

# Rekisteriaineistot nuorten aseman tutkimuksessa

Ulla Hämäläinen

Erilaiset aineistot yhteiskuntatieteissä

- Esimerkkeinä lastensuojelu- sekä lapsi- ja nuorisotutkimus

ReTkin & FSD:n ja Nuorisotutkimusverkoston yhteisseminaari

31.10.2008

## Tässä esityksessä vain rekisteriaineistojen käytöstä

- Yleistä nuoria koskevasta tutkimuksesta
- Mitä rekistereitä ja missä?
- Lupamenettelyt ja prosessi
- Yleisiä huomioita rekistereistä ja niiden käytöstä

## Yleistä nuoria koskevasta tutkimuksesta

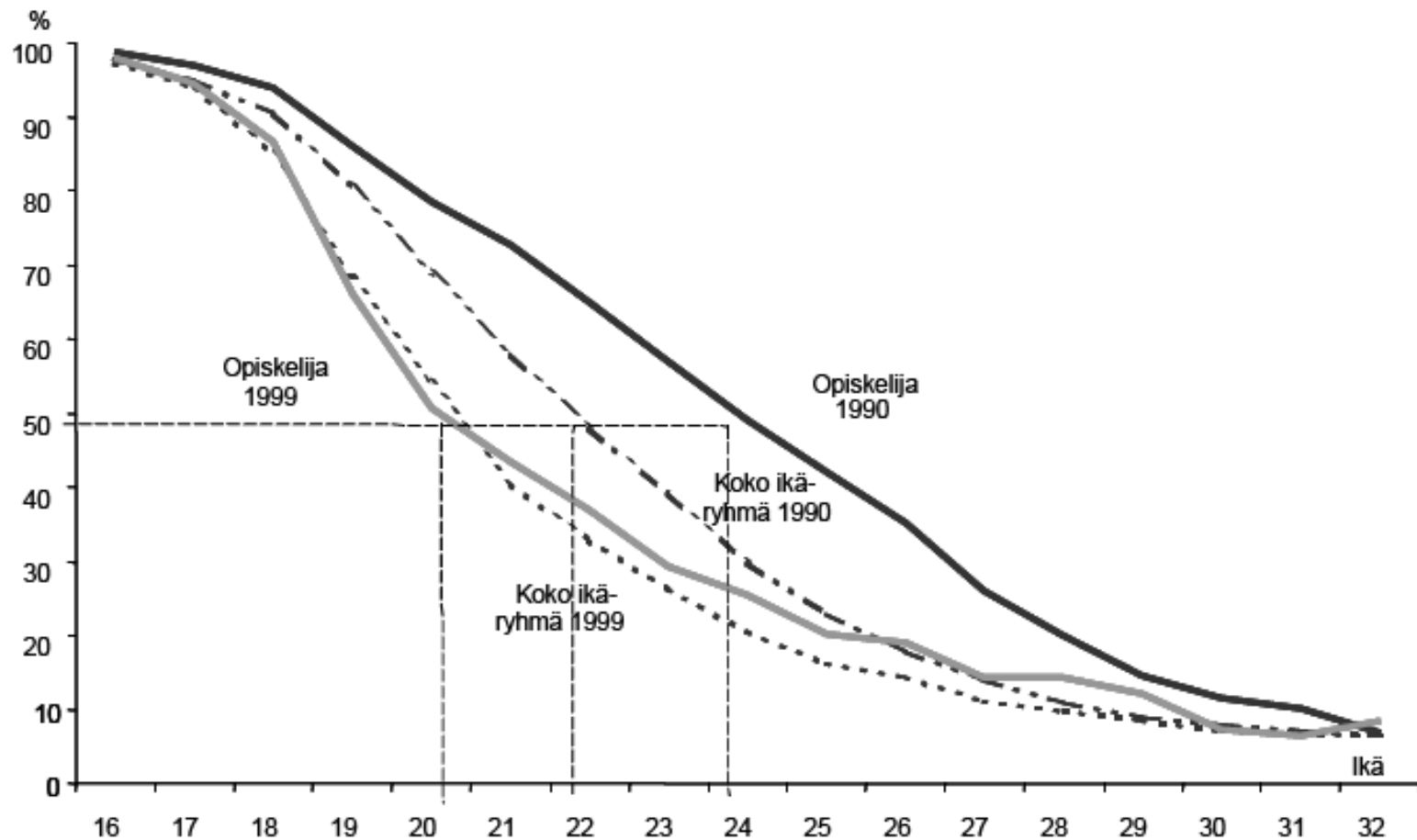
- Nuoret ovat väestöryhmistä heterogeenisin (asumismuotojen, pääasiallisen toiminnan, jne. suhteen) mikä asettaa tutkimukselle haasteita
- Yhden poikkileikkaustilanteen tulkitsemiseen liittyy yleensä riski. Nuorten asema muuttuu useita kertoja vuodessa (työllinen, työtön, opiskelija, perhe-vapaalla, muuten ei-aktiivinen) ja osa näistä tiloista on päällekkäisiä.

