 2004, 15–16.)

Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskus STAKESin julkaisemassa ICD-10 –tautiluokituksen suomalaisessa versiossa ohjeistetaan tilastollisten peruskuolemansyiden valinta- ja soveltamissääntöjen käyttöä, sillä eri maiden kuolinsyytilastojen vertailun mahdollistamiseksi on erittäin tärkeää pyrkiä yhdenmukaisiin ja tarkkaan ohjeistettuihin käytäntöihin. (ICD-10 1999.)

Kuolintodistuksessa mainittu kuolinsyy ei välttämättä ole sama kuin kyseisen kuolintapauksen tilastollinen peruskuolemansyy, sillä peruskuolemansyyin päättely perustuu ensin valintasääntöjen ja sitten soveltamissääntöjen käyttöön. ICD-10:n valintasääntöjen avulla arvioidaan kuolintodistuksesta ilmenevää kuolinsyiden kausaaliketjua edeten välittömästi kuolinsyystä aina peruskuolemansyyhyn asti. Soveltamissäännöt puolestaan ohjaavat menettelyä, jossa joillekin tautimerkinnöille annetaan etusija muihin nähden tai yhdistellään useita eri tautitiloja. ICD-10:n käytön ohjeistuksesta (ICD-10 1999) voidaan mainita esimerkkinä peruskuolemansyy ”vanhuudenheikkous”, josta ohjeistuksen mukaan olisi kuolemansyytilastoissa kuitenkin tarkoituksenmukaisempaa käyttää jotakin muuta vanhenemisprosessiin liittyvää taudin nimeä. Yhdysvaltalaiset Kalifornian yliopiston professorit Geoffrey Bowker ja Susan Leigh Star tosin kritisoiivat ICD-luokituksen vaatimuksia ja sitä, ettei kukaan voi enää kuolla yksinkertaisesti vain vanhuuteen (Bowker & Leigh Star 2006, 90).

Karjala-tietokanta-aineistoon sisältyvien kuolinsyiden luokittelussa sovellettiin WHO:n ICD-10 –tautiluokituksen käyttöohjeita sikäli kuin se historiallisten kuolinsyiden osalta on mahdollista. Esim. kuolinsyy ”vanhuus ja sydäntauti” luokiteltiin koodilla 3030 (= sydänlihaksen taudit). Tosin ohjeita noudatettiin luokituksen yhteydessä tiedostamatta, sillä ohjeistus luettiin vasta noin 50 000 kuolinsyyin luokittelun jälkeen. Tilastokeskuksen tilastotieteen verkkomateriaalissa (Byckling & Salonen 2011) todetaan osuvasti, että aineiston luokittelutarve on syytä selvittää etukäteen. Tähän voisi lisätä, että luokitteluohjeetkin olisi hyvä selvittää etukäteen.

Selvätkään luokittelusäännöt eivät aina auta käytännössä, kuten Kari Pitkänen antaa ymmärtää kuolemansyitä käsittelevässä artikkelissaan teoksessa ”Suomen Vuosisata”. Menneiden aikojen tautioppikäsitysten erilaisuuksien vuoksi vanhoja kuolinsyitä on Pitkäsen mukaan vaikea sovittaa nykyaikaisiin luokituksiin. Kuolemansyyin määrittäminen tuotti vaikeuksia vielä 1900-luvun alussakin, mutta helpottui vuonna 1936 toteutetun nimistöuudistuksen myötä. Vasta vuonna 1951 kuolinsyytilastoinnissa ryhdyttiin käyttämään nykyistä peruskuolemansyitä. (Pitkänen 1999, 20.)

Peruskuolemansyyin määrittäminen Karjala-tietokanta aineistosta ei olekaan välttämättöntä, sillä luovutetun alueen kirkonkirjojen ylläpito päättyi 1949 vuoden lopulla. Luokiteltavaa aineistoa ei siis ole enää 1950-luvulta, jolloin käsite peruskuolemansyy otettiin käyttöön.

6.4 KATIHA-tilastosovelluksen luokiteltu tietokanta-aineisto

KATIHA-henkilöhakusovelluksen verkkoaineisto koostuu tietosuojasyistä vain 100 vuotta vanhemmasta tallennusaineistosta lukuun ottamatta kuolleiden luetteloita. Tämä tarkoittaa sitä, että verkossa olevan aineiston määrä on huomattavasti pienempi kuin Karjala-tietokantaan tallennetun aineiston yhteismäärä. Tällä hetkellä tietokanta sisältää jo yli 9 miljoonaa tallennettua henkilömerkintää. KATIHA-henkilöhakusovelluksen yhteydessä on verkossa julkaistu arviolta viisi miljoonaa rippi- ja lastenkirjoista sekä kirkonkirjaluetteloista poimittua henkilömerkintää.

Tilastotietojen julkaisua eivät sadan vuoden tietosuojasäännökset koske, joten tilastosiota varten on perusteltua luoda uusi MySQL-kanta, joka sisältäisi tilastoitavat tiedot 1600-luvun lopulta aina 1950-luvulle asti. Tilastotietojen laskentaa varten syntyneiden, kuolleiden, vihittyjen ja muuttaneiden luetteloista kootaan siten uudet taulut. Tässä vaiheessa rippi- ja lastenkirja-aineisto jätetään tilastosovelluksen ulkopuolelle.

TAULUKKO 7. Syntyneiden taulu KATIHA-tilastosovelluksessa

Muuttuja	Tyyppi	Pituus	Selite
HENKILO	C	12	Henkilönumero
TSRK	C	4	Tallennusseurakunta
SIDOS	C	4	Sidoksen tunnus
SPP	N	2	Syntymäpäivä
SKK	N	2	Syntymäkuukausi
SVU	N	4	Syntymävuosi
SP	C	1	Sukupuoli
AVIOS	C	20	Aviosynty
MONIS	C	1	Monisynty
L_MONIS	C	1	Monisynty (lukumäärä)
YLIV	C	1	Yliviivaus
KASPP	N	2	Kastepäivä
KASKK	N	2	Kastekuukausi
KASVU	N	4	Kastevuosi
AIKA	N	2	Äidin ikä
N_ISAMM	C	30	Normalisoitu isän ammatti
N_AIAMM	C	30	Normalisoitu äidin ammatti
ISAMMRY	C	5	isän ammattiryhmä
AIAMMRY	C	5	äidin ammattiryhmä

Tilastotietojen tarkistusta varten syntyneiden luetteloiden uuteen tauluun (taulukko 7) voidaan sisällyttää sellaisiakin kenttiä, joita ei varsinaisessa tilastoinnissa tarvita. Seuraavat tiedot ovat kuitenkin välttämättömiä, jotta luvussa 5 mainitut käyttötapaukset voidaan toteuttaa: syntymäpäivämäärä (päivä, kuukausi, vuosi), sukupuoli, monisynty,

kastepäivämäärä (päivä, kuukausi, vuosi), äidin ikä, isän ammattiryhmä ja äidin ammattiryhmä. Ammatit on tallennettu Karjala-tietokantaan lähteen mukaisia. Aiemmin mainittua Sortavalan asukkaiden ammattiryhmäluokitusta varten ammattinimikkeitä normalisoitiin 2000-luvulla ja tulokset tallennettiin erilliseen dBase-pohjaiseen tauluun. Tässä yhteydessä normalisoiminen tarkoittaa ammattinimikkeen muuntamista viralliseen asuun, esim. nimikkeen ”Tal.” normalisoitu asu on ”talollinen”. Ammattien luokittelussa tästä menettelystä on hyötyä, sillä tällöin ei tarvitse luokitella saman nimikkeen kaikkia eri kirjoitusmuotoja vaan vain niiden yksi yhteinen virallinen nimiasu. Koska yksittäisten ammattinimikkeiden normalisointi on tallennettu erilliseen tiedostoon, täytyy tästä erillisestä tiedostosta hakea normalisoidun ammatin tiedot syntyneiden uuteen tauluun. Siirrettävät tiedot ovat ammatin normaaliasu ja ammattiryhmä koodina. Muut tiedot siirretään syntyneiden luetteloiden varsinaisista syntyneiden henkilötiedot –tauluista (vert. taulukko 4 sivulla 25).

Myös kuolleiden luetteloiden käyttötapausten toteuttamista varten laaditaan uusi taulu (taulukko 8), jonka seuraavat kentät ovat tarpeellisia laskentaa varten: kuolinpäivämäärä (päivä, kuukausi, vuosi), hautauspäivämäärä (päivä, kuukausi, vuosi) sukupuoli, monisynty, kastepäivämäärä (päivä, kuukausi, vuosi), ikä (vuosina, kuukausina, päivinä), ammattiryhmä, kuolinsyryhmä.

TAULUKKO 8. Kuolleiden taulu KATIHA-tilastosovelluksessa

Muuttuja	Tyyppi	Pituus	Selite
HENKILO	C	12	Henkilönumero
TSRK	C	4	Tallennusseurakunta
SIDOS	C	4	Sidoksen tunnus
SP	C	1	Sukupuoli
SIV	C	1	Siviilisääty
YLIV	C	1	Ylivivaus
KUOPP	N	2	Kuolinpäivä
KUOKK	N	2	Kuolinkuukausi
KUOVV	N	4	Kuolinvuosi
KUOVU	N	4	Kuolinvuosi (johdettu)
HAUPP	N	2	Hautauspäivä
HAUKK	N	2	Hautauskuukausi
HAUVV	N	4	Hautausvuosi
HAUVU	N	4	Hautausvuosi (johdettu)
N_AMM	C	30	Normalisoitu ammatti
AMMRY	C	5	Ammattiryhmä
N_KSYY	C	60	Normalisoitu kuolinsyry
KSYYRY	C	4	Kuolinsyryryhmä
IKAVV	N	3	Ikä vuosina
N_IKAVV	N	3	Ikä vuosina (johdettu)
IKAKK	N	2	Ikä kuukausina
N_IKAKK	N	2	Ikä kuukausina (johdettu)
IKAPP	N	2	Ikä päivinä
N_IKAPP	N	2	Ikä päivinä (Johdettu)

Uuteen kuolleiden tietojen tauluun luodaan myös ”johdettuja” kenttiä seuraavista tiedoista: kuolinvuosi, hautausvuosi, ikä vuosina, ikä kuukausina ja ikä päivinä. Kyseisten kenttien sisältö on periaatteessa sama kuin sen kentän, mistä tieto johdetaan. Mutta jos esim. kuolinvuosi puuttuu alkuperäisestä kuolleiden luettelosta, tallennetaan kuolinvuoden ”johdettuun” kenttään se vuosi, miltä luettelo on peräisin. Jos taas vainajan tiedoista puuttuu ikätieto kuolinhetkellä, voidaan se laskea kuolinpäivämäärän ja syntymäpäivämäärän erona ja tallentaa tulos ”johdettuihin” ikä-kenttiin. Huom. syntymäpäivämäärän kentät sisältyvät tietokannan kuolleiden luetteloiden henkilötietojen varsinaiseen tauluun. Uuteen kuolleiden tauluun siirretään tietoja kolmesta eri tiedostosta, jotka ovat kuolleiden henkilötiedot –tiedosto, ammattiryhmien tiedosto ja kuolinsyiden tiedosto.

TAULUKKO 9. Vihittyjen taulu KATIHA-tilastosovelluksessa

Muuttuja	Tyyppi	Pituus	Selite
TAPAUUS	C	12	Tapausnumero
TSRK	C	4	Tallennusseurakunta
SIDOS	C	4	Sidoksen tunnus
VPP	N	2	Vihkipäivä
VKK	N	2	Vihkikuukausi
VVV	N	4	Vihki vuosi
YLIV	C	1	Ylivivaus
S_HENKILO	C	12	Sulhasen henkilönnumero
S_NAMM	C	30	Sulhasen normalisoitu ammatti
S_AMMRY	C	5	Sulhasen ammattiryhmä
S_STATUS	C	2	Sulhasen sosiaalinen asema
S_SIV	C	1	Sulhasen siviilisäätty
S_IKAVU	N	2	Sulhasen ikä
S_ASP	C	50	Sulhasen asuinpaikka
S_KYLA	C	25	Sulhasen kylä
M_HENKILO	C	12	Morsiamen henkilönnumero
M_NAMM	C	30	Morsiamen normalisoitu ammatti
M_AMMRY	C	5	Morsiamen ammattiryhmä
M_STATUS	C	2	Morsiamen sosiaalinen asema
M_SIV	C	1	Morsiamen siviilisäätty
M_IKAVU	N	2	Morsiamen ikä
M_ASP	C	50	Morsiamen asuinpaikka
M_KYLA	C	25	Morsiamen kylä

Vastaavasti vihittyjen luetteloiden tilastosovelluksen käyttöön laadittavaan tauluun (taulukko 9) sisällytetään tilastoinnin testausta varten sellaisiakin kenttiä, joita ei varsinaisessa tilastosovelluksessa tarvita. Seuraavat tiedot ovat kuitenkin välttämättömiä, jotta luvussa 5 mainitut käyttötapaukset voidaan toteuttaa: vihkipäivämäärä (päivä, kuukausi, vuosi), sulhasen tiedoista ammattiryhmä, siviilisäätty, ikä, asuinpaikka ja kylä sekä morsiamen vastaavat tiedot.

Muuttaneiden luetteloiden tilastosovelluksen käyttöön laadittavaan tauluun (taulukko 10) täytyy suunniteltujen käyttötapauksen mahdollistamiseksi sisällyttää ainakin seuraavat kentät: syntymäpäivämäärä (päivä, kuukausi, vuosi), seurakuntaan sisäänmuuton päivämäärä (päivä, kuukausi, vuosi), seurakunnasta poisuuton eli ulosmuuton päivämäärä (päivä, kuukausi, vuosi), sukupuoli, ammattiryhmä.

TAULUKKO 10. Muuttaneiden taulu KATIHA-tilastosovelluksessa

Muuttaja	Tyyppi	Pituus	Selite
HENKILO	C	12	Henkilönumero
TSRK	C	4	Tallennusseurakunta
SIDOS	C	4	Sidoksen tunnus
SPP	N	2	Syntymäpäivä
SKK	N	2	Syntymäkuukausi
SVV	N	4	Syntymävuosi
SP	C	1	Sukupuoli
SIV	C	1	Siviilisääty
YLIV	C	1	Ylivivaus
MUTY	C	1	Muuttotyyppi
MILKM	N	2	Miesten lukumäärä
NALKM	N	2	Naisten lukumäärä
YHTLKM	N	2	Yhteislukumäärä
N_YHT	N	2	Normalisoitu yhteislukumäärä
MUSRK	C	30	Muuttoseurakunta
MUUTSRK	C	35	Normalisoitu muuttoseurakunta
MSPP	N	2	Sisäänmuuttopäivä
MKPP	N	2	Ulosmuuttopäivä
MUPP	N	2	Muuttopäivä
MSKK	N	2	Sisäänmuuttokuukausi
MKKK	N	2	Ulosmuuttokuukausi
MUKK	N	2	Muuttokuukausi
MUVV	N	4	Muuttovuosi
N_AMM	C	30	Normalisoitu ammatti
AMMRY	C	5	Ammattiryhmä

Karjala-tietokannan suuren tietuemäärän takia tilasto-osiossa käytettävät päätaulut muodostetaan siten, että kunkin seurakunnan tiedot tulevat erillisiin syntyneiden, vihittyjen, muuttaneiden ja kuolleiden tietojen tauluihin. Täten MySQL-kantaa varten perustetaan 4 (taulutyyppi) x 70 (seurakunnat) taulua eli yhteensä 280 taulua. Seurakuntien lukumäärään sisältyy 50 luterilaista seurakuntaa ja 20 ortodoksista seurakuntaa.

