

Väestötutkimusten tilastolliset haasteet

Tommi Härkönen
Terveysosasto



TERVEYDEN JA HYVINVOINNIN LAITOS



Sisältö

- Terveyden ja hyvinvoinnin laitos (THL)
- Väestötutkimukset
 - Väestötutkimusten tavoitteet
 - Väestötutkimuksen toteutus
 - Tilastotieteeseen liittyvät tehtävät väestötutkimuksissa
- Ennusteet
- Tilastolliset ohjelmistot

LUOTETTAVAA TIETOA

sosiaali- ja terveysalan
päätöksenteon ja
toiminnan tueksi



TERVEYDEN JA HYVINVOINNIN LAITOS

MITÄ THL TEKEE?

- THL tutkii **väestön terveyttä ja hyvinvointia**, näihin liittyvän politiikan ja palvelujen vaikuttavuutta sekä ympäristöterveyttä ja yhteiskunnallisia ongelmia. THL edustaa Suomen tutkimusorganisaatioiden huippua.
 - THL **kerää ja tuottaa tietoa** terveydenhuollon ja hyvinvoinnin tilasta, sosiaali- ja terveystaloudesta sekä sairauksista ja niiden hoidosta.
-

MITÄ THL TEKEE?

- THL kehittää ja ohjaa sosiaali- ja terveydenhuoltoa ja antaa asiantuntijatukea päättäjille ja palvelujen tuottajille kotimaassa ja ulkomailla. Asiantuntijatyön perustana on laadukas **tutkimus, tilasto- ja rekisteritieto** sekä sidosryhmiemme kartoitetut tarpeet.
 - THL ohjaa ja tuottaa valtion sosiaali- ja terveystalouden palveluja ja vastaa alaisestaan viranomaistoiminnasta.
-

THL:N STRATEGIA



TERVEYDEN JA HYVINVOINNIN LAITOS

PRIORISOINNIT

- Hyvinvointiyhteiskunnan kestävyys

- **Eriarvoisuus ja syrjäytyminen**

- **Muuttuva sairauksien kirjo**

- Varautuminen terveysuhkiin

- Palvelujärjestelmän muutos

THL VAIKUTTAA

- **Korkeatasoisella tutkimuksella**
 - käynnissä n. 100 EU:n, Akatemian ja VNK:n rahoittamaa tutkimusta
 - V. 2014 julkaistiin noin 800 tieteellistä artikkelia
 - **Tietovarantojen ylläpitämisellä ja hyödyntämisellä**
 - aiempaa monipuolisempia tietovarantoja asiakkaille
 - **Sosiaali- ja terveydenhuollon kehittämisellä ja ohjauksella**
 - menetelmiä, työkaluja ja tietoa päättäjille ja ammattilaisille
 - v. 2014 asiantuntijamme olivat 81 kertaa kuultavina eduskunnassa
-

THL:N HENKILÖSTÖ 1.1.2015

73/27 %

naisia/miehiä

47 v.

keski-ikä

1 100

monialaista ja
korkeasti
koulutettua
asiantuntijaa

26 %

tutkija-
koulutuksen
saaneita

24

kansallisuutta

65 v. 10 kk

keskimääräinen eläkkeellesiirtymisikä



TERVEYDEN JA HYVINVOINNIN LAITOS

Väestötutkimukset

Tommi Härkönen
Terveysosasto



TERVEYDEN JA HYVINVOINNIN LAITOS



Mikä on väestötutkimus?

- ”Väestöstä kerättäviin tietoihin perustuva tutkimus”
- Tavoitteena usein väestöä edustavat tulokset esim. sairauksien tai elämäntapojen yleisyydestä

Väestötutkimusten tavoitteet

- Tärkeimpiä tavoitteita ovat
 - Tieteellinen tutkimus (esim. sairauksia ennustavat tekijät) ja
 - Poliittisen päätöksenteon tuki (usein erilaiset väestöjakaumat ja ennusteet)
- Tilastokeskus, Väestörekisterikeskus, Kela, Eläketurvakeskus ym. julkaisevat **rekistereihin** perustuvia tilastoja Suomen väestöstä, taloudesta ym.
- Rekistereistä ei kuitenkaan saada tietoja esim. elämäntavoista
 - tarvitaan yksilötason tutkimustietoja, joita kerätään
 - terveystarkastustutkimuksilla ja
 - kysely- tai haastattelututkimuksilla

Väestötutkimuksen toteutus

1. Väestöstä l. **perusjoukosta** poimitaan **otos**
2. Otokseen poimitut tutkimusyksilöt **kutsutaan** tutkimukseen
3. Tutkittavilta pyydetään **suostumus**, ja suoritetaan **mittaukset**, haastattelut ja kyselyt
4. Kenttätutkimustietoja **täydennetään** rekisteriseurantatiedoilla
5. Kerättyä aineistoa **analysoidaan** erilaisilla menetelmillä riippuen mm. tutkimusasetelmasta
 - Poikkileikkausasetelma
 - Kohorttiasetelma
 - Toistomittausasetelma
 - Erilaiset tapaus-verrokkiasetelmat
6. Tuloksia **raportoidaan** mm.
 - Tieteellisissä julkaisuissa (vertaisarviointi)
 - Kotimaisissa ja ulkomaisissa julkaisusarjoissa
 - (Sähköisinä) tilastoina

Tilastotiede väestötutkimuksissa

1. **Otannan** suunnittelu ja toteutus
 - Tutkimuksen tyyppi: terveystarkastus vai kysely/haastattelu?
 - Kustannusten minimointi vs. tulosten tarkkuus
2. Kaikki otokseen poimitut eivät halua osallistua tutkimukseen
→ Haasteena **kadon** vaikutusten korjaaminen
 - Minkälaisia tietoja otoksesta on käytettävissä?
3. Tilastollisten **analyysien** suunnittelu ja toteutus
 - Peruseräraportointi (suuri määrä samanlaisia perusanalyyskejä)
 - Tieteellinen raportti (mahdollisesti räätälöityjä menetelmiä)
 - Ennusteet
4. Tilastollisten **menetelmien kuvaus** raportoinnissa

Väestötutkimusten haasteita: kato

Finriski-tutkimukset 1982-2012, 25-64 vuotiaat

Vuosi	Miehet		Naiset	
	otos	osallistuneita, %	otos	osallistuneita, %
1982	5 827	79	5 568	85
1987	3 962	79	3 970	85
1992	3 965	72	3 962	81
1997	5 000	68	5 000	57
2002	4 972	56	4 980	76
2007	3 984	61	3 979	71
2012	3 960	55	3 961	64

Katoon liittyviä ongelmia

- Havaintoaineiston koko pienenee
→ tulosten **tarkkuus** heikkenee
- Katoon jääneet ovat erilaisia kuin osallistuneet
→ tulokset ovat **harhaisia**
usein esim.
 - Sairaampia
 - Syrjäytyneempiä
 - Vanhimpia ja nuorimpia
 - Vähemmän koulutettuja

Tilastollisia menetelmiä kadon vaikutusten korjaamiseen

- **Painotusmenetelmät**
 - Annetaan suurempi **painoarvo** sellaisille yksilöille, joiden kaltaiset ovat usein jääneet katoon
- **Imputointimenetelmät**
 - Puuttuvan havaintoarvon paikalle asetetaan sopiva **ennustearvo**
 - Useita menetelmiä