- Nuorten kohdalla on tärkeä ymmärtää mm. valmiiksi luotujen muuttujien päättelyjärjestyksiä: aikuisten kohdalla vähemmän ongelmallinen
  - esim. pääasiallinen toiminta perustuu päättelyyn: sen luomiseen on tarvittu tietoja henkilön iästä, työsuhteista, työttömyydestä, opiskelusta, eläkkeen saamisesta, jne., sekä priorisointia näiden tietojen välillä
  - Sosioekonominen asema vastaavasti
- Päättelyjärjestys TULEE TIETÄÄ JA YMMÄRTÄÄ: monet tutkimukset esim. ottavat pääasiallisen toiminnan annettuna, vaikka se vaikuttaa merkittävästi tulosten tulkintaan
  - Esim. yliopisto-opiskelijoista työllisiä vuoden lopussa >60 %: ovatko he opiskelijoita vai työllisiä?
  - Opiskelijoiden määrä on kasvanut valtavasti: vaikuttaa useisiin eri ilmiöihin mitä harvoin edes ajatellaan (epätyypilliset työsuhteet, pienituloisuus, jne.)

## Lakimuutokset ja rekisteröintikäytäntöjen muutokset

- Usein näkee väitettävän, että jokin asia on muuttunut merkittävästi (jonkun ryhmän pienituloisuus, asumismuodot, pääasiallinen toiminta, jne.)
- Kun tällaisen tuloksen saa aineistostaan, kannattaa aina ensimmäisenä miettiä onko muutoksen taustalla lakimuutoksia ja/tai rekisteröintikäytäntöjen muutoksia

## Vanhempiensa taloudessa asuvien nuorten osuus ikäryhmittäin 1990 ja 1999



## Mitä rekistereitä ja missä?

- Yksittäisiä rekistereitä on kymmeniä.
- Hyvän kuvan saa aloittamalla Tilastokeskuksen työssäkäyntitilaston pitkäikäistiedostosta 1987=> : sisältää yhdessä paikassa tietoa mm. Väestörekisterikeskuksen, Eläketurvakeskuksen, Kelan, Stakesin, Verohallituksen, Työministeriön ja Tilastokeskuksen tutkinto-, opiskelija-, yritys- ja toimipaikkarekisteri tietoja
- Tutkimuksessa tarvittava otos ja muuttujat määrittelevät sen mistä lähdetään liikkeelle ja mitä tietoja yhdistetään. Kaikkein ei tarvita Tilastokeskusta.

- Rekistereissä voi ajatella olevan tietoja niistä asioista, joista viranomaiset tarvitsevat tietoa päätöksentekoprosessissaan
  - Rekistereihin ei saa kerätä tietoa, joka ei ole tarpeellista ko. prosessissa, vaikka tieto olisi tutkijalle arvokasta
- Esim. sosiaaliturvassa prosessi tuottaa paljon tietoa eri vaiheissa, mutta kaikkea tietoa ei voida viedä ns. tilastointitietokantoihin
  - ⇒ Kaikkea mitä kysytään esim. Kelan lomakkeella ei välttämättä saada tutkimuskäyttöön helposti
  - ⇒ Jo tilastointitietokantoihin on tehty paljon valintaa



## Lupamenettelyt

- Kaikki rekisteritiedon käyttö tutkimuksessa perustuu yksilöityyn tutkimussuunnitelmaan ja sen perusteella rekisteriviranomaisten myöntämään käyttölupaan
- Tilastokeskuksella ja Stakesilla valmis lomake
- Kelasta vapaamuotoisella hakemuksella, joka sisältää 1) mitä tietoja tarvitaan, 2) käyttö-tarkoitus, 3) tiedon suojauskäytäntöjen kuvaus, 4) tutkimussuunnitelma

- VINKKI: mitä tietoja tarvitaan = pyri tarkkaan yksilöityyn, harkittuun kuvaukseen →  
nopeuttaa prosessin etenemistä
- Ole yhteydessä rekisterinpitäjiin varhaisessa vaiheessa: he tulkitsevat mitä oikeasti haluat 😊
  - esim. henkilötietokanta, päätöstietokanta, maksunjakotietokanta, maksatustietokanta, seurantatietokanta, fraasitietokanta, verotustietokanta
- Ennakoi lupaprosessin vaikeimmat osatekijät: Uusia, tietosuojallisesti hankalia muuttujia? Millainen kokonaisuus muodostuu, jne.