Syntyneiden, kuolleiden, vihittyjen ja muuttaneiden uudet taulut voidaan muodostaa entisistä Karjala-tietokannan tauluista siten, että tiedot kopioidaan kuhunkin tyhjään tauluun ”append from” -komennolla. Komento on käyttökelpoinen, sillä Karjala-tietokannan käytössä on Visual dBase -tiedonhallintaohjelma ja tietokannan taulut ovat dBase-pohjaisia. Tietokannan entisten tietojen lisäksi tilastokannan tiedostoihin täytyy liittää myös tiedot ammatti- ja kuolinsyyryhmistä.

Ammatti- ja kuolinsyyryhmien kooditietojen tilastokantaan siirtämistä varten opinnäytetyön tekijä laati viisi eri dBase-ohjelmaa: Synt_amm.prg, Kuol_amm.prg, Vihi_amm.prg, Muut_amm.prg ja Kuol_syy.prg. Näistä ohjelmista esitellään tarkemmin Kuol_amm.prg (liite 4) ja Kuol_syy.prg (liite 5).

```

SELECT 1
DO WHILE .NOT. EOF()
  t_amm=n_amm
  t_ammry="-2 "
  t_henkilo=henkilo
  IF t_amm <> SPACE(30)
    SELECT 2
    SET ORDER TO ammatti
    IF SEEK(t_amm)
      t_ammry=hisco2
    ENDIF
    SELECT 1
    SET ORDER TO henkilo
    SEEK(t_henkilo)
  ENDIF
  IF amm=SPACE(35)
    t_ammry="-0 "
  ENDIF
  REPL ammry WITH t_ammry
  SKIP
ENDDO

```

KUVA 10. Osa Kuol_amm.prg-ohjelman algoritmista (vert. liite 4)

Ammattiryhmätiedot voidaan liittää kuolleiden uuteen tauluun esimerkiksi kuvan 10 esittämällä tavalla. Ammattiryhmäkoodin alkuarvoksi asetetaan ”-2” (ei luokiteltu). Aluksi tutkitaan myös ammatin normalisoinnin tilanne eli onko N_AMM-kentässä tietoa vai ei. Jos ammatti on normalisoitu, etsitään HISCOSU-tiedostosta normalisoitua ammattinimeä vastaava 5-numeroinen koodi, joka joissakin poikkeustapauksissa voi olla lyhyempikin koodi esim. ”-1” (ei ammattia). Tällaisen koodin saavat esimerkiksi ammattinimikkeet ”kalastajan poika” tai ”talollisen tytär”. Jos henkilötiedoissa ei ole ammattinimikettä ollenkaan, ohjelmoituu koodiksi ”-0” (ei mainittu).

HISCOSU-tiedosto sisältää normalisoidun ammatin ja HISCO-koodin tietokenttien lisäksi myös muutaman lisätiedon kentät esim. sosiaalisen aseman ja sukupuolitiedon tallentamista varten. Tiedostossa on tietueita yhteensä 6 232 eli sen verran ammattinimikkeiden normalisoituja asuja, joiden HISCO-koodit tuotettiin pääasiassa manuaalisesti 2000-luvulla Päivi Repo-Lehikoisen aloitteesta. Luokittelun taustalla oli Karjala-tietokannasta poimittua 29 776 erilaista ammattinimikettä. Koodaustarve saatiin pienennettyä viidennekseen, sillä ammattinimikkeet normalisoitiin ensin ja vasta sen jälkeen toteutettiin ammattinimikkeiden normaaliasujen HISCO-luokittelu.

Tämän opinnäytetyön toteutuksen yhteydessä ilmeni tarve normalisoida ammattinimikkeiden lisäksi myös Karjala-tietokannasta löytyvät erilaiset kuolinsyyt niiden koodaamista varten. Tämän perusteena oli se tosiasia, ettei kuolleiden tilastointi kuolinsyiden mukaan ole järkevää ilman kuolinsyiden luokittelua. Heinäkuussa 2011 selvitettiin, kuinka monta erilaista kuolinsyyn merkintätapaa Karjala-tietokannasta löytyy ja vastaukseksi saatiin 49 883 tapaa. Opinnäytetyön aikataulun takia ei löytynyt henkilö- eikä aikaresurssejaan kuolinsyiden normalisoimiseen, joten lopulta päädyttiin eräänlaiseen oikotiehen. Kuolinsyiden kohdalla luokittelukooditus tehtiin suoraan kunkin erilaisen kuolinsyymerkinnän kohdalle. Käytännössä tämä tarkoitti sitä, että kuolinsyistä koottiin erillinen 49 883 tietueen KSYYKOOD-tiedosto, joka sisälsi kentät kuolinsyylle ja kuolinsyryhmälle eli koodille.

Elo-syyskuussa toteutetun kuolinsyiden manuaalisen koodaustyön jälkeen kuolinsyryhmien koodit liitettiin ohjelmallisesti kuolleiden uuteen tauluun kuvan 11 esittämällä tavalla. Ryhmäkoodin alkuarvoksi asetetaan ”9010” (tuntematon, ilmoittamaton). Jos kuolinsyy-kentässä on tietoa, etsitään KSYYKOOD-tiedostosta kuolinsyytä vastaava 4-numeroinen koodi. Jos henkilötiedoissa ei ole kuolinsyytä ollenkaan, kuolinsyykoodiksi tallentuu automaattisesti jo alussa asetettu koodi ”9010”.

```

SELECT 1
DO WHILE .NOT. EOF()
  t_ksyy=ksyy
  t_ksyyry="9010"
  t_henkilo=henkilo
  IF t_ksyy <> SPACE(60)
    SELECT 2
    SET ORDER TO kuolinsyy
    IF SEEK(t_ksyy)
      t_ksyyry=ksyyry
    ENDIF
    SELECT 1
    SET ORDER TO henkilo
    SEEK(t_henkilo)
  ENDIF
  REPL ksyyry WITH t_ksyyry
  SKIP
ENDDO

```

KUVA 11. Osa Kuol_syy.prg-ohjelman algoritmista (vert. liite 5)

Sekä Kuol_amm.prg:n että Kuol_syy.prg:n ohjelmalistauksista (liitteet 4 ja 5) voidaan havaita, että ohjelmat sisältävät mahdollisia käsiteltäviä seurakuntia yhteensä 82 kappaletta. Eroavuus aiemmin esitettyyn seurakuntien määrään (70) johtuu mm. siitä, että Karjala-tietokannan tallennustyö on vielä kesken.

6.4.1 Luokitellun aineiston tilastokäyttö

Karjala-tietokannan määräaikainen ohjelmoija Teppo Hytönen ohjelmoi kesän 2011 aikana tilastosovelluksen käyttötapaukset PHP-ohjelmointikielellä. Sovelluksen kuolleiden luetteloiden lähdetietoja käsittelevän ja tilastoivan käyttöliittymän osoite on <http://www.karjalatk.fi/tilastot/kuol.php> (kuva 12).

The screenshot shows the 'Karjala-tietokantasäätö' web application. At the top, there is a navigation bar with tabs for 'Syntyneet', 'Kuolleet', 'Vihityt', 'Muuttaneet', and 'Ohje'. The 'Kuolleet' tab is selected. Below the navigation bar, the text 'Valittu kirkonkirjatyyppi: Kuolleet' is displayed. On the left, there is a list of parishes (kirkonkirkot) with 'Lavansaari' selected. On the right, there are radio buttons for selecting the type of query ('Kyselyn tyyppi:'). The selected option is 'Kuolleiden henkilöiden määrät kuolinsyyn mukaan'. Below this, there are input fields for 'Aloitusvuosi:', 'Lopetusvuosi:', and 'Aikaväli:'. At the bottom right, there is a 'Lähetä kysely' button.

KUVA 12. KATIHA-tilastosovelluksen käyttöliittymä kuolleiden tietojen osalta

Demoversion käyttöliittymän seurakuntavalikossa oli aluksi valittavana vain Antrean ja Lavansaaren seurakunnat, mutta jatkotestausta varten listaa täydennettiin 70 valittavaan seurakuntaan. Seurakuntalistan oikealta puolelta voidaan valita tilastokyselyn tyyppi, joita kuolleiden kohdalla on valittavana kuusi erilaista. Tilastokyselylle voidaan lisäksi määrittää aloitus- ja lopetusvuosi sekä aikaväli.

Seuraavana esitetään kaksi esimerkkiä Lavansaaren kuolleiden kuolinsyiden tilastoinnista. Ensin vuosirajoiksi annetaan 1750 ja 1790 sekä aikaväliksi 41 ja vertailun vuoksi vuosirajoiksi muutetaan 1850 ja 1890 aikavälin pysyessä ennallaan.

”Lähetä kysely”-painikkeen painamisen jälkeen näytölle saadaan kuolleiden henkilöiden määrät jaoteltuna kuolinsyiden mukaan 40 vuoden ajalta. Taulukoista 11 ja 12 voidaan vertailla Lavansaaren kuolleiden tietoja 1700- ja 1800-luvulla ja mm. havaita, että sadan vuoden aikana kuolinsyy ”tuntematon, ilmoittamaton” on vähentynyt huomattavasti.

TAULUKOT 11–12. Lavansaaren kuolleiden kuolinsyyt 1750–1790 ja 1850–1890

Lavansaari - kuolleet kuolinsyyntä mukaan		Lavansaari - kuolleet kuolinsyyntä mukaan	
Vuodet 1750 - 1790		Vuodet 1850 - 1890	
Kuolinsyy (koodi)	Kuolleita	Kuolinsyy (koodi)	Kuolleita
kuolleena syntyneet (0002)	1	kuolleena syntyneet (0002)	19
hivutus-, näivetystauti, kuihtuminen (0450)	2	heikkous, vaivaisuus (0400)	56
vanhuus, vanhuuden heikkous (0500)	2	hivutus-, näivetystauti, kuihtuminen (0450)	3
isorokko (1060)	64	vanhuus, vanhuuden heikkous (0500)	42
kuristustauti, kurkkumätä (1110)	1	lavantauti (1000)	16
punatauti (1165)	15	malaria (1050)	1
tarttuva keltatauti (1190)	1	isorokko (1060)	42
polttotauti (1350)	48	tuhkarokko (1080)	30
keuhkotauti, tuberkuloosi (1400)	3	tulirokko (1090)	4
kihti (2300)	1	hinku-, tukehdusyskä (1100)	2
muut hermoston ja aistimien taudit (2820)	7	kuristustauti, kurkkumätä (1110)	21
sydänlihaksen taudit (3030)	1	kolera (1140)	7
sydämen sepelvaltimoiden taudit (3035)	1	punatauti (1165)	6
pistokset (3100)	24	tarttuva keltatauti (1190)	2
katarraalinen keuhkokuume (3520)	5	angiinamyrkytys, kitarisa-, kurkkutulehdus (1215)	1
muut hengityselinten taudit (3585)	2	kuume (1345)	30
muut maha- ja suolitaudit (4070)	4	polttotauti (1350)	40
ihon ja ihonalaisen kudoksen taudit (6530)	1	keuhkotauti, tuberkuloosi (1400)	29
hukkuminen (8500)	10	erittelemätön kulku- ja sukupuolitauti (1610)	3
tukehtuminen (8510)	1	verenvuoto, -syöksy, repeämä (2055)	2
palovammat (8520)	1	muut aineenvaihdunnan ja sisäerityksen taudit (2470)	1
paleltuminen (8530)	1	epilepsia, kouristus-, tauti (2630)	1
nälkä (8610)	1	muut hermoston ja aistimien taudit (2820)	20
muut tapaturmat (8650)	2	pistokset (3100)	19
äkillinen kuolema (9000)	1	kurkunpään ja henkitorven tulehdus (3505)	3
turvotus, pöhö (9002)	15	keuhkoastma (3570)	9
tuntematon, ilmoittamaton (9010)	302	muut hengityselinten taudit (3585)	1
kuolleena löydetty (9040)	1	pitkällinen maha- ja suolitulehdus (4020)	1
Yhteensä	518	muut maha- ja suolitaudit (4070)	10
		ripuli (4510)	4
		muut raskaustilan ja synnytystaudit (5700)	3
		hammastaudit (6100)	23
		ihon ja ihonalaisen kudoksen taudit (6530)	2
		syöpä muissa elimissä tai paikkaa ilmoitt. (7060)	1
		hukkuminen (8500)	32
		paleltuminen (8530)	2
		äkillinen myrkytys (8600)	1
		ravintoaineiden aiheuttama myrkytys (8604)	1
		muut tapaturmat (8650)	1
		turvotus, pöhö (9002)	37
		tuntematon, ilmoittamaton (9010)	119
		Yhteensä	647

Syy sille, että tuntemattoman tai ilmoittamattoman taudin määrä on vähentynyt suhteellisen paljon, löytyy kuolinsyiden tunnistamisen ja kirjaamisen historiasta. 1800-luvun puolivälissä nimittäin havaittiin jo tässä opinnäytetyössäkin aiemmin mainittu ilmiö, että ilmoittamaton tauti esiintyi kuolinsyyntä liian usein. Tämän seurauksena pyrittiin epäselvissä kuolintapauksissa pyytämään kirjallinen lääkärintuomio tai määrittämään ainakin todennäköisin kuolinsyy (Forsius 2007b). Karjala-tietokannan aineisto osoittaa, että tuolloin annettuja ohjeita noudatettiin.

6.4.2 Luokittelussa käytettyjen koodistojen metatiedot

Julkisen hallinnon tietohallinnon neuvottelukunnan (JUHTA) alaisuudessa toimiva JHS-jaosto julkaisi 19.9.2011 verkkosivuillaan suositusluonnoksen ”JHS Koodistot ja luokitukset”. Suositusluonnoksen (JHS 2011a) mukaan Suomesta puuttuu yhteinen julkishallinnon koodistopalvelu, jonka kautta hallinnossa yleisesti käytettävät koodistot olisivat yhdenmukaisesti saatavissa ja teknisesti suoraviivaisesti hyödynnettävissä.