Terveys 2000 ja 2011 tutkimukset



TERVEYDEN JA HYVINVOINNIN LAITOS



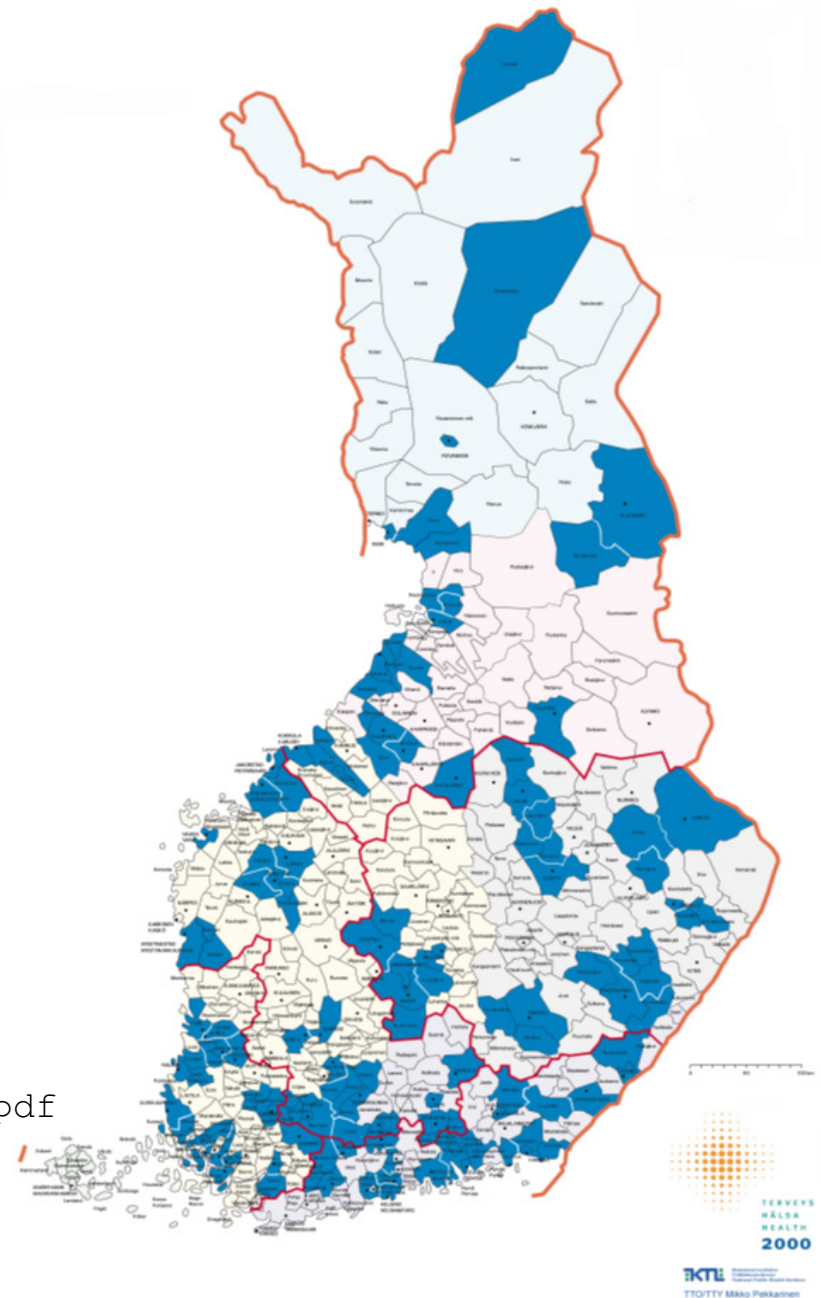
Otanta-asetelma vuonna 2000: ositus ja ryvästys

- Suomi jaettiin 20 *ositteeseen*
- Suurimmat kaupungit muodostivat **15** ositetta
- Lopuista 5 ositteesta (*miljoonapiiristä*) poimittiin **65** terveyskeskuspiiriä (*ryvästä*) 234:stä
- 8028 havaintoyksilöä poimittiin
 - yli 80-vuotiaita kaksinkertaisella todennäköisyydellä

Tutkimuspaikkakunnat

Yhteensä 80 terveyskeskuspiiriä

Lähde: Terveys 2000 menetelmäraportti.
Kansanterveyslaitoksen julkaisu B 6 / 2005.
<http://terveys2000.fi/julkaisut/2005b6.pdf>



Kaksitasoisen otannan vaikutuksia

- Kustannussäästöjä aineiston keräämisessä
 - Terveystarkastuksissa vähemmän tila- ja muuttokustannuksia
- Samasta terveyskeskuspiiristä poimittujen henkilöiden välillä riippuvuuksia:
 - Maantieteelliset etäisyydet pieniä
 - Sukulaisuussuhteita enemmän
 - Monet palvelut yhteisiä
- Riippuvuuksien huomioiminen analyyseissä välttämätöntä

Painokertoimien taustaa: otanta

Yksilöillä erilaiset poimintatodennäköisyydet

- Yli 80-vuotiailla kaksinkertainen poimintatodennäköisyys
→ otoksessa suhteellisesti enemmän vanhuksia kuin väestössä
- Esim. suora keskiarvo havainnoista tuottaa harhaisen arvion väestökeskiarvosta
- Ratkaisu: painotettu keskiarvo, jossa
 - yli 80-vuotiaille painoiksi asetetaan $\frac{1}{2}$
 - nuoremmille 1

Painokertoimien taustaa: kato

- Yksilöiden osallistumisaktiivisuus vaihtelee
- Jos osallistumisaktiivisuuteen vaikuttavat tekijät tiedetään
 - ... **sekä osallistuneista että katotapauksista**, erot voidaan korjata hyvin (esim. ikä ja sukupuoli)
 - *oletus havaittujen ja katotapausten samankaltaisuudesta*
 - ... **vain osallistuneista**, erojen huomioiminen on vaikeaa (esim. terveyteen ja toimintakykyyn liittyvät muuttujat)
 - *vahvoja oletuksia kadon luonteesta*

Painokertoimien taustaa

- Painokertoimilla voidaan korjata tunnetuista tekijöistä johtuvan kadon vaikutusta
- Terveys 2000 –painokertoimissa huomioitu
 - alkuperäinen poimintatodennäköisyys
 - ikä
 - terveyskeskuspiiri
 - sukupuoli
 - miljoonapiiri
 - äidinkieli

Terveys 2011 –tutkimus

- Toistomittaus Terveys 2000 –otokseen kuuluneille
 - 30 vuotta täyttäneitä kutsuttiin 7 964
 - 59 % osallistui terveystarkastukseen
 - 73 % osallistui johonkin tutkimusosioon
 - Perustulokset:
<http://www.julkari.fi/handle/10024/90832>
- Uusi otos nuoria aikuisia
 - Otokoko 1 994 18-28 vuotiasta

Kato

- Terveys 2011 –tutkimuksen osallistumisaktiivisuus on merkittävästi matalampi verrattuna Terveys 2000 – tutkimukseen:
93 % vs. 67 %
- Toistomittausasetelma:
 - Kaksi mittausta samoista yksilöistä
- Vuoden 2000 havaintojen hyödyntämisen puuttuvaa tietoa arvioitaessa:
 1. Vuonna 2000 mitattiin sairauksien (ja alentuneen toimintakyvyn) riskitekijöitä
 2. Sairaudet yms. lisäävät katoa
 3. Vuoden 2000 tiedoilla voidaan ennustaa terveydentilaa 2011
- → Esim. sairauksien esiintyvyyttä voidaan arvioida tarkemmin hyödyntämällä aiempia havaintoja

Kuolleisuuden vaikutus arvioitaessa puristusvoiman muutosnopeutta



TERVEYDEN JA HYVINVOINNIN LAITOS

Long-term Changes in Handgrip Strength in Men and Women - Accounting the Effect of Right Censoring Due to Death.