- Luovaa ongelmanratkaisua tietosuojakysymyksissä: laadi huolellisesti perusteltu esitys uusien, tietosuojallisesti hankalien muuttujien kohdalla ja pyri itse tarjoamaan ratkaisu siihen, miten tarvitsemasi tieto voitaisiin saada käyttöön
- Esimerkkejä salatuista tiedoista esim. kunta tai oppilaitos, jotka sellaisenaan yhdessä muiden muuttujien kanssa identifioisivat henkilöitä

## Otanta

- Suunnittele otos huolellisesti. Virheet ovat kalliita ja hidastavat asioita.
- Pyri yksinkertaiseen otokseen millä pystyt vastaamaan ko. kysymykseesi
  - X % vuosina xxxx-xxxx syntyneistä
  - X % tutkinnon vuosina xxxx-xxxx suorittaneista
  - x % työttömiksi tulleista (varovaisuutta...)
- Varo rajauksia, joita ei voi myöhemmin korjata

- Ihmisten elämä on rekistereissäkin monimutkaista
  - Useita jaksoja, takaisinperintää, rekisterivirheitä
  - Kohdentumisessa erilaisia ajankohtia: jaksot jakautuvat useille vuosille, eri tietolähteet eivät kuvaa tarkalleen samaa ajanhetkeä:
    - vuosidataa,
    - päivädataa,
    - jaksodataa,
    - vuoden viimeisen viikon poikkileikkaus-dataa, jne.

- Esimerkiksi Työssäkäyntitilaston pääperiaatteena on, että
  - tiedot ovat vuoden lopun tietoja: ikä, koulutus, pääasiallinen toiminta, työnantajan tiedot, työttömyysjakson tiedot, jne.
  - lisäksi kertymätietoja vuoden ajalta: eri tulolajit verotuksesta, työllisyyskk, työttömyyskk, työttömyysjaksot
  - myös päivämäärätietoja vuoden varrelta saatavilla: työttömyysjaksoista, työllisyydestä, jne.

## Hyvää suomalaisissa rekistereissä...

- Rekisterit ovat laadukkaita!
- Huolimatta alkuperäisestä tarkoituksesta (hallinnollinen käyttö) soveltuvat hyvin sekä nopeiden selvitysten että syvällisten tutkimusten tekoon
- Vaativat kuitenkin tutkijalta alkuperäisen aineiston rajoitteiden ymmärtämistä

## Hyvää (jatkuu...)

- Pitkiä paneeleja löytyy jo => avaa aivan uusia mahdollisuuksia tutkimusasetelmien tekoon
- Pitkät, luotettavat paneeliominaisuudet (vrt. survey-aineistojen kykyyn koota pvm-tiedot, tulotiedot, jne.)
- Tietoja saman vuoden sisällä eri tavoilla kerättynä
- Paranee vain vanhetessaan (tuo meille myös otosteknisiä ongelmia)
- otospoiminta periaatteessa helppoa ja kato vähäinen
- rekisterit ovat osittain päällekkäisiä = sama tieto useammasta lähteestä



## Tutkijoiden huomioitava

- Metatiedot usein puutteellisia: metatiedoissa paljon hiljaista tietoa, joka katoaa ihmisten mukana.
- Metatietojen puutteellisuus erityisen suuri ongelma pitkittäistarkasteluissa: todella paljon on käyttäjän vastuulla ja käyttäjän on vastuunsa ymmärrettävä ja kannettava!
  - Lakimuutokset: suoraan aiheeseen liittyvät ja välilliset
  - Rekisteröintikäytäntöjen muutokset

## Tutkijoiden huomioitava...

- Rekisteriviive on merkittävä ajantasaisen tutkimuksen haitake
- Henkilötilastot ovat monimutkaisia ja täysin uudet aineistot virhealttiita: aloita suunnittelu ajoissa ja varaudu virheisiin aineistossa.

## Millaista on tutkimus rekisteripohjaisilla mikroaineistoilla?

- Kaunis kuva:  
hypoteesi => testaus => tulos
- Todellisuus: Aineisto ja tutkija keskustelevat keskenään.
  - Prosessi on kaikkea muuta kuin yksioikoinen: Tutkijan alkuperäinen näkemys on ollut väärä; aineisto ei taivukaan teknisistä syistä vastaamaan kysymykseen juuri sillä tavalla kuin alun perin suunnittelit; muuttujat ovat muuta kuin mitä sen nimi sanoo; menetelmällisiä yllätyksiä, jne. jne.

## Nuoria koskevassa tutkimuksessa

- Tarvitaan ehdottomasti rekistereiden lisäksi myös kyselyaineistoja. Näkökulmat täydentävät toisiaan.