Suositusluonnoksessa kuvataan hallintamalli eli prosessi, jolla julkisen hallinnon koodistot ja luokitukset yhtenäistetään sekä semanttisen sisällön että teknisen toteutuksen osalta julkaistavaksi JHS Koodistopalvelimella. Luonnoksessa esitetty suositus liittyy JHS-suositusten laajempaan JHS-sanastotyö ja JHS XML -skeemojen kokonaisuuteen ja sen kohderyhmänä ovat sekä julkisen hallinnon tietoarkkitehtuurin kehittäjät että koodistojen omistajat. Koodistojen muodostamis-, ylläpito- ja käyttömallien sekä keskitetyn teknisen ratkaisun avulla pyritään vähentämään koodistojen päällekkäistä hallintointia ja niiden käyttöön liittyvää manuaalista työtä eri organisaatioissa. Suositusluonnoksessa todetaan, että hallintamallin soveltamiseksi on välttämätöntä toteuttaa kansallinen keskitetty ratkaisu koodistojen ylläpitoa ja jakelua varten. Tarve keskitettyyn ratkaisuun on ilmeinen, sillä julkishallinnon eri organisaatioissa käytetään runsaasti erilaisia koodistoja ja luokituksia. (JHS 2011a.)

JHS Koodistot ja luokitukset -suositusluonnos koskee sekä yksilöiviä että luokittelevia koodistoja. Yksilöivää koodistoa käytetään objektien yksilöinnissä ja eriyttämisessä, jolloin koodiarvo toimii objektin yksiselitteisenä tunnisteena. Sen sijaan luokittelevaa koodistoa käytetään objektien luokituksessa, jolloin objektit pyritään luokittelemaan koodiston koodien avulla. Koodistot voidaan jakaa myös yksitasoisiin ja hierarkkisiin koodistoihin. Koodisto on hierarkkinen, jos koodit muodostavat monitasoisia rakenteita ja koodeja voidaan tarkentaa alemman tason koodien avulla. (JHS 2011a.)

Useat käytössä olevat koodistot ovat yhteisiä hallinnon eri toimialoilla ja organisaatioissa. Koodistojen käyttö- ja esitystavat vaihtelevat kuitenkin organisaatioittain, mikä vaikeuttaa tietojen vaihtoa eri organisaatioiden välillä. Luokitukset on yleensä dokumentoitu tekstimuotoisina asiakirjoina, joista luokituksen käyttöönotto vaatii luokistusten manuaalista syöttämistä luokitusta käyttäviin järjestelmiin. (JHS 2011a.)

Karjala-tietokantasäätiön käytössä olevat ammatti- ja kuolinsyyluokitteluihin liittyvät koodistot (liitteet 2 ja 3) on yleisestä käytännöstä poiketen dokumentoitu sekä teksti- muotoisina asiakirjoina että dBase-tilauina. Koodistot toteutettiin ensin tietokanta- muotoisina ja vasta sen jälkeen Ammry.dbf- ja Ksyyry.dbf-nimisissä tiedostoissa oleva tieto siirrettiin tekstimuotoon dBase-ohjelmaan sisältyvän Export-toiminnon avulla.

Kyseiset KATIHA-tilastosovelluksen koodistot eivät periaatteessa sisälly JHS-jaoston koodistoja koskevien suositusten vaikutuspiiriin, sillä Karjala-tietokantasäätiö on yksityinen toimija eikä osa julkishallintoa. Siitä huolimatta Karjala-tietokannan koodistojen dokumentoinnissa (taulukko 13) voidaan hyödyntää suunniteltua mallia koodistopalvelun koodistojen metatiedosta (JHS 2011b).

TAULUKKO 13. Metatiedot ammatti- ja kuolinsyyryhmien koodistoista.

Koodistokohtainen metatieto	Ammattiryhmien koodisto	Kuolinsyyryhmien koodisto
Yksilöivä tunnus	ei URI-tunnusta (ei ole URL-osoitetta)	ei URI-tunnusta (ei ole URL-osoitetta)
Nimi	HISCOSU-ammattiryhmät	KUOLKOOD-kuolinsyyryhmät
Kuvaus	Suomenkielinen HISCO-koodisto	Kuolinsyyt vuoden 1935 tautinimistöä (asetus vuodelta 1936)
Versio	Ammry.txt (Ammry.dbf)	Ksyyry.txt (Ksyyry.dbf)
Viimeinen päivitys	16.11.2009	23.9.2011
Voimassaolon alkamisaika	vuosi 2000 (kolmen seurakunnan tietojen ammattiluokitus aloitettiin Päivi Repo-Lehikoisen toimesta)	vuosi 2000 (kolmen seurakunnan tietojen kuolinsyyryluokitus aloitettiin Päivi Repo-Lehikoisen toimesta)
Voimassaolon päättymisaika	vuosi 2050 (verrattava vuosittain HISCO-päivityksiin)	vuosi 2050 (tarkistukset, korjaukset ja päivitykset vuosittain)
Käyttöohje	Koodit muodostetaan HISCO-koodista siten, että koodi muodostuu vain viidestä numerosta (ei sis. HISCO-koodin välimerkkejä).	Koodi muodostuu neljästä numerosta vuoden 1935 tautinimistön mukaisesti.
Käsite	ammattiluokitus	kuolinsyyryluokitus
Kohdealue	Karjala-tietokantasäätiön tietokanta-aineisto	Karjala-tietokantasäätiön tietokanta-aineisto
Kohdealueen osa-alue	syntyneiden, kuolleiden, vihittyjen ja muuttaneiden luetteloiden tiedot	kuolleiden luetteloiden tiedot
Toimintaympäristö	KATIHA-tilastosovellus	KATIHA-tilastosovellus
Tarkentaa koodistoa	HISCOSU-koodiston aikaisempi versio	KUOLKOOD-koodiston aikaisempi versio
Tila	Koodisto käytössä, koodisto saatavilla sekä teksti- että tietokantamuotoisena.	Koodisto käytössä, koodisto saatavilla sekä teksti- että tietokantamuotoisena.
Huomioitava koodisto	JUHTA: JHS 150 Ammattiluokitus	WHO: ICD-10 version 2010
Koodiston lähde	HISCO-classification, Leuven University Press 2002	Kuolemansyynimistö, Valtioneuvosto 1935
Sitovuustaso	ei määritelty	ei määritelty
Omistaja	Karjala-tietokantasäätiö	Karjala-tietokantasäätiö
Kieli	suomi	suomi

JHS-jaosto huomauttaa, että samasta koodistosta ylläpidetään monia eri kopiota usein kahdessa tai useammassa eri organisaatiossa, mistä aiheutuu resursseja hukkaavaa

päällekkäistä työtä. Myös koodistojen linkittämisessä käytössä oleviin tietorakenteisiin sovelletaan useita eri toteutusmalleja, sillä esim. koodistojen linkittämistä XML-skeemoihin ei ole ohjeistettu. (JHS 2011a.)

KATIHA-tilastosovelluksen tietokannan muodostamisen yhteydessä ammatti- ja koulinsykkoodistoista haettiin tietoja erityisten dBase-muotoisten linkitysohjelmien avulla. Ohjelmien laadinnassa ei ole noudatettu valmiita toteutusmalleja, sillä ohjelmat ovat vain kyseiseen tarkoitukseen laadittuja (liitteet 4 ja 5) Karjala-tietokantasäätiön omia sovelluksia tietokannan tilastokäytön edistämiseksi.

Tilastokäytössä koodistoja tarvitaan, minkä takia Tilastokeskus hallinnoi Suomessa useimpia JHS-suositusten kohteena olevia luokituksia, esim. JHS 150 Ammattiluokitus -suositusta. Tosin joillakin hallinnonaloilla ylläpidetään omaa koodistopalvelua. Esimerkiksi THL:n eli Terveyden ja hyvinvoinnin laitoksen tarjoama koodistopalvelu sisältää koodistoja, joita hyödynnetään sosiaali- ja terveydenhuollon tietojärjestelmissä. Koodistojen ja luokitusten suositusluonnos vaikuttaa monien julkisen hallinnon ylläpitämien koodistojen hallintaan, mutta suositus ei kuitenkaan koske ISO OID -yksilöintitunnusten hallintaa. (JHS 2011a.)

JHS 159 ISO OID-yksilöintitunnuksen soveltaminen julkishallinnossa -suosituksen tarkoituksena on yhdenmukaistaa OID-yksilöintitunnuksen käyttöä. Yhtenäiset yksilöintikäytännöt ovat tarpeellisia silloin, kun eri tietojärjestelmät vaihtavat tietoa avointen rajapintojen kautta. Tällöin käsiteltävät tietosisällöt mm. koodistot ja niiden versiot on tunnistettava eli yksilöitävä. ISO OID-yksilöintitunnus on objektiin liitettävä numeroarvo, joka yksilöi objektin yksiselitteisesti mm. yksilöintitunnuksille varatussa kansainvälisessä ISO-järjestelmässä. (JUHTA 2010.)

JHS Koodistot ja luokitukset – suositusluonnoksen tekstissä (JHS 2011a) mainitaan kuitenkin, että suosituksen sisältämän hallintamallin yhtenäinen tekninen toteutus voi tukea myös ISO OID -järjestelmään perustuvien yksilöivien koodistojen tallentamista palveluun.

6.4.3 Luokitusten ja koodistojen laatuvaatimukset

Eurostat ja Euroopan kansalliset tilastoviranomaiset ovat sitoutuneet noudattamaan luvussa 3 esitettyjä tilastojen käytännesääntöjä, jotka perustuvat Euroopan tilastoviranomaisten yhdessä määriteltyihin periaatteisiin. Nämä 15 eri periaatetta voidaan jakaa kolmeen pääryhmään. Ensimmäisen ryhmän muodostavat tilastojärjestelmän toimintaedellytyksiin liittyvät periaatteet. Toisen ryhmän periaatteita yhdistää se, että ne liittyvät tilastollisiin prosesseihin. Kolmas ryhmä puolestaan muodostuu tilastotuotteiden laatuun liittyvistä periaatteista. (Tilastokeskus 2008.)

Tilastojen käytännesäännöt sisältävät näkökulman myös tilastoinnin laatuvaatimuksista. Seuraavassa esityksessä (taulukot 14, 15 ja 16) sovelletaan Euroopan tilastoviranomaisten käytännesääntöjen laatuvaatimuksia Karjalasäätiön tietokanta-aineiston luokittelun laatuvaatimuksiin.

TAULUKKO 14. Luokitteluedellytysten laatuvaatimukset

Eurostatin periaate	Käytännesäännön laatuvaatimus	Karjala-tietokantasäätiön luokitusten laatuvaatimus
Ammatillinen riippumattomuus	Tilastotuotanto ja jakelu ovat riippumattomia ulkopuolisista eduista.	Historiallisten aineistojen luokittelu on riippumatonta nykyisistä eri tieteen alojen intresseistä.
Tiedonkeruuvaltuudet	Tarvittava lähdeaineisto on saatavissa.	Luokittelun lähteenä on Karjala-tietokannan aineisto, jonka koaamiseen on tietosuojalautakunnan poikkeuslupa.
Resurssien riittävyys	Käytettävissä olevien resurssien on riitettävä täyttämään tilastoja koskevat vaatimukset.	Luokittelulle on varattava riittävästi resursseja.
Laatuun sitoutuminen	Tilastojen tiedonkeruun, muokkauksen ja jakelun laadun seurannassa noudatetaan sovittuja menettelyjä.	Koodistojen sisältöjen ja luokitteluperusteiden määrittämistä seurataan.
Tilastosalaisuus	Tilastosalaisuuden suojaamisesta on annettu julkisesti nähtävillä olevat määräykset ja ohjeet.	Tietokanta-aineistoon sisältyviä henkilötietoja ei julkisteta, mutta luokitteluperusteet (koodistot) ovat julkisia.
Puolueettomuus ja objektiivisuus	Tilastomenetelmien tiedot ovat julkisia ja julkaistuissa tilastoissa havaitut virheet korjataan ja julkistetaan mahdollisimman nopeasti.	Koodistoissa ja luokittelussa havaitut virheet korjataan ja tilastosovelluksen tietokanta päivitetään mahdollisimman pian.

Käytännesääntöjen periaatteisiin sisältyvän vaatimuksen ammatillisesta riippumattomuudesta voidaan luokitusten kohdalla käsittää tarkoittavan riippumattomuutta mm. väestö- ja tilastotieteen nykyisistä suuntauksista ja tutkimusintresseistä. Tämän vaati-

muksen noudattamatta jättämisestä Mikkeli ja Pakkasvirta (2007, 184) esittävät esimerkin historiasta. 1730-luvulla ilmestyneen englantilaisen lääkärin George Cheyenen teoksen ”The English Malady” esipuheen lopussa oleva matemaattinen arvio Englannin taudin levinneisyydestä viittaa Mikkelin ja Pakkasvirran mukaan tuolloin vallinneeseen ihanteeseen tieteen edistyneisyydestä. Isaac Newtonin esimerkin myötä tuolloin pyrittiin nimittäin eri tieteenaloilla tekemään matemaattisia laskelmia, koska niiden uskottiin lisäävän tutkimusten uskottavuutta. Tuonaikaiset mittalaitteet eivät kuitenkaan olleet tarpeeksi kehittyneitä, jotta saadut tulokset olisivat olleet riittävän luotettavia. (Mikkeli & Pakkasvirta 2007, 184.)

Vaatus tiedonkeruuvaltuuksista täyttyy Karjala-tietokantasäätiön luokituksissa hyvin, sillä luokiteltava aineisto on säätiö omaisuutta ja sen kokoamiseen on saatu tietosuojaalautakunnalta poikkeuslupa vuonna 1992.

Resurssien riittävyys on Karjala-tietokannan aineiston luokittelussa haasteellinen vaatimus. Tietojen luokittelu vaatii asiantuntemusta ja aikaa, jotta luokittelun tuloksena syntyisi tieteelliselle tutkimukselle hyödyllinen aineisto. Ketokivi painottaa, että aineisto on tutkijalle kaikki kaikessa ja toteaa samalla, että korkealuokkaisten aineistojen hankkiminen on työlästä, kallista ja ammattitaitoa vaativaa (Ketokivi 2009, 9).

Luokitusten laatuun tulisi kiinnittää erityistä huomiota. Laatuun sitoutuminen Karjala-tietokannan luokitusten kohdalla vaatii seurantaa sekä koodistojen sisältöjen että luokitteluperusteiden osalta.

Tilastosalaisuus toteutuu luokitellun aineiston julkistamisessa siten, ettei tietokanta-aineistoon sisältyviä henkilötietoja voi edes julkistaa, sillä ne eivät edes sisälly tilastosovelluksen tietokantaan. Luokitteluperusteena olevat koodistot ovat sen sijaan julkisia.

Puolueettomuus ja objektiivisuus edellyttävät, että koodistoja ja luokituksia on mahdollisuus tarkistaa. Jos havaitaan virheitä, korjataan ne asianmukaisesti ja tilastosovelluksen tietokanta päivitetään mahdollisimman pian.

TAULUKKO 15. Luokitteluprosessien laatuvaatimukset

Eurostatin periaate	Käytännösäännön laatuvaatimus	Karjala-tietokantasäätöön luokitusten laatuvaatimus
Pätevät menetelmät	Vakiokäsitteitä, määritelmiä ja luokituksia käytetään johdonmukaisesti.	Koodistoja käytetään luokituksissa johdonmukaisesti.
Asianmukaiset tilastolliset menettelyt	Tiedonkeruuta, tiedon tallennusta ja koodausta valvotaan jatkuvasti ja korjataan tarpeen mukaan.	Koodien käyttöä ja luokittelua valvotaan ja korjataan tarpeen mukaan.
Kohtuullinen vastaustaakka	Jos tarkkoja tietoja ei ole vaivatta saatavilla, hyväksytään estimaatit ja likiarvot.	Luokitteluperusteiden täsmälliseen määrittelyyn käytettävän ajan rajallisuuden takia luokittelussa hyväksytään todennäköisimmän koodin käyttö.
Kustannustehokkuus	Tiedonsyötön, koodauksen ja validoinnin automatisointia parannetaan ja seurataan.	Koodauksen ja luokittelun automatisointia pyritään parantamaan ja seuraamaan.