- *J Gerontol A Biol Sci Med Sci.* 2012 Mar 15. [Epub ahead of print]
- **Stenholm S, Härkänen T, Sainio P, Heliövaara M, Koskinen S.**
- **Abstract**
- **BACKGROUND:** Age-related decline in muscle strength is among of the most important factors in the aging process leading to disability. This study examines age-related changes in handgrip strength through a 22-year follow-up in men and women. Because handgrip strength is associated with mortality, this study also accounts for the selection effect of right censoring by comparing the estimates of handgrip strength decline based either on only the handgrip strength data or on the data of both the handgrip strength and survival times.
- **METHODS:** Data are from 1,890 men and women aged 30 years or more at baseline participating in the population-based Mini-Finland Health Examination Survey (1978-1980) with repeated handgrip strength measurement in 2000-2001.
- **RESULTS:** In men aged 31-41 years, the annual decrease in handgrip strength was approximately 3.5 Newtons (N). After that, the decrease accelerated and stabilized around the age of 75 years, being approximately 7.3 N per year. In women, respectively, prior to 45 years, the annual decrease was approximately 2 N and after age 80 years approximately 4 N per year. The estimates for the handgrip strength decline were more pronounced when the right censoring due to death was accounted for, especially for persons aged 65 years and older.
- **CONCLUSIONS:** Our work confirms that the right censoring, which depends on the outcome of interest, should be accounted for in analyses.

Puristusvoima Mini-Suomi –uusintatutkimus (MSU) -aineistossa

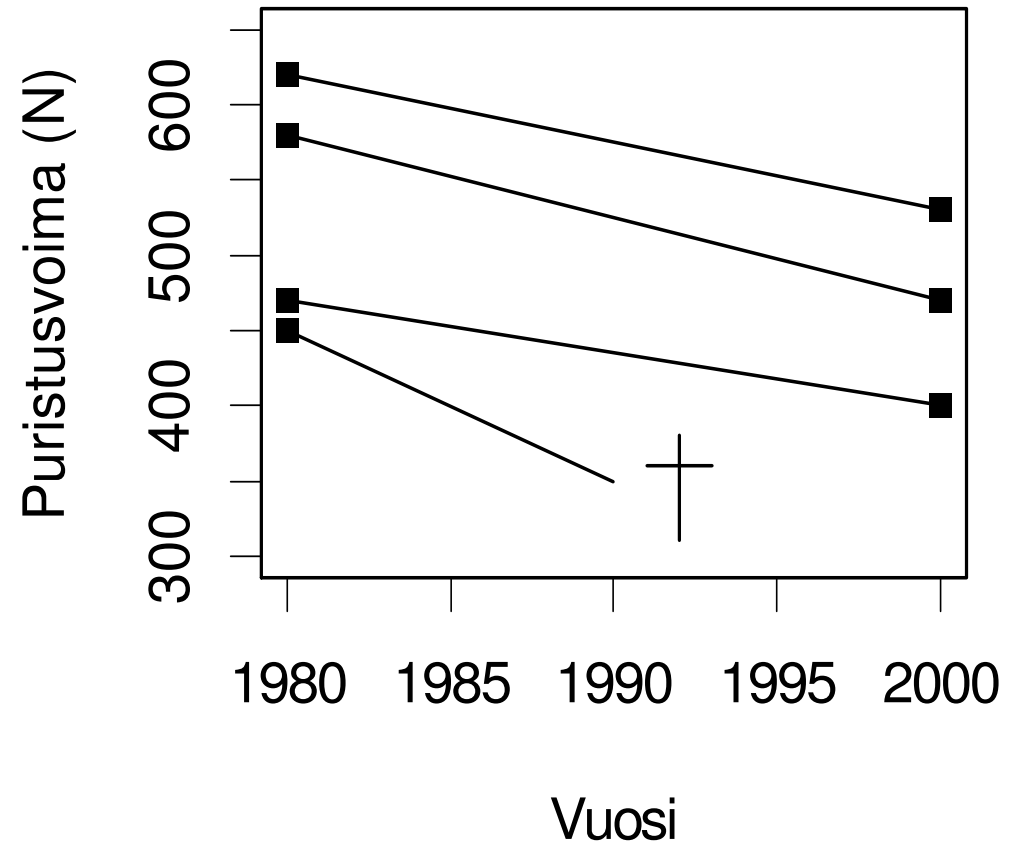
- MSU toteutettiin 9 paikkakunnalla Terveys 2000 –tutkimuksen yhteydessä
 - 1278 kutsuttiin MSU:een
- Puristusvoima: dominoivan käden paras tulos
- Pitkittäistutkimusasetelmaa approksimoitiin valitsemalla MS-aineistosta 9 paikkakunnalla asuvat (n=2049)
 - **1890** tutkittavalla baseline-puristusvoimamittaus
 - 369 kuoli ennen MSU:sta
 - 501 ei asunut enää 9 paikkakunnalla
 - 241 ei halunnut / kyennyt osallistumaan
 - **882** henkilöltä saatiin MSU-puristusvoimamittaus
- 530 kuoli MSU:n jälkeen, ennen seurannan päättymistä 31.12.2008

Oikealta sensurointi yksilötason seuranta-aineistossa

- Harhaton tulos saadaan, jos kaikki osallistuvat, ja saadaan sekä alku- että loppumittaus
- Harhan lähteitä voivat aiheuttaa mm.
 - puuttuva tieto (tutkittava ei osallistunut)
 - **oikealta sensurointi** (usein esim. kuolema)
- Tässä keskitytään oikealta sensuroinnin vaikutukseen:
 - heikkokuntoisimmat kuolevat herkemmin
→ muutosnopeus saadaan vain hyväkuntoisilta
- Muutosnopeus =
puristusvoiman muutos / mittausvälin pituus

Esimerkki sensuroinnin vaikutuksesta: 4 kuvitteellista yksilöä, kaksi mittausta

- Tutkittavat ovat samanikäiset ja samaa sukupuolta
- Lähtötilanteessa heikoin tutkittava kuoli seurannan aikana (1990)
- Toiseksi heikoin pysyi elossa, mutta heikkenemisnopeus oli hitaampi kuin muilla tutkittavilla
- Jos estimoidaan muutosnopeus vain elossa pysyneistä → heikkenemisnopeus aliarvioidaan



Puristusvoima ja kuolinriski

- Puristusvoima on yhteydessä kuolinriskiin:
 - Mitä heikompi puristusvoima, sitä suurempi kuolinriski
- → Jos tutkittava on kuollut ennen seurantapistettä, hänellä on ollut luultavasti heikompi puristusvoima kuin hengissä selvinneellä
- Käyttämällä tietoa kuolinajasta (tai hengissä selviämisestä) voidaan arvioida tutkittavan puristusvoimaa mittausten välillä
 - On mahdollista poistaa tuloksista harhaa, jonka heikkokuntoisimpien kuoleminen aiheuttaa