Pätevät menetelmät liittyvät niin tutkimusprosesseihin kuin lähdeaineistojen luotettavuuden todentamiseenkin. Pirjo Markkola mainitsee lähdekritiikistä kertovassa kirjoituksessaan (Markkola 2008, 169) käsityksen, että ”kaiken historianitutkimuksen perustalla ovat ensisijaisesti lähdekritiikki ja jatkuva kirjoittaminen”. Jotta Karjala-tietokannan luokiteltu aineisto olisi käyttökelpoista ja luotettavaa tutkimusaineistoa, täytyy mm. koodistoja käyttää luokituksissa johdonmukaisesti.

Tietojen luokitusten tulee pohjautua menetelmiin, jotka edesauttavat aineiston asianmukaisia tilastollisia käsittelyjä. Koodien käyttöä ja luokittelua tulisi valvoa, jotta välttyttäisiin mm. systemaattisilta luokitusvirheiltiltä. Mikkelin ja Pakkasvirran mukaan sairausluokitusten osalta tulisi huomioida esimerkiksi se, että lähteistä löytyvät erilaiset sairauskäsitykset eivät ole yleispäteviä vaan ne heijastavat tiettyä historiallista tulkintatapaa, joka ilmenee selvästi sairausluokitusten muuttumisena (Mikkeli & Pakkasvirta 2007, 129).

Vaatimus kohtuullisesta vastaustaakasta antaa Karjala-tietokannan aineiston luokituksissa mahdollisuuden käyttää harkinnan mukaista luokkaa, jos yksiselitteistä luokitteluperustetta ei löydy. Hyrkkänen (2008, 189) määrittelee historianitutkimuksen tärkeimmäksi lähteeksi historianitutkijan järjenkäytön. ”Lähteet eivät kerro, mikä on oikea tapa lukea niitä; lähteet eivät sano, mikä on tärkeintä ja oikeaa historiaa.” Markku Hyrkkäsen mukaan lähteet jättävätkin historioitsijan aina pulaan, sillä tutkija joutuu koko ajan luottamaan vain itseensä lähteisiin perustuvien päätelmien teossa. Näin myös luokituksissa.

Kustannustehokkuuden nimissä luokituksissa olisi pyrittävä koodauksen ja luokittelun automatisoinnin parantamiseen. Tehokkuusvaatimus heijastaa eräällä tavalla tietoyhteiskunnan sitä vaihetta, joka on nimetty tehokkuuden tietoyhteiskunnaksi. Tämän hetkistä vuosituhannen vaiheessa alkanutta vaihetta kutsutaan sen sijaan jakamisen tietoyhteiskunnaksi (Eduskunta 2009, 9). Karjala-tietokanta aineiston luokitteluvaatimuksissa nämä voidaan yhdistää: luokittelu tulee toteuttaa kustannustehokkaasti ja luokiteltua aineistoa tulee aktiivisesti tarjota tutkijoiden käyttöön.

TAULUKKO 16. Luokittelun tulosten laatuvaatimukset

Eurostatin periaate	Käytännösäännön laatuvaatimus	Karjala-tietokantasäätiön luokitusten laatuvaatimus
Relevanssi	Tilastojen käyttäjien tyytyväisyydestä sekä tulevista tarpeista kerätään tietoa säännöllisesti.	Tilastosovelluksen käyttäjiltä pyritään saamaan tietoa luokitusten hyödyllisyydestä ja muutostarpeista.
Tarkkuus ja luotettavuus	Tilastojen on kuvattava todellisuutta tarkasti ja luotettavasti.	Koodistojen sisältöjen tulisi olla riittävän tarkkoja ja luokitusten tulisi perustua luotettavasti Karjala-tietokannan aineistoon.
Ajantasaisuus ja oikea-aikaisuus	Tilastot on levitettävä ajantasaisina ja oikea-aikaisina.	Luokitusten päivitykset siirretään viipymättä tilastosovellukseen.
Yhtenäisyys ja vertailukelpoisuus	Yhteisön tilastojen tulee olla sisäisesti ja ajallisesti yhtenäisiä sekä alueiden ja maiden kesken vertailukelpoisia.	Ammattiluokituksissa käytetään kansainvälistä HISCO-luokitusta ja kuolinsyiden luokituksissa Suomen ensimmäistä virallista kuolemansyynimistöä v. 1935.
Saatavuus ja selkeys	Tilastojen esitystapa tukee niiden oikeata tulkintaa ja tarkoituksenmukaista vertailua.	Luokituksissa käytettävien koodistojen sisältö ja koodien määrä ovat tarkoituksenmukaisia.

Luokitusten relevanssivaatimuksen huomioimiseksi tilastosovelluksen käyttäjiltä kerätään tietoa luokitusten hyödyllisyydestä ja muutostarpeista sekä palautetta tilastosovelluksen kehittämistarpeista yleensäkin. Eduskunnan tulevaisuusvaliokunta esittää vuosien 2007–2011 yhteenvedossaan (Eduskunta2011, 13), että olemassa olevien tietovarantojen hyödyntämiseen tulisi pyrkiä tehokkaammin ja tietoteknologiaa tulisi hyödyntää luovemmin esimerkiksi kehittämällä sähköisiä palveluita.

Tarkkuuden ja luotettavuuden vaatimus on laatuvaatimuksista yksi tärkeimpiä. Koodistojen sisältöjen tulisi olla riittävän tarkkoja ja luokitusten tulisi perustua luotettavasti Karjala-tietokannan aineistoon. Luokittelussa, kuten historian tutkimuksessa yleensäkin, on se vaikeus, että lähdeaineisto ei niinkään koostu menneistä tapahtumista vaan niistä esitetyistä tulkinnoista ja väitteistä (Kalela 2002, 180).

Ajantasaisuuteen ja oikea-aikaisuuteen tietokanta-aineiston luokituksissa pyritään siten, että luokitusten päivitykset siirretään tilastosovellukseen viivytyksittä eli mahdollisimman nopeasti.

Luokitusten yhtenäisyys ja vertailukelpoisuus muihin vastaavien aineistojen luokituksiin toteutetaan noudattamalla jo olemassa olevia koodistoja ja luokittelumenetelmiä. Ammattiluokituksissa käytetään kansainvälisiä HISCO-koodeja ja kuolinsyiden luokituksissa Suomen ensimmäistä virallista kuolemansyynimistöä vuodelta 1935. Lisäksi epäselvien luokittelutapausten kohdalla tutkitaan muita luetteloita ja tulkintoja kyseisen aikakauden nimistöistä. Markkola muistuttaa, että epäilyttävien tutkimusaineistojen kohdalla on syytä tutkia tarkoin asiayhteyttä, lähdetekstin sisältöä, kieltä ja käsitteitä (Markkola 2008, 171).

Saatavuus ja selkeys luokittelussa saavutetaan mm. tarkoituksenmukaisilla koodistoilla ja koodien määrillä. Vaikka luokittelulla voidaan menettää osa alkuperäisen aineiston informaatiosta, niin toisaalta luokittelu yleensä lisää tulosten havainnollisuutta (Byckling & Salonen 2011). Valmiit luokitellut aineistot voivat kuitenkin tutkijan kannalta olla hankalia tutkimuskäytössä, jos tietoja ei löydy halutussa muodossa.

Pitkänen (2003, 2) huomauttaa, että väestötilastot ovat kokoelma valmiiksi laadittuja taulukoita ja vaikka niitä olisi tarjolla runsaastikin, tutkija ei aina välttämättä löydä tietoja halumassaan muodossa. Sen sijaan väestötilaston pohjana olevia rekistereitä käyttäessä hän ei ole sidottu tilastonlaatijan ratkaisuihin. Pitkänen toteaa valmiiksi taulukoidun aineiston käytön ongelmalliseksi myös silloin, kun halutaan tutkia sellaisten tekijöiden välisiä riippuvuuksia, joiden ajatellaan vallitsevan yksilötasolla.

7 PÄÄTÄNTÖ

Tilastosovelluksen tietokannan määrittämisen yhteydessä ilmeni, ettei Karjala-tietokannan KATIHA-henkilöhakuohjelman verkossa olevaa tietokantaa voi suoraan käyttää tilastosovelluksen tietokantana. Perussyitä löytyi kaksi. Ensiksikin verkossa olevasta tietokannasta puuttuu osa tietokantaan tallennetuista 1900-luvun kirkonkirjojen tiedoista, sillä henkilöhakuohjelman verkkoversio ei tietosuojasyistä johtuen sisäl-

lä syntyneitä, vihittyjä eikä muuttaneita vuosilta 1911–1950 (kuolleiden tiedot vastaavalta ajalta ovat julkisia, sillä sadan vuoden tietosuojarajoitukset eivät koske niitä). Toinen syy sille, miksei KATIHA-henkilöhakuohjelman tietokantaa voi sellaisenaan hyödyntää tilastosovelluksessa, on tietojen luokittelutarve ammattinimikkeiden ja kuolinsyiden osalta.

Tilastosovelluksen tietokantavaatimuksista huolimatta määräaikainen ohjelmoija Teppe Hytönen sai ohjelmoitua ammattiluokituksiin ja kuolinsyylokituksiin liittyvät tilastolliset kyselyt. Ohjelmoinnin testiaineistona käytettiin 2000-luvun alkupuolella työstettyä Lavansaaren seurakunnan tietokanta-aineistoa, joka sisältää tuolloin tehdyt ammatti- ja kuolinsyylokitukset. Ohjelmointityön edetessä todettiin, että kaikkien muidenkin tilastosovellukseen kuuluvien seurakuntien aineistojen ammattinimikkeet ja kuolinsyyt olisi normalisoitava ja luokiteltava tai entiset luokitukset olisi ainakin tarkistettava, jotta tilastosovellus olisi julkaisukelpoinen.

Ammattinimikkeiden luokittelua varten oli jo valmiina 29 776 ammattinimikkeen käsittely ja koodaus. Ammattinimikkeiden koodeja sisältävästä HISCOSU-tiedostosta siirrettiin koodaustiedot ohjelmallisesti uusiin tietokantatauluihin, mutta samalla epäiltiin aineistosta löytyvän vielä ammattinimikkeitä, joita ei ole luokiteltu. Asian varmistamiseksi suoritettiin ohjelmallinen laskelma, joka laski syntyneiden, kuolleiden, vihittyjen ja muuttaneiden tauluista löytyvien ”-2” (ei luokiteltu) -koodien lukumäärät yhteen. Vastaukseksi saatiin 732 981, mikä osoitti, että ammattien luokittelua täytyy jatkaa ja täydentää. Koska tilastosovelluksen tietokanta sisältää noin 3 660 000 henkilömerkintää, on vielä viidennes ammattinimikkeistä käsittelemättä. Luokittelemattomien ammattien lukumäärä ei kuitenkaan tarkoita, että niin monta eri ammattinimikettä täytyisi luokitella manuaalisesti erikseen. Luokittelemattomista nimikkeistä poimitaisiin luokiteltaviksi ainoastaan erilaiset kirjoitusasut, joita laskennassa ilmeni tietokannassa olevan 190 699 kappaletta. Näiden nimikkeiden kaksoiskappaleet voitaisiin myöhemmin luokitella ohjelmallisesti.

Kuolinsyiden osalta luokitus on pidemmällä. Siitä huolimatta on tarvetta luokituksen tarkistukselle, sillä 49 883 kuolinsyyn luokitus tehtiin kiireellisellä aikataululla elossyyskuun aikana. Tämän opinnäytetyön tekijä käytti tuolloin kuolinsyiden luokituksen yhteensä noin 120 tuntia eli noin 20 työpäivää, jos tehokasta työaikaa ajatellaan työpäivässä olevan noin 6 tuntia. Tosin luokitteluun ei käytännössä käytetty yhtäkään

työpäivää vaan luokittelu tehtiin työajan ulkopuolella kesäloman ja opintovapaan aikana. Jottei kuolinsyiden luokittelu jäisi vain yhden henkilön harkinnan varaan, täytyy koodiston käyttöä tarkistaa sekä arvioida yksittäisten kuolinsyiden luokitusperusteita.

Kun kaikki ammattinimikkeet on saatu normalisoitua ja luokiteltua sekä tarkastettua, voidaan luokitusten tulokset liittää tilastosovelluksen tietokannan tauluihin. Se on välttämätöntä myös tilastosovelluksen testausta varten. Testausvaiheessa voitaisiin olla yhteydessä tutkijoihin ja pyytää palautetta tilastosovelluksen toimivuudesta ja hyödyllisyydestä väestötietojen tutkimisessa.

Jo ennen varsinaista luokitellun tietokanta-aineiston testausta voisi esittää hypoteesin, että koodistojen osalta esiintyy muutospaineita. Koodistot kannattaisi nimittäin muuttaa hierarkkisiksi. Kansainvälisen HISCO-luokittelun lisäksi ammattien koodit voisi ryhmitellä myös esim. sosiaalista asemaa ilmaisevan ryhmittelyn mukaan, jota Kansallisarkiston tutkimusjohtaja Päivi Happonen on käyttänyt väitöskirjatutkimuksessaan. Vaihtoehtoisesti ryhmittely voitaisiin toteuttaa JHS 150 Ammatiluokitus –suosituksen mukaisesti.

Kuolinsyykoodistoon sisältyvät koodit voitaisiin luokitella ja ryhmitellä Tilastokeskuksen käytössä olevan kuolinsyyluokituksen mukaan 20 pääryhmään, jotka ovat nähtävissä esim. vuoden 2008 kuolinsyiden tilastoinnissa (liite 6). Tilastokeskuksen käyttämä koodisto pohjautuu WHO:n laatimaan kansainväliseen IDC-10 –tauti-luokitukseen, jonka päivitetty versio on saatavilla STAKESin koodistopalvelimella (ICD-10 2010). Vaihtoehtoisesti kuolinsyyt voitaisiin luokitella hierarkkisesti kuuteen pääluokkaan Pensolan, Ahosen ja Notkolan esityksen mukaisesti.

Kaiken kaikkiaan väestötietojen tilastointiin liittyvä aihe on osoittautunut yllättävän ajankohtaiseksi, sillä näinä päivinä on uutisoitu, että maailman väkiluku on marraskuun alkuun 2011 saavuttanut 7 miljardin rajan ja väkiluvun arvioidaan vuonna 2025 olevan jo 8 miljardia. Tilastotiedoille on siis käyttöä sekä väestön menneisyyden analysoinnissa että tulevaisuuden ennustamisessa.