Ideaalitilanne: Mittauksia tehdään usein

- Mitä useammin mittauksia tehdään, sitä tarkemmin muutosnopeus pystytään estimoimaan
- **Ongelmia:**
 - vain harvoin resursseja on tarpeeksi riittävän monen mittauksen tekemiseen tai
 - tutkittavat eivät jaksakaan osallistua, jos tutkimuksia on paljon
- **Ratkaisu:**
 1. rakennetaan tilastollinen malli realististen oletusten pohjalta,
 2. hyödynnetään (rekisteri-) tietoja, joita saadaan mittausten väliseltä ajalta,
 3. täydennetään puuttuvat arvot ja
 4. analysoidaan täydennetty aineisto

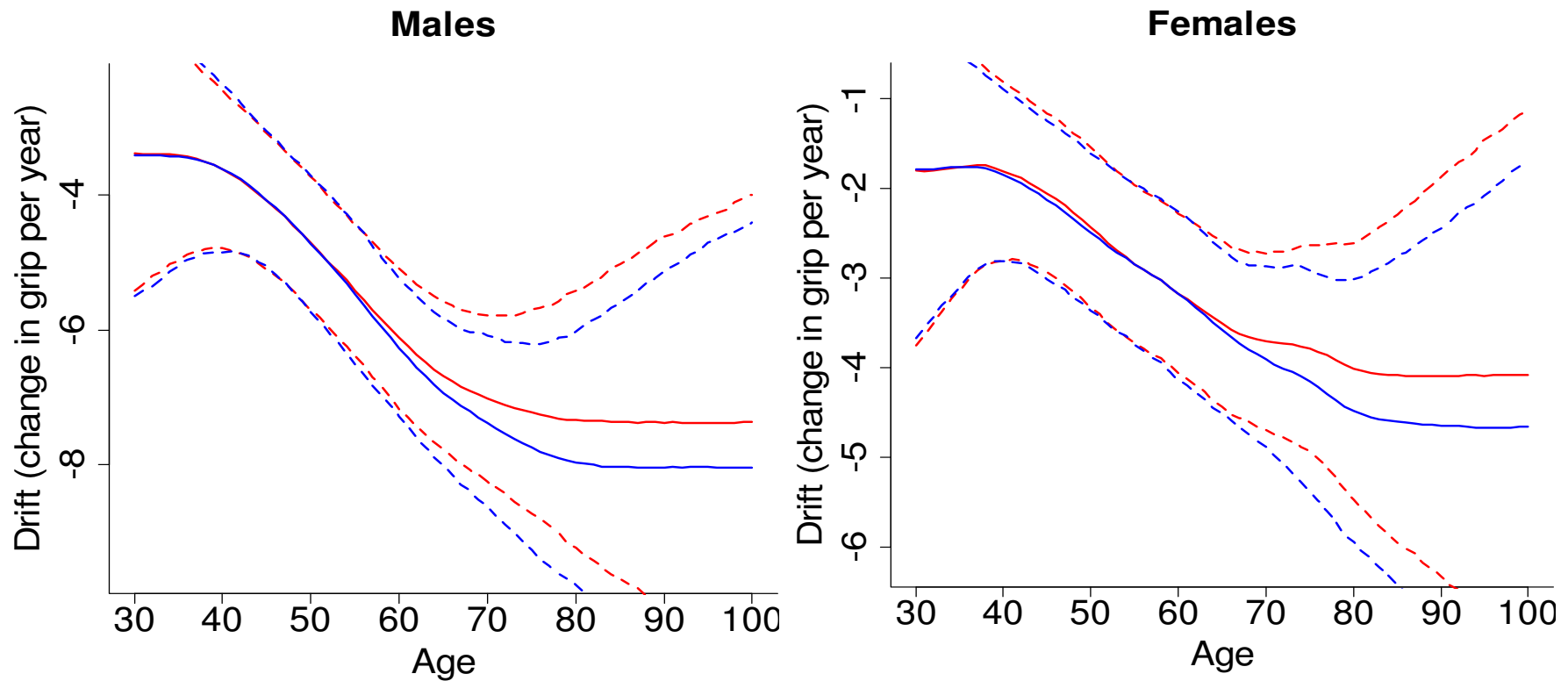
Kuolinaikojen ja puristusvoiman yhteismallitus

- **Valikoitumismallin** avulla voidaan huomioida oikealta sensuroitumisen (kuoleman) vaikutuksia
- Valikoitumismallissa käytetään kahta mallia:
 - **(Logistinen) regressiomalli**, jolla mallitetaan kuoleman (**{0, 1}-arvoinen vaste**) todennäköisyyttä varsinaisella vasteella (puristusvoima)
 - **(Lineaarinen) regressiomalli**, jolla mallitetaan varsinaista vastetta (puristusvoima, **jatkuva-arvoinen vaste**)
- Ilman puuttuvaa tietoa logit-mallia ei tarvittaisi, mutta meiltä puuttuu yleensä kuolinhetkeä edeltävä(t) puristusvoimamittaus (-mittaukset)
- Kätevä tapa käyttää valikoitumismallia on hyödyntää **Bayes-päätelyä** ja **data-augmentointia**

Mallit puristusvoiman muutoksille ja kuolinriskille

- Aika baseline-mittauksesta seurannan päättymiseen (tai kuolemaan) jaetaan esim. 1 vuoden pituisiin väleihin
- Jokaista elinvuotta kohden on olemassa
 - **puristusvoima**, mutta havaittu (mitattu) arvo on korkeintaan baseline- ja seurantatutkimuspisteissä
 - binäärinen **tieto elossaolosta** (=0) tai kuolemasta(=1)
- Regressiomalli **puristusvoimalle**:
 - Puristusvoima vuonna t =
edellisen vuoden $t-1$ puristusvoima + muutosnopeus
+ residuaalivirhe
- Logistinen malli **kuolinriskille**:
 - ikä + sukupuoli + kyseisen vuoden puristusvoima

Puristusvoiman muutosnopeus eri ikäisillä miehillä ja naisilla



- Sininen viiva: oikealta sensurointi huomioitu, valikoitumismalli
- Punainen viiva: tavanomainen analyysi
- Katkoviivat: 95 % todennäköisyysvälit



Puristusvoiman ja kuolinriskin välinen yhteys

- Logistinen malli, jossa vasteena kuolema kalenterivuoden aikana
- Selittäjät:
 - ikä,
 - sukupuoli ja
 - yhdysvaikutus:
 - puristusvoima (100 N)
 - ikäluokka
 - sukupuoli
- Oheisessa taulukossa puristusvoimainteraktion tulokset
- Suurempi puristusvoima on yhteydessä pienempään kuolinriskiin (OR < 1)

Sukupuoli	Ikäluokka	OR	Todennäköväli
M	30-64	0.87	0.70-1.07
M	65+	0.78	0.67-0.91
N	30-64	0.65	0.43-0.93
N	65+	0.82	0.65-1.00



Lopuksi

- Muutosten arviointi väestötutkimuksissa voidaan jakaa karkeasti kahteen ryhmään:
 - Populaatiotason muutokset esim. ikä- ja sukupuoliryhmittäin
 - Yksilötasolla tapahtuvat muutokset
- **Populaatiotasolla** kuolleisuus, maastamuutto ym. valikoitumismekanismit vaikuttavat samalla tavalla sekä tutkimusotoksessa että väestössä
→ otoksen edustavuus säilyy
- **Yksilötason** muutosten tarkastelussa **oikealta sensurointi erityisesti vanhusväestössä** voi vaikuttaa tuloksiin, jos tutkittava ilmiö on yhteydessä oikealta sensuroitumisen aiheuttavaan mekanismiin

Prediction projects

Tommi Härkänen



TERVEYDEN JA HYVINVOINNIN LAITOS



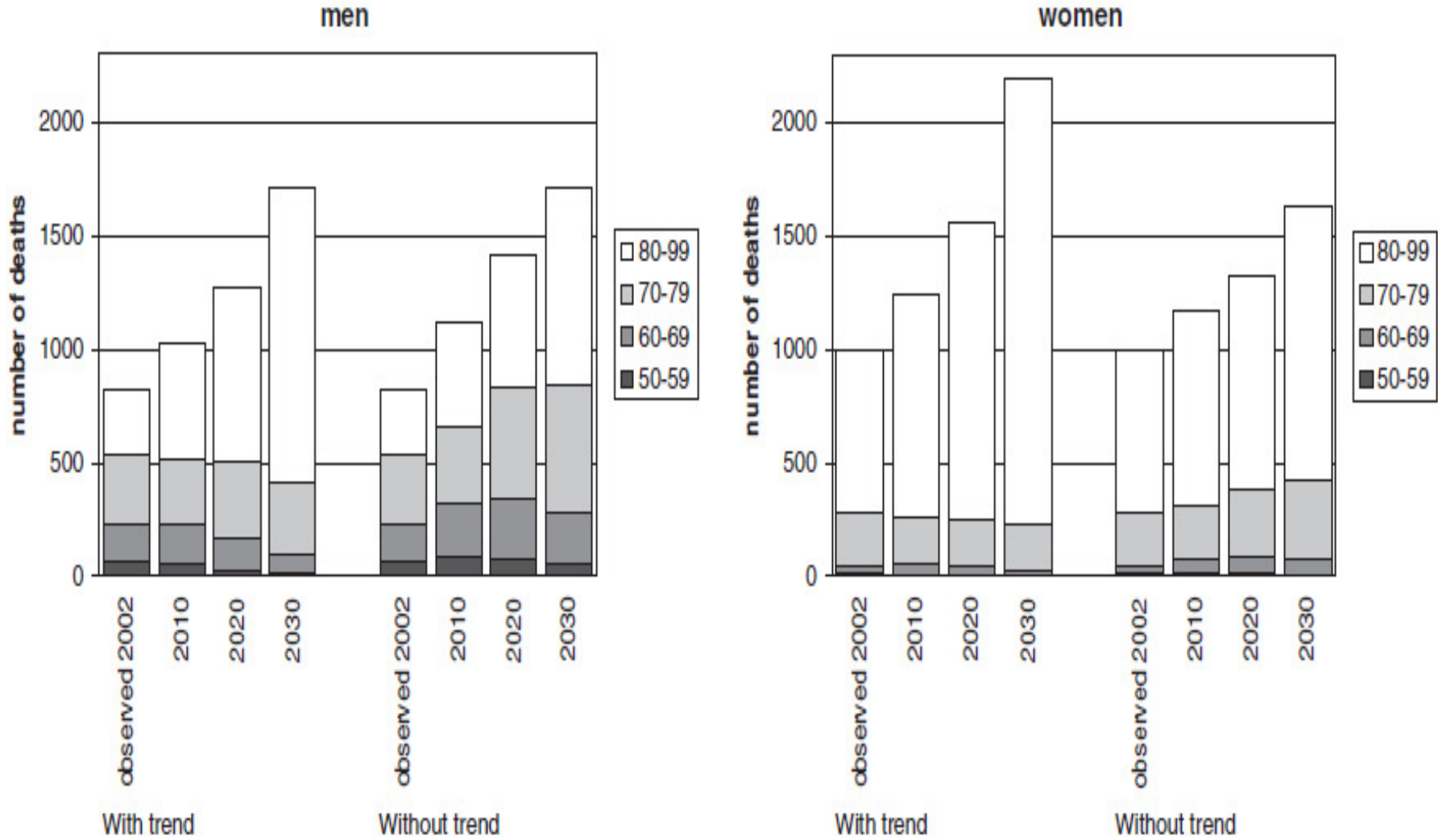
Contents

- Coronary heart disease mortality 2030 ([Huovinen et al. 2006](#))
- Elämä pelissä (Koskinen et al. 2008)
- Ageing report ([Prime Minister's Office 2009](#), in Finnish)
- Population Attributable Fraction (PAF, [Laaksonen et al. 2010](#))
- Sugar tax ([Härkänen et al. 2014](#))
- Educational differences in sickness absence (Kaikkonen 2015)
- Unpublished:
 - Healthy life years
 - Mobility predictions until 2044

Coronary heart disease mortality

- Data:
 - North Karelia and Kuopio in and
 - Eastern Finland and the Turku/Loimaa
 - Annual mortality data 1972–2002 by gender and age (5 year age groups from 30 years up to 99 years)
- Age-period-cohort model for
 - Population sizes (normal distribution) and
 - Number of deaths (binomial distribution, logistic model)
- Bayesian inference with smoothing (random-walk) priors