LÄHTEET

Borg, Sami & Kuula, Arja 2007. Julkisrahoitteisen tutkimusdatan avoin saatavuus ja elinkaari. Valmisteluraportti OECD:n datasuosituksen toimeenpanomahdollisuuksista Suomessa. Yhteiskuntatieteellisen tietoarkiston julkaisuja 6, 2007. Tampere: Yhteiskuntatieteellinen tietoarkisto, Tampereen yliopisto.

Bowker, Geoffrey C. & Leigh Star, Susan 2002. *Sorting Things Out. Classification and Its Consequences*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Byckling, Timo & Salonen, Riku 2011. Johdatus tilastolliseen ajatteluun. Tilastokeskuksen verkkomateriaali. <http://www.stat.fi/tup/verkkokoulu/data/tt/index.html>. Päivitetty 2011. Luettu 7.3. 2011.

CSC 2011. IT Center for Science – Tieteen tietotekniikan keskus. CSC:n verkkosivut. <http://www.csc.fi>. Päivitetty 2011. Luettu 3.10.2011.

DDB 2011. Demografiska databasen. Umeå universitet. Yliopiston WWW-sivut. <http://www.ddb.umu.se>. Päivitetty 2011. Luettu 15.3.2011.

DDSS 2011. Temasidor: Dödsorsaker och sjukdomsnamn. DDSS Demografisk Databas Södra Sverige. WWW-sivusto. <http://www.ddss.nu>. Päivitetty 15.9.2011. Luettu 15.9.2011.

Eduskunta 2009. Tehokkuudesta jakamiseen – tietoyhteiskunnan uusi vaihe. Tulevaisuusvaliokunnan tietoyhteiskunta-akatemia 19.-20.11.2008. Toim. Pirkka Åman, Tuuli Kaskinen, Outi Kuittinen, Roope Mokka & Aleksi Neuvonen. Eduskunnan tulevaisuusvaliokunnan julkaisu 2/2009. Helsinki: Eduskunta. PDF-dokumentti. <http://www.eduskunta.fi>. Päivitetty 2009. Luettu 28.9.2011

Eduskunta 2011. 10 tulevaisuuden kipupistettä. Tulevaisuusvaliokunnan 2007-2011 yhteenvetoa ja arviointia. Toim. Sirkka Heinonen, Olli Hietanen, Juha Kaskinen, Jyrki Luukkanen, Paula Tiihonen & Marku Wilenius. Eduskunnan tulevaisuusvaliokunnan julkaisu 6/2011. Helsinki: Eduskunta. PDF-dokumentti. <http://www.eduskunta.fi>. Päivitetty 2011. Luettu 28.9.2011.

Forsius, Arno 2007a. Tautinimistöjen kansainvälinen kehittämistyö. WWW-dokumentti. <http://www.saunalahti.fi/arnoldus/tautini3.htm>. Päivitetty 26.1.2007. Luettu 22.9.2011.

Forsius, Arno 2007b. Tautinimistöjen kehitys Ruotsissa ja Suomessa vuosina 1749-1809 ja Suomessa sen jälkeen. WWW-dokumentti. <http://www.saunalahti.fi/arnoldus/tautini2.htm>. Päivitetty 26.1.2007. Luettu 22.9.2011.

Forsius, Arno 2010a. Kuolinsyyt vuosien 1749-1877 väkilukutauluissa. WWW-dokumentti. http://www.saunalahti.fi/arnoldus/kuol_syy.html. Päivitetty 22.8.2010. Luettu 20.8.2011.

Forsius, Arno 2010b. Tautien nimistöistä ja luokituksista. WWW-dokumentti. <http://www.saunalahti.fi/arnoldus/tautinim.htm>. Päivitetty 20.7.2010. Luettu 4.10.2011.

Forssell, Jarno 2010. Tulevaisuuden toivot. Virpi Lummaa: Biologin kohde löytyi läheltä. Artikkelisarja Tutustu nuoriin tutkijoihin. Tiede 10, 61-62.

Fougstedt, Gunnar & Hyppölä, Jorma & Purola, Tapani & Strömmer, Aarno (toim.) 1962. Multilingual demographic dictionary. Monikielinen väestötieteen sanakirja. Suomenkielinen laitos. Suomen tilastoseuran julkaisuja 1. Helsinki: Suomen tilastoseura.

Grönroos, Matti 2004. Johdatus tilastotieteeseen - Kuvailu, mallit ja päättely. Helsinki: Finn Lectura.

Haimi, Olavi 1987. Väestöntutkimus: Periaatteet ja menetelmät. Väestöntutkimuslaitoksen julkaisusarja D, nro 19/1987. Helsinki: Väestöntutkimuslaitos.

Happonen, Päivi 2004. Sukututkimuksen asiakirjaopas: väestöhistorialliset arkistolähteet. Mikkeli: Karjala-tietokantasäätiö.

Happonen, Päivi 2009. Kaksi todellisuutta? Kirkonkirjat ja henkikirjat Sortavalan kaupungin väestöllisen profiilin kuvaajina 1800-luvun alusta vuoteen 1940. Joensuun yliopisto. Yhteiskunta- ja aluetieteiden tiedekunta. Yhteiskuntatieteellisiä julkaisuja nro 96. Arkistolaitoksen toimituksia 8. Väitöskirja. Joensuu: Joensuun yliopisto.

Heikkilä, Tarja 2010. Tilastollinen tutkimus. Helsinki: Edita Publishing Oy.

Henkilötietolaki 1999. Suomen säädöskokoelma: Henkilötietolaki 22.4.1999/523. Oikeusministeriö. Helsinki: Edita Publishing Oy. Ajantasainen lainsäädäntö: Finlex-verkkosivusto. <http://www.finlex.fi/fi/laki/ajantasa/1999/19990523>. Päivitetty 2011. Luettu 3.10.2011.

Hirvensalo, Lauri 1966. Saksalais-suomalainen sanakirja. Deutsch-Finnisches Wörterbuch. Porvoo: Werner Söderström Osakeyhtiö.

HISCO, 2002. Historical International Standard Classification of Occupations. Toim. Marco H.D. van Leeuwen, Ineke Maas & Andrew Miles. Belgium, Leuven: Leuven University Press.

HISCO, 2011. History of Work Information System. International Institute of Social History. Verkkopalvelu. <http://historyofwork.iisg.nl/>. Päivitetty 2011. Luettu 15.3.2011.

Historismi 2006. Kuolinsyyt-luettelo. Toim. Arno Forsius & Harri Hirvelä. WWW-sivusto. http://www.saunalahti.fi/hirvela/historismi_sivut/deathsiivu.html. Päivitetty 16.7.2006. Luettu 20.8.2011.

Holopainen, Martti & Pulkkinen, Pekka 2008. Tilastolliset menetelmät. Helsinki: WSOY Oppimateriaalit Oy.

- Hyrkkänen, Markku 2008. Historian arvo ja viehätys. Teoksessa Tutkijan kirja. Toim. Kirsti Lempiäinen, Olli Löytty & Merja Kinnunen. Tampere: Vastapaino, 187-203.
- ICD-10 1999. Tautiluokitus ICD-10. Suomalainen 2. uudistettu painos Maailman Terveysjärjestön (WHO) luokituksesta ICD-10. Helsinki: Sosiaali- ja terveysalan tutkimus- ja kehittämiskeskus STAKESin Sosiaali- ja terveydenhuollon luokituskeskus.
- ICD-10 2010. ICD-10 on päivitetty: Muutosinformaatio terveydenhuollon ammattilaisille. Stakesin sivusto. http://sty.stakes.fi/FI/koodistopalvelu/ICD-10/icd10_2010.htm. Päivitetty 29.9.2010. Luettu 5.11.2011.
- Jalovaara, Marika & Koskinen, Seppo & Martelin, Tuija 2007. Demografiset mittaluvut. Teoksessa Suomen väestö. Helsinki: Gaudeamus, 23-40.
- JHS 2011a. Julkisen hallinnon suositus: JHS Koodistot ja luokitukset, luonnos. JUHTA Julkisen hallinnon tietohallinnon neuvottelukunta. Helsinki. PDF-dokumentti. <http://www.jhs-suositukset.fi>. Päivitetty 15.9.2011. Luettu 19.9.2011.
- JHS 2011b. Koodistopalvelun metatiedot, luonnos. JHS Koodistot ja luokitukset, liite 3. JUHTA Julkisen hallinnon tietohallinnon neuvottelukunta. Helsinki. PDF-dokumentti. <http://www.jhs-suositukset.fi>. Päivitetty 15.9.2011. Luettu 19.9.2011.
- JUHTA 2010. Julkisen hallinnon suositus JHS 159: ISO OID-yksilöintitunnuksen soveltaminen julkishallinnossa. JUHTA Julkisen hallinnon tietohallinnon neuvottelukunta. Helsinki. PDF-dokumentti. <http://www.jhs-suositukset.fi/web/guest/jhs/recommendations/159>. Päivitetty 17.6.2010. Luettu 27.10.2011.
- Kalela, Jorma 2002. Tutkijan vapaus historiantutkimuksessa. Teoksessa Tutkijan eettiset valinnat. Toim. Sakari Karjalainen, Veikko Launis, Risto Pelkonen & Juhani Pietarinen. Helsinki: Gaudeamus, 177-194.
- Kamensky, Mika 2008. Strateginen johtaminen – Menestyksen timantti. Talentum Media Oy. Kariston Kirjapaino, Hämeenlinna.
- Kanji, Gopal K. 2006. 100 Statistical Tests. Thousand Oaks, California: Sage Publications.
- Karjalainen, Leila 2010. Tilastotieteen perusteet. Ristiina: Pii-Kirjat.
- Karjala-tietokanta 1990. Koeprojektin loppuraportin lyhennelmä 3.1.1990. Pöytäkirjan liite, 6. pykälä, Karjala-tietokantatoimikunnan kokous 8.3.1990.
- Karjala-tietokantasäätiö 2011. Säätiön verkkosivut. <http://www.karjalatk.fi>. Päivitetty 19.1.2011. Luettu 14.3.2011.
- Karttunen, Hannu 2001. Datan käsittely. Helsinki: CSC – Tieteellinen laskenta Oy.
- Keckman-Koivuniemi, Hannele 2010. Menneisyys ja nykyisyys kohtaavat: luovutetun Karjalan kirkonkirjat tutkimuksen elektronisina lähteinä. Tietoarkisto 29 1/2010. Verkkolehti. <http://www.fsd.uta.fi/tietoarkistolehti/29/katiha.html>. Päivitetty 11.3.2010. Luettu 14.3.2011.

Kecskeméti, István 2009. Kansallisarkiston digitointihankkeen loppuraportti 1.5.2009-31.12.2009. Arkistolaitos, Helsinki. PDF-dokumentti. [http://www.arkisto.fi/uploads/Arkistolaitos/Tehtävät ja toiminta/Hankkeet/Digitointihankkeen loppuraportti.pdf](http://www.arkisto.fi/uploads/Arkistolaitos/Tehtävät_ja_toiminta/Hankkeet/Digitointihankkeen_loppuraportti.pdf). Päivitetty 11.6.2010. Luettu 14.3.2011.

Ketokivi, Mikko 2009. Tilastollinen päättely ja tieteellinen argumentointi. Helsinki: Gaudeamus.

Koskinen, Seppo & Martelin, Tuija 2007. Kuolleisuus. Teoksessa Suomen väestö. Toim. Seppo Koskinen, Tuija Martelin, Irma-Leena Notkola, Veijo Notkola, Kari Pitkänen, Marika Jalovaara, Elina Mäenpää, Anne Ruokolainen, Markku Ryyänen & Ismo Söderling. Helsinki: Gaudeamus, 169-238.

Koskinen, Seppo & Martelin, Tuija & Notkola, Irma-Leena & Notkola, Veijo & Pitkänen, Kari 2007. Väestö tutkimuksen kohteena. Teoksessa Suomen väestö. Helsinki: Gaudeamus, 15-21.

Korkiasaari, Jouni & Söderling, Ismo 2007. Muuttoliike. Teoksessa Suomen väestö. Helsinki: Gaudeamus, 239-270.

Lamberg, Marko 2008. Karjala-tietokannan tutkimukselliset mahdollisuudet. Sukutieto 3, 24-25.

Markkola, Pirjo 2008. Lähdekritiikki – niin hiljaista ettei sitä huomaa. Teoksessa Tutkijan kirja. Toim. Kirsti Lempiäinen, Olli Löytty & Merja Kinnunen. Tampere: Vastapaino, 168-177.

Mikkeli, Heikki & Pakkasvirta, Jussi 2007. Tieteiden välissä? Johdatus monitieteisyyteen, tieteidenvälisyyteen ja poikkitieteellisyteen. Helsinki: WSOY Oppimateriaalit Oy.

Nemnich, Philipp Andreas 1801. *Lexicon Nosologicum Polyglotton: Omnium morborum symptomatum vitiorumque naturae et affectionum propria nomina decem diversis linguis explicata continens*. Hamburg: Ex officina Conradi Müller. Digitoitu lääketieteellinen sanakirja. <http://books.google.fi>. Päivitetty 2011. Luettu 12.9.2011.

Nieminen, Mauri 1999. Väestötilastoja 250 vuotta. Katsaus väestötilaston historiaan vuosina 1749-1999. Helsinki: Tilastokeskus.

Nieminen, Mauri 2011. Johdatus väestötieteen perusteisiin. Tilastokeskuksen verkkomateriaali. <http://www.stat.fi/tup/verkkokoulu/data/vt/index.html>. Päivitetty 2011. Luettu 7.3. 2011.

Nummenmaa, Lauri 2009. Käyttätymistieteiden tilastolliset menetelmät. Helsinki: Tammi.

Pelo, June 2007. Medical dictionary. Genealogy Research by June Pello. WWW-dokumentti. <http://sydaby.eget.net/swe/genealogy.html>. Päivitetty 29.6.2007. Luettu 12.9.2011.

- Pensola, Tiina & Ahonen, Hilikka & Notkola, Veijo 2004. Ammatit ja kuolleisuus. Työllisten ja työttömien ammattiryhmittäinen kuolleisuus 1996-2000. Helsinki: Tilastokeskus.
- Pitkänen, Kari 1999. Kuolemansyyt. Teoksessa Suomen vuosisata. Toim. Kristiina Andreasson & Vesa Helin. Helsinki: Tilastokeskus, 20-25.
- Pitkänen, Kari 2003. Suomen väestökisterit ja väestötilastot. PDF-dokumentti. <http://www.valt.helsinki.fi/sosio/vaesto/pdf/pitkanen.pdf>. Päivitetty 4.3.2003. Luettu 7.3.2011.
- Pitkänen, Kari 2007. Sanasto. Teoksessa Suomen väestö. Helsinki: Gaudeamus, 338-350.
- Pitkänen, Kari & Jalovaara, Marika 2007. Perheet ja perheenmuodostus. Teoksessa Suomen väestö. Helsinki: Gaudeamus, 115-167.
- Pitkänen, Kari & Sihvonen Ari-Pekka, 2003. Johdatus väestötieteeseen. Helsingin yliopisto, Sosiologian laitos. Helsingin yliopiston Tutkimus- ja koulutuskeskus Palmenia. Verkkomateriaali. <http://www.valt.helsinki.fi/sosio/vaesto/>. Päivitetty 11.3.2003. Luettu 7.3.2011.
- Poliklinikka 2011. Terveystieto A-Ö. Lääketieteen sanasto. Terve.fi-verkkopalvelut. <http://www.poliklinikka.fi>. Päivitetty 2011. Luettu 12.9.2011.
- Pressat, Roland 1972. Väestö ja väestötiede. Population, suom. Anja Aaltonen. Helsinki: Otava.
- Ropponen, Jari 2008. Karjala-tietokanta internetissä. Monen vuoden odotus on nyt ohi. Sukutieto 3, 22-23.
- Ruokolainen, Anne & Notkola, Irma-Leena 2007. Hedelmällisyys. Teoksessa Suomen väestö. Helsinki: Gaudeamus, 77-113.
- Ryynänen, Markku & Notkola, Veijo 2007. Väestönkehitys ja sen ennustaminen. Teoksessa Suomen väestö. Helsinki: Gaudeamus, 271-298.
- Rääf, Leonhard Fredrik 1859. Ydre-Målet eller folkdialekten i Ydre Härad af Östergöthland. Ordbok samt förteckning på alla oregelbundna och starka verber, som i ydre begagnas, jemte gamla dopnamn. Digi-toitu sanakirja. <http://books.google.fi>. Päivitetty 2011. Luettu 12.9.2011.
- Salkind, Neil J. 2008. Statistics for people who (they think) hate statistics (3rd edition). Thousand Oaks, California: Sage Publications.
- Schmidt, Rudy 2011. A Glossary of Archaic Medical Terms, Diseases and Causes of Death. The Genealogist's Recourse for Interpreting Causes of Death. WWW-sivusto. <http://www.antiquusmorbus.com>. Päivitetty 2011. Luettu 15.9.2011.
- Suonttavaara Lappby 2010. Liten lektion i Meänkieli/Kvänska. Kompletterat från Inkeri Tuovinen 1929-1937. WWW-dokumentti. <http://www.suonttavaara.se/dokument/Likhet%20mellan%20spraken.html>. Päivitetty 30.4.2010. Luettu 16.9.2011.

SVT 1940. Suomen virallinen tilasto 6:92. Kuolemansyyt vuonna 1938. Helsinki: Tilastollinen päätoimisto. PDF-dokumentti. http://www.doria.fi/bitstream/handle/10024/67226/ksyyt92_1938.pdf. Päivitetty 7.3.2011. Luettu 15.9.2011.

SVT 2010. Suomen virallinen tilasto. Suomen tilastollinen vuosikirja 2010. Helsinki: Tilastokeskus.

Tietosuoja 2010. Tietosuoja ja tieteellinen tutkimus henkilötietolain kannalta. Tietosuojaavaltuutetun toimisto. PDF-dokumentti. <http://www.tietosuoja.fi/uploads/8evlt8qrbsw3ud.pdf>. Päivitetty 17.11.2010. Luettu 3.10.2011.

Tilastokeskus 2008. Euroopan tilastojen käytäntesäännöt. EU:n komissio 25.5.2005. Eurostat. PDF-dokumentti. http://www.tilastokeskus.fi/org/periaatteet/cop_esitys.pdf. Päivitetty 3.11.2008. Luettu 3.10.2011.

Tilastokeskus 2011. Tilastokeskuksen verkkosivut. <http://www.tilastokeskus.fi>. Päivitetty 2011. Luettu 27.4.2011.

Valtioneuvosto 1935. Kuolemansyynimistö: Asetus kuolleisuustilastoa varten toimitettavista tiedoista. Lääkintöhallituksen vahvistama kuolemansyynimistö. Helsinki: Valtioneuvosto.

VaskiVakka 2009. Sanasto – Kuolinsyyt. Toim. Helena Karppinen, Juha Takala & Päivi Aalto. WWW-dokumentti. http://vaskivakka.net/tietovakka/sanastot/yleinen/sanasto_kuolinsyyt.htm. Päivitetty 24.7.2009. Luettu 20.8.2011.

Warner, Rebecca M. 2008. Applied Statistics: from Bivariate Through Multivariate Techniques. Thousand Oaks, California: Sage Publications.

Wright, Daniel B. & London, Kamala 2009. First (and Second) Steps in Statistics. Thousand Oaks, California: Sage Publications.

YK 2007. Tilastotietojen luottamuksellisuus & yksikötason tietojen käyttö. Hyviä käytäntöjä koskevat periaatteet ja suuntaviivat. Euroopan tilastokonferenssi. YK:n Euroopan talouskomissio. Yhdistyneiden Kansakuntien julkaisu Sales No. E.07.II.E.7. New York & Geneve: United Nations. PDF-dokumentti. http://www.tilastokeskus.fi/org/periaatteet/ece_lopullinen.pdf. Päivitetty 2007. Luettu 3.10.2011.

Tilastojen käytännesäännöt (Tilastokeskus 2008), taulukointi: Satu Soivanen

Periaate	Käytännesäännön sisältö	Laatuvaatimus
Ammatillinen riippumattomuus	Tilastoviranomaisen riippumattomuus politiikasta, hallinnosta sekä yksityisen sektorin toimijoista.	Tilastotuotanto ja jakelu ovat riippumattomia ulkopuolisista eduista.
Tiedonkeruuvaltuudet	Tilastoviranomaisella selvät juridiset valtuudet kerätä tilastotarpeisiin tarkoitettuja tietoja.	Tarvittava lähdeaineisto on saatavissa.
Resurssien riittävyys	Tilastoviranomaisella on riittävästi henkilöstöä, rahoitusta ja teknisiä välineitä tehtäviensä suorittamiseksi.	Käytettävissä olevien resurssien on riitettävä täyttämään tilastoja koskevat vaatimukset.
Laatuun sitoutuminen	Tilastojen laatuohjeistus on dokumentoitu, julkisesti saatavilla ja tilastohenkilöstö tuntee sen.	Tilastojen tiedonkeruun, muokkauksen ja jakelun laadun seurannassa noudatetaan sovittuja menettelyjä.
Tilastosalaisuus	Tietojen antajien yksityisyys, tiedon salassapito ja tiedon käyttäminen ainoastaan tilastotarkoitukseen on taattava.	Tilastosalaisuuden suojaamisesta on annettu julkisesti nähtävillä olevat määräykset ja ohjeet.
Puolueettomuus ja objektiivisuus	Tilastoviranomaisten on tuotettava ja levitettävä tilastoja noudattaen objektiivisesta ja avointa tapaa, joka kohtelee kaikkia käyttäjiä tasavertaisina.	Tilastomenetelmien tiedot ovat julkisia ja julkaistuissa tilastoissa havaitut virheet korjataan ja julkistetaan mahdollisimman nopeasti.
Pätevät menetelmät	Menetelmät ovat kansainvälisten standardien, ohjeiden ja hyvien käytänteiden mukaisia.	Vakiokäsitteitä, määritelmiä ja luokituksia käytetään johdonmukaisesti.
Asianmukaiset tilastolliset menettelyt	Aineiston keruuseen ja käsittelyyn käytetään teknisiä apuvälineitä, joiden laatua arvioidaan ja parannetaan säännöllisesti.	Tiedonkeruuta, tiedon tallennusta ja koodausta valvotaan jatkuvasti ja korjataan tarpeen mukaan.
Kohtuullinen vastaustaakka	Euroopan tilastojen määrä ja yksityiskohtaisuus rajoitetaan vain kaikkein tarpeellisimpaan.	Jos tarkkoja tietoja ei ole saatavilla, hyväksytään estimaatit ja likiarvot.
Kustannustehokkuus	Tiedonkeruun, prosessoinnin ja jakelun tehokkuutta parannetaan tietotekniikan avulla.	Tiedonsyötön, koodauksen ja validoinnin automatisointia parannetaan ja seurataan.
Relevanssi	Käytössä on menettelyt, joilla tietojen käyttäjiltä saadaan palautetta, arvioidaan tilastojen relevanssia ja hyödyllisyyttä.	Tilastojen käyttäjien tyytyväisyydestä sekä tulevista tarpeista kerätään tietoa säännöllisesti.
Tarkkuus ja luotettavuus	Lähdetiedot ja tulokset arvioidaan, otanta- ja muut virheet mitataan ja dokumentoidaan järjestelmällisesti.	Tilastojen on kuvattava todellisuutta tarkasti ja luotettavasti.
Ajantasaisuus ja oikea-aikaisuus	Tilastojen ajantasaisuus täyttää sovitut standardit ja tilastot julkaistaan vakioaikaan.	Tilastot on levitettävä ajantasaisina ja oikea-aikaisina.
Yhtenäisyys ja vertailukelpoisuus	Tilastojen kattavuudessa, määritelmässä, yksiköissä ja luokituksissa noudatetaan yhtenäisiä standardeja.	Yhteisön tilastojen tulee olla sisäisesti ja ajallisesti yhtenäisiä sekä alueiden ja maiden kesken vertailukelpoisia.
Saatavuus ja selkeys	Tilastot olisi esitettävä tarkoituksenmukaisella tavalla ja niiden sekä niitä tukevan metatiedon ja ohjeiden tulisi olla tasapuolisesti saatavilla.	Tilastojen esitystapa tukee niiden oikeata tulkintaa ja tarkoituksenmukaista vertailua.

Ammattinimikkeiden HISCO-luokitus, suomennos: Karjala-tietokantasäätiö

01110 kemistit	13920 lastentarhan johtajat	22610 työnjohtajat
01330 geologit	13940 rehtorit, yliopettajat	22630 metallityönjohtajat
02000 insinöörit	13950 koulutarkastajat	22650 huoltoasemanhoitajat
02120 arkkitehdit	13960 kotiopettajat	22655 sähkötyönjohtajat
02210 siviili-insinöörit	13990 muut opettajat	22665 leipomon, meijerin johtajat
02220 rakennusinsinöörit	14120 kirkkoherrat, papit	22670 käsityönjohtajat
02240 rautatieinsinöörit	14130 lähetystyö	22675 rakennustyönjohtajat
02305 sähköinsinöörit	14140 diakoniatyö	22690 muut työnjohtajat
02420 koneinsinöörit	14190 saarnaajat	30000 asiamiehet, toimitsijat
02710 vuori-insinöörit	14990 kellonsoittajat	31000 arkistovirkailijat
02930 maanviljelysinsinöörit	15120 kirjailijat	31020 verokirjurit
03020 maanmittarit	15900 sanomalehtimiehet	31030 verotarkastajat
03110 piirtäjät	15920 lehtitoimittajat	31040 tullivirkailijat
03120 konepiirtäjät	15990 libristit	31090 voudit
03160 kartanpiirtäjät	16120 kuvanveistäjät	32120 sihteeri
03200 teknikit	16130 graafikit, taiteilijat	32140 konekirjoittajat
03310 rakennusteknikot	16150 taidekaivertajat	33110 kirjanpitäjät
03405 sähkötekniikit	16200 mainosnukkien valmistajat	33120 kauppa- ja kirjuri
03430 lennätintekniikit	16220 mainospiirtäjät	33130 kunnan kassanhoitajat
03510 konetekniikit	16240 huonekalupiirtäjät	33135 kassanhoitajat
03520 sahatekniikit	16310 valokuvaajat	33140 pankin kassanhoitajat
03530 lentokonetekniikit	16390 retuseeraajat	33160 metsäkassanhoitajat
03550 lämpötekniikit	17000 taiteilijat	33170 postivirkailijat
04125 lentäjät	17120 säveltäjät	33940 pankkivirkailijat
04200 pursimiehet	17135 kapellimestarit	36000 autonrahastajat
04215 merikapteenit	17140 muusikit	36020 junailijat, konduktöörit
04217 laivaliikenne	17145 laulajat	36030 makuuvaununhoitajat
04230 perämiehet	17150 kanttorit	36040 linja-auton rahastajat
04240 luotsit	17190 musiikinopiskelijat	36090 raitiovaunun konduktöörit
05320 agronomit	17320 näyttelijät	37000 postiapulaiset
05360 maa- ja metsätalous	17330 näyttämönjohtajat	37020 postikirjurit
05390 metsätiede	17340 elokuvateatterin hoito	37030 postinteknikot
05400 laborantit	17420 teatterin johtajat	37040 lähetit
06100 lääketiede, lääkärit	17590 sirkuslaiset	38000 puhelinvirkailijat
06105 lääkärit	17990 voimistelijat	38020 puhelinkeskuksenhoitajat
06110 kirurgia	18020 painijat	38040 lennätinvirkailijat
06190 haavurit, välskärät	18030 urheilujoukkueen johto	38090 viestimestarit
06210 lääketiede, apulaiset	18050 liikuntaneuvojat	39120 tavaranylähtäjät
06310 hammaslääketiede	18090 voimistelijat	39140 varastonhoitajat
06320 hammastekniikit	19260 historia ja kielitiede	39150 vaakamestarit
06510 eläinlääketiede	19290 filosofia	39190 sahakirjurit
06710 apteekki	19310 nuorisotyö ja kansanhuolto	39310 konttoristit, kirjurit
06810 apteekki, harjoittelijat	19320 sosiaalihuolto	39340 kärjäkirjurit, apulaiset
07100 sairaanhoitoharj.	19390 kasvattajat	39350 vakuutusvirkailijat
07110 sairaan-, terv.hoito	19430 opinto-ohjaajat	39390 muut sihteeri
07120 melisairaanhoitajat	19490 kerhoneuvojat	39530 kirjaajat
07130 terveydenhoitajat	19530 kielenkääntäjät	39940 oikolukijat
07190 asastonhoitajat	20110 virkamiehet, lautamiehet	39960 rautatievirkailijat
07210 apuhoitajat	20210 neuvokset, kunnan johto	39990 muut kirjurit
07310 kättilöt	21000 johtajat, päälliköt	41020 tukkukauppiat
07620 lääkintävoimistelijat	21110 johtajat, tehtailijat	41025 kauppiat
07640 hierojat	21220 johtajat, toimitusjohtajat	41030 liikkeenharjoittajat
07990 terveydenhoito	21230 pehtorit, tilanhoitajat	42200 ostajat
09010 taloustiede	21240 rakennusmestarit	42220 hankintamiehet
09020 ekonomit	21300 myymälänhoitajat	43120 tekniset johtajat
11010 kamreerit	21330 liikkeenjohtajat	43130 huoltoesimiehet
11020 tilintarkastajat	21340 johtajat, myyntipäälliköt	43200 edustajat
12000 juristit	21490 sot.kodin, ampumaradan hoito	43220 kauppa- ja matkustajat
12210 tuomarit	21940 kunnalliset johtajat	43230 kauppaedustajat
12310 notaarit	21950 talousjohtajat	44100 vakuutusmiehet
12410 asianajajat	21960 liikennepäälliköt	44120 henkivakuutusasiamiehet
13000 opettajat	21970 radioasemanhoitajat	44230 mainosmiehet
13020 koulunopettajat	21990 muut johtajat	45125 myyjät
13030 lehtorit	22000 tarkastajat	45130 kauppa-apulaiset
13100 professorit	22110 konttorinhoitajat	45140 muotiliikkeen apulaiset
13190 seminaarin opettajat	22190 isännöitsijät	45150 myyjät, esittelijät
13200 oppikoulun opettajat	22210 asemanhoitajat	45220 kaupustelijat
13215 kieltenopettajat	22220 postinhoitajat	51000 majoitusmestarit
13220 matematiikan opettajat	22230 aseman esimiehet	51020 kestäkievarit
13250 taideaineenopettajat	22240 vaunun tarkastajat	51030 ravintolanhoitajat
13260 kauppakoulun opettajat	22250 liikennetarkastajat	51040 majoittajat
13270 maatalouskoulun opettajat	22275 lennätintarkastajat	51050 kahvilanhoitajat
13275 talous-, käsityönopettajat	22290 muut tarkastajat	53100 kokit
13280 puu-, metallityönopettajat	22390 liike-, kassatarkastajat	53130 keittäjät
13290 toisen asteen opettajat	22420 emännöitsijät	53190 keittiöapulaiset
13320 kansakoulunopettajat	22425 taloudenhoitajat	53210 tarjoilijat
13390 pyhäkoulunopettajat	22430 emännät	53290 ravintola-apulaiset
13420 lastentarhanopettajat	22490 hovi-, muonitusmestarit	54010 apulaiset, palvelijat
13490 lastenkodin opettajat	22520 maatalouden työnjohtajat	54020 kotiapulaiset, piit