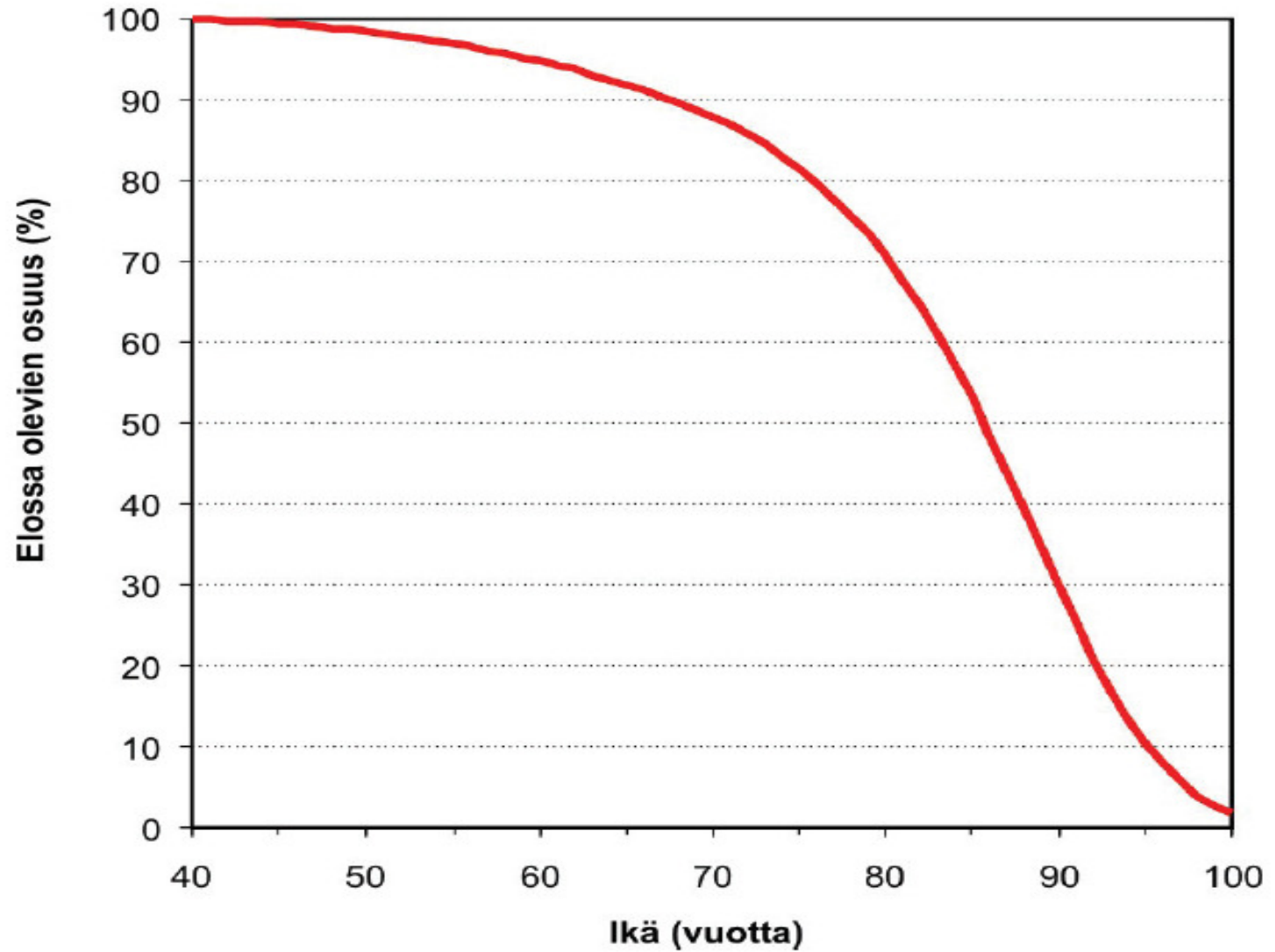
Coronary heart disease mortality



Elämä pelissä

- Data: Finrisk 1982-1997, mortality follow-up (until 2001)
 - 39 predictors: demographic, lifestyle and biologic risk factors
- Piecewise exponential models:
 - Model with only age and gender (**population risk**)
 - 4 models containing all risk factors with/without cholesterol or blood pressure (**individual risk**)
- Under age **75** we used individual risk
 - After age **105** we used population risk
 - Between ages 75 and 105 the annual risk was **average** of individual risk and population risk

Elämä pelissä: Survival function for 40 year old women



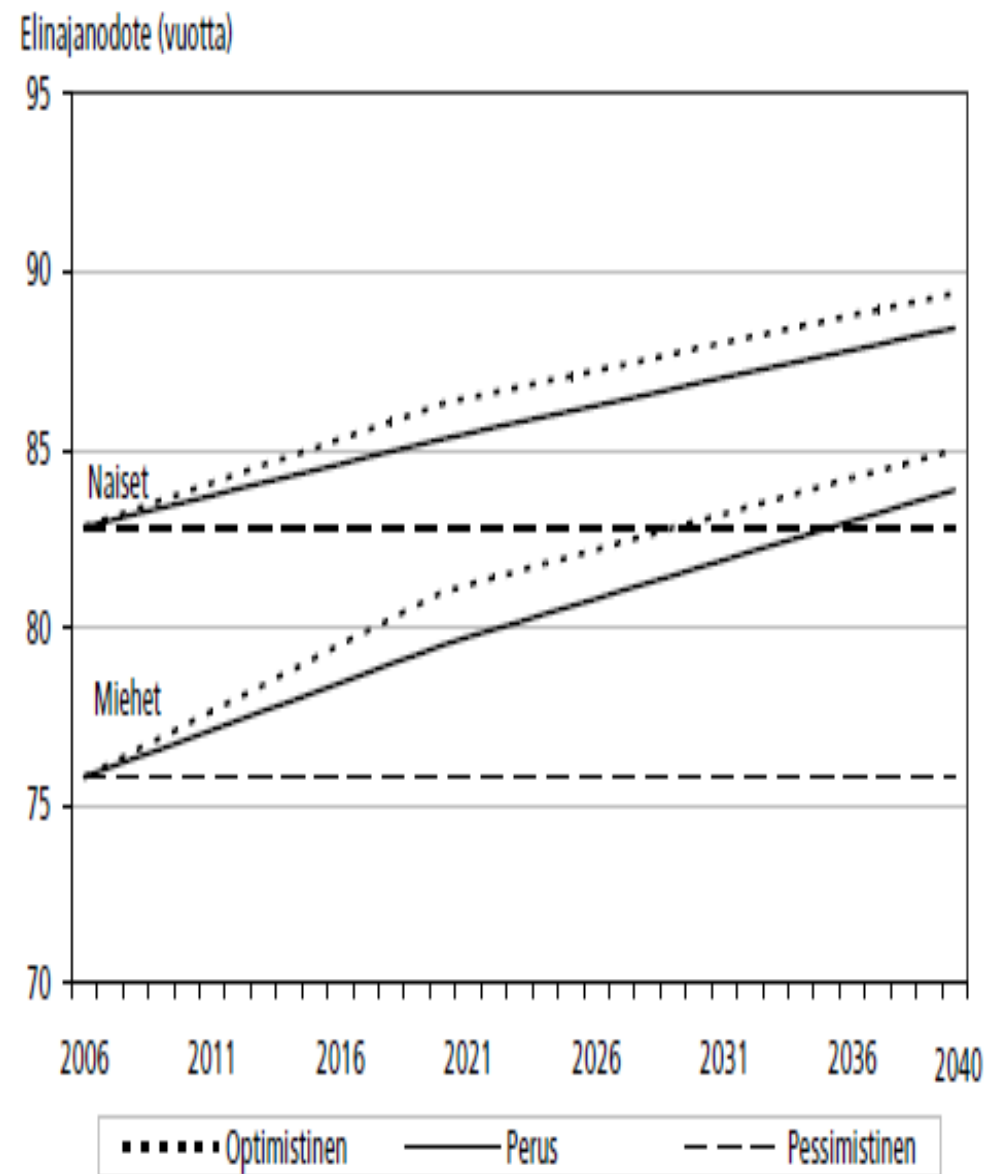
Ageing report

- Data, Statistics Finland:
 - Mortality (age and gender)
 - Mortality trend between 1987-91 and 2002-6
 - Cause of death in 2006 (5-year age groups)
- Removal of a cause of death 'x' (in age and gender groups):
 $\text{mortality} - \text{death from 'x'} / \text{all deaths}$
- Population predictions and life expectancies were calculated using life table methods