Ammattinimikkeiden HISCO-luokitus, suomennos: Karjala-tietokantasäätiö

54030 palvelijat, rengit	77510 meijeriköt	89210 savenvalajat
54035 lastenhoitajat	77540 juustomestarit	89242 kiveneistotehtailijat
54055 hotellin työntekijät	77610 leipurit	89300 lasimestarit
54090 muut palvelusapulaiset	77630 kondiittorit	89360 tiilenpolttajat
55100 stuerit	77690 karamellityöntekijät	90200 kumikorjaajat
55130 vahtimestarit, talonmiehet	77810 oluenpanijat	92110 kirjapainotyöläiset
55140 suntiot, lukkarit	77850 viinintekijät	92120 kirjapainon latojat
55190 näyttämönhoitajat	77890 limonadimestarit	92130 konelatojat
55220 siivoojat	78200 sikarintekijät	92200 painajat
55240 nuohoojat	79100 räätälit, vaatturit	92620 käsinsitojat
55290 puhdistajat, siivoojat	79200 turkistyyöntekijät	92625 kirjansitojat
56010 pesijät	79220 turkisompelijat	92690 sitomoapulaiset
56065 silittäjät	79310 hatuntekijät	93120 maalarit
57020 kampaajat	79320 modistit	93920 muut maalarit
57025 parturi-kampaajat	79400 leikkaajat	93960 automaalarit
57030 parturit	79500 vaattetustyöntekijät	93990 konemaalarit
57050 manikyristit	79510 ompelijat	94100 soitteknikot
57070 saunottajat	79620 verhoilijat	94150 harmonin tekijät
58110 palomiehet	79990 käsilaukkutyöntekijät	94160 urkujen tekijät
58220 poliisit	80110 jalkinetyöläiset, suutarit	94170 pianoteknikot
58230 etsivät	80310 nahkathyömiehet	94180 pianovirittäjät
58300 armeijan työntekijät	80320 satulasepät	94220 korintekijät
58320 upseerit	81000 puutyömiehet	94230 harjatyöläiset
58330 aliupseerit	81208 puunleikkaaja	95000 rakennustyömiehet
58340 sotilaat	81230 sorvarit	95120 muurarit
58930 vanginvartijat	81275 höylääjät	95130 uunisepät
58940 vartijat	81925 vaunumestarit	95210 betonimiehet
58990 järjestysmiehet	81940 mallinikkarit	95230 betoniraudoittajat
61110 talolliset	81945 veistäjät	95410 kirvesmiehet, nikkarit
61115 torpparit	81955 huonekalukuihlottajat	95415 rakennuspuusepät
61240 tallimestarit	82000 kivityömiehet	95420 konepuusepät
62105 piiat, rengit	83000 metallityömiehet	95450 veneenrakentajat
62110 muonamiehet	83110 sepät	95510 rapparit
62120 maatilain palkolliset	83190 erikoissepät	95910 rakentajat
62210 rehumeestarit	83220 sahasepät, viilaaajat	95950 vesijohtoasentajat
62400 karjanhoitajat	83320 metallisorvaajat	96150 turbiinihoitajat
62410 tallimiehet	83450 koneporaajat	96910 koneenhoitajat
62430 lampaanhoitajat	83490 pumpunhoitajat	96930 lämmittäjät
62460 hevosenhoitajat	83530 hiojat	96970 jäähdytyskoneasentajat
62490 kuoharit	83550 sahanteroittajat	96980 keskuslämmityksenhoitajat
62510 karjakot	83920 asesepät	96990 öljysäiliönhoitajat
62610 siipikarjanhoitajat	83930 lukkosepät	97120 satamatyöläiset
62700 puutarhityöläiset	83990 koneenkäyttäjät	97125 lastaajat
62740 puutarhurit	84100 asentajat	97130 kuormaajat
63110 tukkimiehet	84105 koneapulaiset	97145 varastomiehet, -työläiset
63120 uittopäälliköt	84190 konemekaanikot	97150 leipomon pakkaajat
63220 metsänhoitajat, -vartijat	84222 kellosepät	97152 pakkaajat
63230 metsänhoidon työläiset	84230 hienomekaanikot	97170 paalaajat
63240 mittamiehet	84320 autonasentajat	97190 kantajat
63290 tervanpolttajat	84410 lentokonemekaanikot	97400 tiettyömiehet
64100 kalastajat	84900 konepajan työntekijät	97415 rautatieläiset
64990 riistanvalvojat	84910 konemestarit	98100 moottorivienemiehet
71105 kaivosmiehet	84965 konttorikonemekaanikot	98120 merivartiomiehet
71220 kivenhakkaajat	84975 polkupyörämeekaanikot	98130 ylimatruusit
72000 rautatyömiehet	84980 vaununvoittelijat	98135 laivamiehet
72100 masuunimestarit	84985 katsastajat	98140 laivapoajat
72200 jyr sijät	84990 koneenpuhdistajat	98190 lossinhoitajat
72420 valurit	85110 sähköasentajat	98220 höyrylaivanlämmittäjät
72500 kaavaajat	85120 autosähköasentajat	98300 junamiehet
73100 vanerityöläiset	85345 käämijät	98320 veturinkuljettajat
73210 sahurit	85420 radiomekaanikot	98330 veturinämmittäjät
73290 hakkurikuljettimenhoitajat	85510 sähkötyömiehet	98420 jarrumiehet
73400 paperityöläiset	85600 puhelintyömiehet	98430 junanlähettäjä
74290 liimamestarit	85620 sähkölennätinasentajat	98440 vaihdemiehet
74400 tislajaajat	85630 lennätinmekaanikot	98500 ajomiehet
75135 karstaajat	85740 puhelinasentajat	98600 ajurit
75220 kehrääjät	86240 elokuvakoneenhoitajat	98620 hevostmiehet
75290 huopamestarit	87105 putkiasentajat	98920 telakkamestarit
75400 kutojat	87210 hitsaajat	98940 majakanvartijat
75440 konekutojat	87310 levysepät	98990 ratavartijat
75500 neulojat	87330 kuparisepät	99900 työläiset
75622 värjärit	87340 peltisepät	99910 työntekijät, sekatyömiehet
76145 karvarit, nahkurit	87350 kattilasepät	99920 tilapäistyöläiset, irtolaiset
77120 myllärit	88010 jalokivisepät	99930 tehdastyöläiset
77200 sokeritehtaan työläiset	88030 kiillottajat	99990 muut työläiset
77310 teurastajat	88050 hopeasepät, kultasepät	-1 ei ammattia
77340 makkaran tekijät	88080 kaivertajat	-2 ei luokiteltu
77390 teuraskarjan hankkijat	88090 korutyöläiset	-0 ei mainittu

Kuolemansyynimistö 1935 (Valtioneuvosto 1935), koodiston päivitys: Karjala-tietokantasäätiö

0001	kehitysvirheet	2450	tyymyksen taudit
0002	kuolleena syntyneet	2470	muut aineenvaihdunnan ja sisäerityksen taudit
0150	ennenaikainen syntymä, synnyttämisen heikkous	2480	keripukki ja muut puutostaudit
0250	lasten syntymävammat	2600	aivohalvaus, -verenvuoto, -kalvontulehdus
0300	vastasyntyneiden muut kuolemansyyt	2605	aivoveritulppa
0350	äidin tukahduttama	2610	märkäpesäke aivoissa
0400	heikkous, vaivaisuus	2615	märkivä aivokalvontulehdus
0450	hivutus-, näivetyystauti, kuihtuminen	2620	selkäydinkato
0500	vanhuus, vanhuuden heikkous	2630	epilepsia, kouristukset, krampit
0600	vanhuudenkuolio	2640	pikkulasten kouristukset
1000	lavantauti	2700	halvaava tylsistyminen
1010	pikkulavantauti	2750	muut mielisairaudet
1015	hiirilavantauti	2800	korvatulehdus ja muut korvataudit
1020	piikkukuume	2820	muut hermoston ja aistimien taudit
1030	toisintakuume	2900	kehitysvammaisuus
1035	undulaattikuume	3000	sydänpussin tulehdus
1050	malaria, vilutauti	3010	äkillinen sydämen sisäkalvontulehdus
1060	isorokko	3020	pitkällinen sydämen sisäkalvontulehdus
1070	vesirokko	3030	sydänlihaksen taudit
1080	tuhkarokko	3035	sydämen sepelvaltimoiden taudit
1090	tulirokko	3040	verisuonten kalkkeutuminen
1100	hinku-, tukehdusyskä	3043	aortan ja muut valtimon laajentumat
1110	kurkkumätä, kuristustauti	3045	kohonnut verenpaine
1125	influenssa	3055	laskimotulehdus, veritulppa
1130	sikotauti	3070	muut verenkiertoelinten taudit
1140	kolera	3100	pistokset
1150	äkillinen maha- ja suolitulehdus	3505	kurkunpään ja henkitorven tulehdus
1165	punatauti	3520	katarraalinen keuhkokuume
1190	keltatauti	3530	lohkokeuhkokuume
1200	äkillinen nivelleini	3532	äänijänteiden vesipöhö
1210	ruusu	3535	keuhkopussin tulehdus
1215	angiinamyrrytys, kitarisa-, kurkkutaudit	3545	keuhkopussin märkiminen
1220	yleinen verenmyrrytys	3555	keuhkokuolio, märkäpesäke keuhkoissa
1230	lapsihalvaus, polio	3570	keuhkoastma
1240	aivokuume, unitauti	3585	muut hengityselinten taudit
1250	kulkutaudinluonteinen aivokalvontulehdus	4005	maha- ja pohjukaissuolihaava
1260	vastasyntyneiden rakkoihottuma	4020	pitkällinen maha- ja suolitulehdus
1270	räkätauti	4030	lapsinäivetyt
1280	pernarutto	4040	umpilisäkkeen tulehdus
1290	vesikauhu	4050	tyrä, kohju
1300	jäykkäkouristus	4060	suolitukkeuma
1310	sädesieni-, hiivasienitauti	4070	muut maha- ja suolitautit
1345	kuume	4075	madot, suolistolaiset
1350	polttotauti	4100	äkillinen maksan surkastuminen
1400	keuhko-, kurkunpätuberkuloosi, keuhkotauti	4110	maksankovettuma
1410	aivokalvo- ja aivotuberkuloosi	4120	sappikivet, sappirakon tulehdus
1420	suoli- ja vatsakalvotuberkuloosi	4135	muut maksan ja sappiteiden taudit
1430	virtsa- ja sukupuolielinten tuberkuloosi	4200	haiman taudit
1440	luu- ja niveltuberkuloosi	4300	vatsakalvon tulehdus
1450	miliaarituberkuloosi	4500	muut ruoansulatuselinten taudit
1480	muiden elinten tuberkuloosi	4510	ripuli
1500	pahanlaatuinen imusolmuketauti	4600	sisäelinten sairaus
1510	spitaali	4700	sisäinen verenvuoto
1520	synnyttämisen kuppatauti	5000	äkillinen munuaistauti
1521	saatu kuppatauti	5010	pitkällinen munuaistauti
1530	nivustaivegranulooma	5020	munuais- ja rakkokivitauti
1570	muut matotaudit	5030	virtsaiteiden tulehdus
1580	muut loistaudit	5040	muut virtsaelinten taudit
1610	muut tartuntataudit	5210	eturauhasen taudit
2000	verinäivetyystauti, anemia	5215	muut miesten sukupuolielinten taudit
2010	leukemia	5230	munatorven ja munasarjan tulehdus
2020	pseudoleukemia	5240	muut naisten sukupuolielinten taudit
2030	verenvuototauti	5500	eteinen istukka
2040	punatäplätauti	5510	muut verenvuodot
2050	muut veren ja vertamuodostavien elinten taudit	5520	keskenmeno
2055	verenvuoto, -syöksy, repeämä	5550	emän repeämä
2060	verentungos	5560	muut synnytyshäiriöt
2300	kihti, luusärky	5600	raskaus- ja synnytyskouristustauti
2325	avitamiinoosit, vitamiinipuutokset	5610	synnytyksen jälk. yl. verenmyrrytys
2350	riisitauti	5615	lapsivuode-, kuume
2360	diabetes, vesitauti	5620	keskenmenon jälk. yl. verenmyrrytys
2370	aivolisäkkeen taudit	5630	emänulkopuolinen raskaus
2390	struuma, kilpirauhastaudit	5650	lapsivuoteisen veritulppa
2410	Addisonin tauti	5700	muut raskaustilan ja synnytystaudit
2430	lisäkilpirauhasen taudit	6000	luuytimen tulehdus

Kuolemansyyntimistö 1935 (Valtioneuvosto 1935), koodiston päivitys: Karjala-tietokantasäätiö