MIEHET	Elinajan- odote	65–79-v.		80–89-v.		90+ -v.	
		Lkm	%	Lkm	%	Lkm	%
Lähtötilanne 2006	75,80	285 871		59 313		5 998	
2040							
Tilastokeskuksen ennuste	83,86	447 082		202 084		46 762	
Kuolemansyyntä poistamisen vaikutus: elinajanodotteen ja iäkkäiden määrän poikkeama Tilastokeskuksen ennustamasta							
Kasvalmet	2,28	20 971	4,7	36 355	18,0	21 675	46,3
- keuhkosityöpä	0,23	5 950	1,3	9 184	4,5	4 175	8,9
- eturauhassyöpä	0,43	1 781	0,4	5 215	2,6	4 171	8,9
Dementiat	0,57	446	0,1	3 479	1,7	6 325	13,4
Alkoholikuolemansyyt	0,29	9 686	2,2	5 492	2,7	992	2,1
Sepelvaltimotauti	2,13	18 143	4,1	38 510	19,1	39 737	84,6
Aivoverisuonisairaudet	0,98	4 609	1,0	10 964	5,4	9 248	19,7
Liikennetapaturmat	0,08	594	0,1	487	0,2	194	0,4
Kaatumiset	0,18	2 234	0,5	3 172	1,6	2 384	5,1
Itsemurhat	0,23	2 833	0,6	1 819	0,9	421	0,9
NAISET							
Lähtötilanne 2006	82,83	361 916		132 614		23 080	
2040							
Tilastokeskuksen ennuste	88,42	480 256		277 979		94 292	
Kuolemansyyntä poistamisen vaikutus: elinajanodotteen ja iäkkäiden määrän poikkeama Tilastokeskuksen ennustamasta							
Kasvalmet	1,78	14 022	2,9	24 841	8,9	19 741	20,9
- keuhkosityöpä	0,18	1 957	0,4	3 002	1,1	1 682	1,8
- rintasyöpä	0,35	2 936	0,6	3 836	1,4	2 466	2,6
Dementiat	0,65	168	0,0	4 046	1,5	15 388	16,3
Alkoholikuolemansyyt	0,24	2 269	0,5	1 668	0,6	452	0,5
Sepelvaltimotauti	2,04	4 105	0,9	20 382	7,3	46 104	48,8
Aivoverisuonisairaudet	0,91	2 461	0,5	10 093	3,6	17 929	19,0
Liikennetapaturmat	0,06	168	0,0	278	0,1	94	0,1
Kaatumiset	0,17	498	0,1	1 863	0,7	3 399	3,6
Itsemurhat	0,18	498	0,1	834	0,3	283	0,3

Ageing report: Life expectancy scenarios

- **Pessimistic** scenario:
Mortality remains on 2006 level
- **Basic** scenario:
Mortality trend used by
Statistics Finland continues
- **Optimistic** scenario:
Half of CHD mortality gets
removed



Population Attributable Fraction (PAF)

- "Avoidable proportion of cases if a risk factor could be eliminated"
- Three versions available for different designs:
 - Cross-sectional
 - Cohort (survival outcome)
 - Cohort (disease occurrence) with informative right-censoring (death)
- Under development:
 - Repeated measurements
- Related concept: Predictive margins (model adjustment)

PAF for disease occurrence

- The same risk factor can cause both disease and death
 - Elimination of the risk factor can prolong life
→ an individual can still contract the disease, but probably later
 - The stronger association between the risk factor and death, the more a standard cohort-PAF overestimates the PAF

RR for disease	RR for death	PAF _{1,5}	PAF _{2,5}
2	2	0.38 (0.20, 0.53)	0.41 (0.22, 0.55)
2	4	0.35 (0.15, 0.50)	0.41 (0.22, 0.55)
4	2	0.64 (0.52, 0.73)	0.66 (0.54, 0.75)
4	4	0.62 (0.49, 0.71)	0.66 (0.54, 0.75)

Table IV. Comparison of the estimates of PAF for disease incidence (95 per cent confidence intervals) accounting (PAF_{1,t}) and not accounting (PAF_{2,t}) for censoring due to death in simulated data for 50 000 individuals according to relative risk (RR) of the modified binary risk factor for disease and death.

PAF accounting or not for death

Table III. Comparison of the estimates of PAF for the incidence of type 2 diabetes (95 per cent confidence intervals) accounting ($PAF_{1,t}$) and not accounting ($PAF_{2,t}$) for censoring due to death in Mini-Finland Health Survey according to modification of potential life-style-related risk factors.

Risk factor modification	$PAF_{1,10}$	$PAF_{2,10}$
Smoking (3,4 → 1)*	0.07 (−0.03, 0.17)	0.09 (−0.02, 0.19)
Alcohol consumption (3 → 2)*	0.04 (−0.03, 0.10)	0.04 (−0.03, 0.10)
Physical exercise (1 → 2)*	0.14 (−0.01, 0.26)	0.15 (0.00, 0.28)
BMI (3 → 2)*	0.40 (0.28, 0.49)	0.40 (0.29, 0.50)
All four variables	0.50 (0.35, 0.61)	0.51 (0.37, 0.63)

Sugar tax

Data:

- The Household Budget Survey of Statistics Finland (1995–6, 1998, 2001, 2006) → cross-price elasticities
- The Health 2000 Survey → distribution of food consumption
- Fineli → average nutrition content of different foods

Food groups in the model:

- Bread, Meat, Fish, Fruit and veg, Sugar and sweets, Others

Sugar tax: Literature (meta-analyses)

- Risk ratios (RR) of an obese (BMI > 30) vs. normal weight
 - **diabetes** is **7.2** (Abdullah et al., 2010).
 - **coronary heart disease** is **1.8** (Bogers et al., 2007)
- Nutrition
 - **CHD death** is **0.64** average 29 g of fatty **fish** or 48 g of less fatty fish per day RR is **0.64** (Mozaffarian and Rimm 2006)
 - Additional **106 g / day** of **vegetables and fruit** RR of
 - **cardiovascular mortality** is **0.74**, and
 - **CHD** is **0.96** (fruit and vegetable intake) and **0.93** (fruit intake)
- Daily reduction of **20 kcal for men** and **12 kcal for women** was associated with a **one kilogram** reduction in body weight

Tax reforms

- **Sugar tax:** a tax of one euro per kilogram of added sugar applied to each food category based on its sugar content.
- **Cut in VAT:** abolition of the current VAT on fresh fruit, vegetables and fish.
- **Combined reform:** both of the reforms above.

	Change
<i>Sugar tax</i>	
Bread	-0.0377*
Meat	0.0224*
Fish	0.0392
Fruit and veg	-0.0057
Sugar and sweets	-0.2331*
<i>VAT cut</i>	
Bread	-0.0128
Meat	0.0442*
Fish	0.1155*
Fruit and veg	0.0537*
Sugar and sweets	-0.0057
<i>Both reforms</i>	
Bread	-0.0505*
Meat	0.0666*
Fish	0.1547*
Fruit and veg	0.0480*
Sugar and sweets	-0.2389*

Results for sugar tax: body weight and T2D

Change in body weight (kgs) as a result of the sugar tax.

All	-3.19 (-4.89, -1.44)		
By sex	-2.54 (-3.89, -1.13) (males)	-3.79 (-5.81, -1.73) (females)	
By education	-3.02 (-4.73, -1.30) (basic education)	-3.17 (-4.87, -1.40) (secondary)	-3.44 (-5.20, -1.63) (tertiary)
By household income	-5.41 (-8.59, -2.53) (low income)	-0.78 (-3.7, 2.11) (middle income)	-2.63 (-5.4, 0.28) (high income)

Change (negative PAR_{2C}, %) in the incidence of T2D as a result of the sugar tax.