6010 pitkällinen niveltulehdus	8520 palovammat
6020 muut luuston ja nivelten taudit	8530 paleltuminen, kylmettyminen
6025 jäsentauti, jäsentulehdus	8540 auringonpisto
6030 kuolio, mätä	8550 sähköisku
6035 särky	8560 salamanisku
6100 hammastaudit	8570 ruhje- ja murtumavammat
6530 ihon ja ihonalaisen kudoksen taudit	8580 pisto-, leikkuu- ja iskuhaavat
7000 syöpä huulessa	8590 ampumahaava
7001 syöpä suun limakalvossa tai nielussa	8600 äkillinen myrkytys
7002 syöpä nenän limakalvossa	8602 myrkyllisten eläinten puremat
7004 syöpä kurkunpäässä	8604 ravintoaineiden aiheuttama myrkytys
7005 syöpä keuhkoissa	8606 vieraat esineet
7006 syöpä ruokatorvessa	8610 nälkä, uupumus
7010 syöpä mahalaukussa	8650 muut tapaturmat
7020 syöpä suolistossa	8700 hukuttautuminen
7022 syöpä peräsuolella	8710 hirttäytyminen
7023 syöpä mahasykirauhasessa	8720 ruhje- ja murtumavammat (itsemurha)
7025 syöpä muissa ruoansulatuselimissä	8730 pisto-, leikkuu- ja iskuhaava (itsemurha)
7026 syöpä virtsaelimissä	8740 ampumahaava (itsemurha)
7028 syöpä eturauhasessa	8750 myrkyttäytyminen
7030 syöpä emässä ja munasarjoissa	8790 muunlaiset itsemurhat
7035 syöpä muissa sukupuolielimissä	8800 lapsenmurha
7040 syöpä rintarauhasessa	8900 muut murhat ja tapot
7050 syöpä ihossa	8950 sodassa kuolleet
7060 syöpä muissa elimissä tai paikkaa ilmoitt.	8960 kuolleet julkisen viranomaisen toimesta
7500 sarkooma ihossa ja ihonalaisessa kudoksessa	9000 äkillinen kuolema
7510 sarkooma luissa	9002 tulehdus, turvotus, pöhö
7520 sarkooma sisäelimissä	9001 tiedottomuus, tunnottomuus
7550 sarkooma muissa elimissä tai paikkaa ilmoitt.	9005 väärä hoito
7700 muut kasvaimet	9010 tuntematon, ilmoittamaton
7800 lähemmin määrittelemättömät kasvaimet	9030 kuolleeksi julistettu
8000 pitkällinen alkoholisairaus	9040 kuolleen löydetty
8150 muut pitkälliset myrkytystaudit	9050 muu
8500 hukkuminen	tyhjä ei luokiteltu
8510 tukehtuminen	

Luokitusohjelma Kuol_amm.prg, dBase-ohjelmointi: Satu Soivanen

```

DECLARE tied[1,82]
tied[1,1]="0013Antr"   tied[1,2]="0061Harl"   tied[1,3]="0068Hein"   tied[1,4]="0092Hiit"   tied[1,5]="0109Ihan"
tied[1,6]="0116Ilme"   tied[1,7]="0119Impi"   tied[1,8]="0122Inke"   tied[1,9]="0126Jaak"   tied[1,10]="0132Joha"
tied[1,11]="0152Jääs"  tied[1,12]="0163Kann"  tied[1,13]="0182Kauk"  tied[1,14]="0206Kirv"  tied[1,15]="0212Kive"
tied[1,16]="0215Koiv"  tied[1,17]="0225Korp"  tied[1,18]="0245Kuol"  tied[1,19]="0252Kurk"  tied[1,20]="0263Käml"
tied[1,21]="0264Käki"  tied[1,22]="0285Lava"  tied[1,23]="0293Lepp"  tied[1,24]="0309Lumi"  tied[1,25]="0329Mets"
tied[1,26]="0339Muol"  tied[1,27]="0392Pets"  tied[1,28]="0427Pyhä"  tied[1,29]="0433Päk"   tied[1,30]="0449Raut"
tied[1,31]="0461Rusk"  tied[1,32]="0464Räis"  tied[1,33]="0470Sakk"  tied[1,34]="0471Salm"  tied[1,35]="0485Seis"
tied[1,36]="0498Soan"  tied[1,37]="0504Sort"  tied[1,38]="0507Sort"  tied[1,39]="0516Suoj"  tied[1,40]="0524Suur"
tied[1,41]="0527Säkk"  tied[1,42]="0549Teri"  tied[1,43]="0579Tytä"  tied[1,44]="0592Uusi"  tied[1,45]="0597Vahv"
tied[1,46]="0600Valk"  tied[1,47]="0617Vibo"  tied[1,48]="0618Viip"  tied[1,49]="0625Viip"  tied[1,50]="0627Viip"
tied[1,51]="0636Vuok"  tied[1,52]="0637Vran"  tied[1,53]="0660Äyrä"  tied[1,54]="1001Anna"  tied[1,55]="1015Kell"
tied[1,56]="1016Kel"   tied[1,57]="1018Kite"  tied[1,58]="1020Korp"  tied[1,59]="1022Kuok"  tied[1,60]="1024Kurk"
tied[1,61]="1025Kyyr"  tied[1,62]="1026Kaki"  tied[1,63]="1032Mant"  tied[1,64]="1033Matk"  tied[1,65]="1036Palk"
tied[1,66]="1037Pets"  tied[1,67]="1040Pitk"  tied[1,68]="1041Pälk"  tied[1,69]="1042Raiv"  tied[1,70]="1046Räis"
tied[1,71]="1047Sakk"  tied[1,72]="1048Salm"  tied[1,73]="1051Sort"  tied[1,74]="1052Suis"  tied[1,75]="1053Suoj"
tied[1,76]="1054Taip"  tied[1,77]="1056Teri"  tied[1,78]="1057Tiur"  tied[1,79]="1061Uusi"  tied[1,80]="1064Viip"
tied[1,81]="1065Viip"  tied[1,82]="1066Viip"
STORE 0 TO num
num=1
DO WHILE num < 83
  seur=tied[1,num]
  @ 13,9 SAY "Kuolleiden "+seur+" ammatit"
  tied1='KUOL'+right(left(seur,4),3)
  ammat='HISCOSU'
  polku='D:\0000data\koetilasto'
  SET DIRECTORY TO &polku
  SELECT 1
  IF .NOT. FILE (tied1+'.mdx')
    USE exclusive (tied1)
    INDEX ON henkilo TAG henkilo
    SET ORDER TO henkilo
  ELSE
    IF FILE (tied1+'.dbf')
      USE exclusive (tied1)
    ENDIF
  ENDIF
  SELECT 2
  IF .NOT. FILE(ammatt+'.mdx')
    USE exclusive (ammatt)
    INDEX ON amm TAG ammatti
    SET ORDER TO ammatti
  ELSE
    IF FILE (ammatt+'.dbf')
      USE exclusive (ammatt)
    ENDIF
  ENDIF
  STORE SPACE(5) TO t_ammry
  STORE SPACE(12) TO t_henkilo
  STORE SPACE(30) TO t_amm
  SELECT 1
  DO WHILE .NOT. EOF()
    t_amm=n_amm
    t_ammry="-2 "
    t_henkilo=henkilo
    IF t_amm <> SPACE(30)
      SELECT 2
      SET ORDER TO ammatti
      IF SEEK(t_amm)
        t_ammry=hisco2
      ENDIF
      SELECT 1
      SET ORDER TO henkilo
      SEEK(t_henkilo)
    ENDIF
    IF amm=SPACE(35)
      t_ammry="-0 "
    ENDIF
    REPL ammry WITH t_ammry
    SKIP
  ENDDO
  num=num+1
ENDDO

```

Luokitusohjelma Kuol_syy.prg, dBase-ohjelmointi: Satu Soivanen

```

DECLARE tied[1,82]
tied[1,1]="0013Antr"   tied[1,2]="0061Harl"   tied[1,3]="0068Hein"   tied[1,4]="0092Hiit"   tied[1,5]="0109Ihan"
tied[1,6]="0116Ilme"   tied[1,7]="0119Impi"   tied[1,8]="0122Inke"   tied[1,9]="0126Jaak"   tied[1,10]="0132Joha"
tied[1,11]="0152Jääs"  tied[1,12]="0163Kann"  tied[1,13]="0182Kauk"  tied[1,14]="0206Kirv"  tied[1,15]="0212Kive"
tied[1,16]="0215Koiv"  tied[1,17]="0225Korp"  tied[1,18]="0245Kuol"  tied[1,19]="0252Kurk"  tied[1,20]="0263Käml"
tied[1,21]="0264Käki"  tied[1,22]="0285Lava"  tied[1,23]="0293Lepp"  tied[1,24]="0309Lumi"  tied[1,25]="0329Mets"
tied[1,26]="0339Muol"  tied[1,27]="0392Pets"  tied[1,28]="0427Pyhä"  tied[1,29]="0433Pälk"  tied[1,30]="0449Raut"
tied[1,31]="0461Rusk"  tied[1,32]="0464Räis"  tied[1,33]="0470Sakk"  tied[1,34]="0471Salm"  tied[1,35]="0485Seis"
tied[1,36]="0498Soan"  tied[1,37]="0504Sort"  tied[1,38]="0507Sort"  tied[1,39]="0516Suoj"  tied[1,40]="0524Suur"
tied[1,41]="0527Säkk"  tied[1,42]="0549Teri"  tied[1,43]="0579Tytä"  tied[1,44]="0592Uusi"  tied[1,45]="0597Vahv"
tied[1,46]="0600Valk"  tied[1,47]="0617Vibo"  tied[1,48]="0618Viip"  tied[1,49]="0625Viip"  tied[1,50]="0627Viip"
tied[1,51]="0636Vuok"  tied[1,52]="0637Vran"  tied[1,53]="0660Äyrä"  tied[1,54]="1001Anna"  tied[1,55]="1015Kell"
tied[1,56]="1016Kell"  tied[1,57]="1018Kite"  tied[1,58]="1020Korp"  tied[1,59]="1022Kuok"  tied[1,60]="1024Kurk"
tied[1,61]="1025Kyyr"  tied[1,62]="1026Kaki"  tied[1,63]="1032Mant"  tied[1,64]="1033Matk"  tied[1,65]="1036Palk"
tied[1,66]="1037Pets"  tied[1,67]="1040Pitk"  tied[1,68]="1041Pälk"  tied[1,69]="1042Raiv"  tied[1,70]="1046Räis"
tied[1,71]="1047Sakk"  tied[1,72]="1048Salm"  tied[1,73]="1051Sort"  tied[1,74]="1052Suis"  tied[1,75]="1053Suoj"
tied[1,76]="1054Taip"  tied[1,77]="1056Teri"  tied[1,78]="1057Tur"   tied[1,79]="1061Uusi"  tied[1,80]="1064Viip"
tied[1,81]="1065Viip"  tied[1,82]="1066Viip"
STORE 0 TO num
num=1
DO WHILE num < 83
  seur=tied[1,num]
  @ 13,9 SAY "Kuolinsyiden "+seur+" koodit"
  tied1='KUOL'+right(left(seur,4),3)
  ksyyt='KSYYKOOD'
  polku='D:\0000data\koe\tilasto'
  SET DIRECTORY TO &polku
  SELECT 1
  IF .NOT. FILE (tied1+'.mdx')
    USE exclusive (tied1)
    INDEX ON henkilo TAG henkilo
    SET ORDER TO henkilo
  ELSE
    IF FILE (tied1+'.dbf')
      USE exclusive (tied1)
    ENDIF
  ENDIF
  SELECT 2
  IF .NOT. FILE (ksyyt+'.mdx')
    USE exclusive (ksyyt)
    INDEX ON ksyy TAG kuolinsyy
    SET ORDER TO kuolinsyy
  ELSE
    IF FILE (ksyyt+'.dbf')
      USE exclusive (ksyyt)
    ENDIF
  ENDIF
  STORE SPACE(4) TO t_ksyyry
  STORE SPACE(12) TO t_henkilo
  STORE SPACE(60) TO t_ksyy
  SELECT 1
  DO WHILE .NOT. EOF()
    t_ksyy=ksyy
    t_ksyyry="9010"
    t_henkilo=henkilo
    IF t_ksyy <> SPACE(60)
      SELECT 2
      SET ORDER TO kuolinsyy
      IF SEEK(t_ksyy)
        t_ksyyry=ksyyry
      ENDIF
      SELECT 1
      SET ORDER TO henkilo
      SEEK(t_henkilo)
    ENDIF
    REPL ksyyry WITH t_ksyyry
    SKIP
  ENDDO
  num=num+1
ENDDO

```

Esimerkki kuolinsyiden tilastoinnista (SVT 2010, 530), Tilastokeskus

Kuolleet kuolemansyyn mukaan, 2008				
Kuolemansyy, Nro (ICD-10)		Yht.	Miehet	Naiset
I	Tartunta- ja loistaudit	437	187	250
	tuberkuloosi (A15-A19, B90)	60	25	35
	immuunikato (HIV-tauti) disease (B20-B24)	7	7	-
II	Kasvaimet	11 214	5 782	5 432
	mahalaukun pahanlaatuiset kasvaimet (C16)	525	285	240
	ohut-, paksu- ja peräsuolen pahanlaatuiset kasvaimet (C17-C21)	1 129	569	560
	henkitorven, keuhkoputken ja keuhkojen pahanlaatuiset kasvaimet (C33-C34)	1 998	1 386	612
	nisän pahanlaatuiset kasvaimet (C50)	827	2	825
	eturauhasen pahanlaatuiset kasvaimet (C61)	811	811	-
III	Vertamuodostavien elinten ja veren sairaudet	50	19	31
IV	Umpieritys-, ravitsemus- ja aineenvaihduntasair.	667	319	348
	sokeritauti (E10-E14)	540	269	271
V	Mielenterveyden ja käyttäytymisen häiriöt	2 369	755	1 614
VI-VIII	Hermoston ja aistimien sairaudet	4 082	1 537	2 545
IX	Verenkiertoelinten sairaudet	20 281	9 623	10 658
	verenpainetaudit (I10-I15)	768	257	511
	iskeemiset sydänsairaudet (I20-I25)	11 761	5 913	5 848
	muut sydänsairaudet (I30-I52)	2 121	1 037	1 084
	aivoverisuonien sairaudet (I60-I69)	4 246	1 707	2 539
X	Hengityselinten sairaudet	1 980	1 202	778
	influenssa (J10-J11)	13	5	8
	keuhkokuume – lunginflammation – pneumonia (J12-J18)	457	227	230
	keuhkoputkentulehdus ja keuhkolaajentuma (J40-J44)	1 069	751	318
	astma – asthma (J45-J46)	105	27	78
XI	Ruoansulatuselinten sairaudet	2 460	1 433	1 027
	mahahaava ja pohjukaissuolihaava (K25-K26)	167	84	83
	krooniset maksasairaudet (K70, K73-K74)	1 219	900	319
XII	Ihon ja ihonalaiskudoksen sairaudet	36	12	24
XIII	Tuki- ja liikuntaelinten sairaudet	265	77	188
XIV	Virtsa- ja sukupuolielinten sairaudet	402	148	254
	munuaisen ja virtsajohtimien sairaudet (N00-N29)	309	106	203
XV	Raskaus, synnytys ja lapsivuoteus	5	-	5
XVI	Eräät perinataaliaikana alkaneet tilat	74	45	29
XVII	Synnynnäiset epämuodostumat, kromosomipoikkeavuudet	175	97	78
XVIII	Oireet ja epätäydellisesti määritetyt tilat	170	111	59
XX	Tapaturmat ja väkivalta	4 277	3 012	1 265
	maakuljetustapaturmat (V01-V89)	347	261	86
	muut tapaturmat (V90-X59)	2 599	1 736	863
	itsemurhat (X60-X84, Y87.0)	1 033	801	232
	muut ulkoiset syyt (X85-Y86, Y87.1-Y89)	298	214	84
	Ei kuolintodistusta	146	92	54
	Yhteensä	49 090	24 451	24 639
	Tauteihin kuolleet	44 667	21 347	23 320