All	-13.4 (-6.3, -19.9)		
By sex	-10.8 (-5.2, -15.7) (males)	-15.9 (-7.1, -23.6) (females)	
By education	-12.5 (-5.5, -19.0) (basic education)	-13.8 (-6.7, -20.0) (secondary)	-14.4 (-7.2, -21.3) (tertiary)
By household income	-20.8 (-10.3, -30.5) (low income)	-3.2 (9.6, -16.0) (middle income)	-11.8 (2.3, -22.6) (high income)

Educational differences in sickness absence

- We predicted population average of sickness absence days per working year (DWY)
- Based on expected outcome values using Poisson and gamma regression models
- Health 2000 Survey

Men	Education			
	High	SE	Low	SE
M1: Age (reference model)	3.19	0.51	7.97	0.51
M13: Age + health behavior and obesity + physical + psychosocial working conditions	3.19	0.51	6.09	0.49
M14: All main effects	3.19	0.5	6.19	0.53

Unpublished work

- **Healthy life years:** Individual predictions using
 - Finnish Mobile Clinic and Mini Finland survey data
 - a small number of risk factors
 - T2D, CHD, cancer and death as outcomes
 - Illness-death model using Bayesian intensity models
- **Functional capacity:** Population predictions
 - The Health 2000 and 2011 surveys (repeated measurements on risk factors, outcome and mortality)
 - Prevalence and count predictions for 2022, 2033 and 2044

References

- Huovinen E, Härkänen T, Martelin T, Koskinen S, Aromaa A. Predicting coronary heart disease mortality-assessing uncertainties in population forecasts and death probabilities by using Bayesian inference. *Int J Epidemiol*. 2006.
- Koskinen S, Härkänen T, Jousilahti P, Kuulasmaa K, Martelin T, Sihvonen A-P, Vanhalakka N, Virtala E, Puska P. Elämä Pelissä. *Yleislääkäri* 2008(1): 11-19.
- Laaksonen MA, Härkänen T, Knekt P, Virtala E, Oja H. Estimation of population attributable fraction (PAF) for disease occurrence in a cohort study design. *Stat Med*. 2010 Mar 30;29(7-8):860-74.
- Härkänen, T., Kotakorpi, K., Pietinen, P., Pirttilä, J., Reinivuo, H., & Suoniemi, I. (2014). The welfare effects of health-based food tax policy. *Food Policy*, 49, 196-206.
- Kaikkonen, Härkänen, Rahkonen, Gould, Koskinen (2014). Explaining educational differences in sickness absence: a population-based follow-up study. *Scand J Work Environ Health* – online first.

Tilastolliset ohjelmistot



TERVEYDEN JA HYVINVOINNIN LAITOS

Tilastollisten ohjelmistojen ominaisuuksia

- **Aineisto**
 - tallennetaan yleensä matriisimuodossa
 - käsitellään erilaisten funktioiden avulla
- Tilastolliset **analyysit**
 - kuvaillaan aineistoa ja
 - yleistetään tuloksia koko perusjoukkoon
- **Tulokset**
 - voidaan **esittää** monipuolisesti, esim. taulukoina tai graafisesti (ja tallentaa kuvatiedostoiksi)
 - voidaan **viedä** erilaisissa muodoissa käytettäväksi muissa ohjelmissa
- **Laajentaa** ohjelmiston toiminnallisuutta ohjelmoimalla uusia funktioita

Muutamia ohjelmistoja joita THL:lla käytetään

	SAS	SPSS	Stata	R
Aineistonhallinta	+	0	+	+
Tilastolliset analyysit	+	+	+	+
Uusimmat menetelmät	0	0	+	+
Tulosten esittäminen	+	+	0	+
Laajennettavuus	0	-	0	+
Monipuolisuus	+	0	+	+
Hinta	-	0	0	+
Käyttäjäystävällisyys	0	+	+	0

Muita tilasto-ohjelmistoja: Sudaan, Mplus, (Win-/Open-)Bugs, Jags, Survo, ...

R-ohjelmistosta

- **S kieli**
 - **Old S** S on yksi tilastolaskentaan kehitetyistä kielistä (Bell Laboratories), ensimmäinen versio 1975-1976
 - **New S** vuonna 1991, Statistical Models in S (White Book), jossa esiteltiin mm. datakehikko-olio, julkaistiin
- **Kaupallinen ohjelmisto S-plus** ilmestyi 1988
- **Avoimen lähdekoodin R-ohjelmisto**
 - Versio 0.16 Ensimmäinen julkinen versio, 1. 4.1997
 - Versio 1.0.0 1. tuotantoversio, 29.2.2000
 - Versio 3.1.1 Tämänhetkinen versio, 10.7.2014

R-aiheisia verkkosivustoja

Pääsivusto: <http://www.r-project.org/>

- CRAN: The Comprehensive R Archive Network
 - **Download** R-ohjelmiston lataus Windows-, Mac OS X- ja Linux-käyttöjärjestelmille
 - **Extension packages** Lisäpaketit, joiden avulla voi laajentaa tai parantaa R:n toiminnallisuutta (tällä hetkellä 5,274 pakettia)
 - **CRAN Task Views** Eri aihealueiden mukaan esitelty tärkeimpiä lisäpaketteja. Tällä hetkellä aihealueita on 33.

Parannettu käyttöliittymä on **RStudio**:

- <http://www.rstudio.com/> Tutustumisen arvoinen(!)

Graafinen käyttöliittymä **R commander**:

- <http://www.rcommander.com/>

Tehokkain tapa käyttää R:ää on komentorivi ja **S-kieli!**

Muita R-verkkosivustoja

- R-aiheisia **blogi**kirjoituksia löytyy <http://www.r-bloggers.com/> monia aiheita, esim.
 - opastukset R:n ominaisuuksiin, uudet ominaisuudet, uudet/päivitetyt lisäpaketit.
- Blogi **suomeksi**: <http://www.r-ohjelmointi.org/>
- Esimerkkejä R:n avulla tehtyä **grafiikkaa**: <http://gallery.r-enthusiasts.com/>

S-kielen ominaisuuksia

- Tietotyyppejä: numeerinen, merkkimuotoinen, looginen
- Tietorakenteita: skalaari, vektori, matriisi, datakehikko, lista
- Monia funktioita tietorakenteiden ja –tyyppien käsittelemiseen
- **Olio-orientoitunut** kieli
 - R esim. tunnistaa automaattisesti funktion argumenttina annetun olion tyyppin, ja valitsee käytettävän funktion
Esim. `print(x)`-funktiolle on tehty yli 182 versiota erityyppisiä olioita varten, kuten
`print.data.frame()`; `print.Date()`; `print.table()`; ...
- **Funktionaalinen** kieli: funktioita voidaan
 - antaa argumentteina toisille funktioille ja
 - palauttaa funktion tuloksena

Esimerkkejä S-kielestä

- Tietotyyppejä:
`x <- c(1, 3, 4, 6) # vektori, jossa 4 elementtiä`
- Muunnetaan vektori merkkimuotoiseksi:
`class(x) <- "character"`
- Muunnetaan vektori matriisiksi (annetaan ulottuvuudet):
`dim(x) <- c(2, 2) # rivit ja sarakkeet`
- Esimerkki funktionaalisuudesta: vektori, jossa kolme funktiota:

```
fn <- c(mean, median, sd)
for (i in fn) { # silmukka
  print(i(1:4)) # argumenttina vektori c(1,2,3,4)
}
[1] 2.5          ## keskiarvo, mean
[1] 2.5          ## mediaani, median
[1] 1.290994    ## keskihajonta, sd
```

R-ohjelmiston käyttökohteita THL:lla

- Aineistojen käsittelyminen
 - Rajapintoja: tietokannat, xml
- Tilastolliset analyysit
- Raportointi
 - Tilastografiikka
 - Web-raportit
 - Pdf-raportit