

Opaste kuolleisuus- ja eloonjäämislukujen tulkintaan

Risto Kolari

18.1.1991

Sisällys

	Sivu
1. Mitä ovat kuolleisuus- ja eloonjäämisluvut	3
2. Laskennan kulku	5
3. Suureiden tulkinnasta	8
A. Kokonaiskuolleisuus	8
B. Todennäköisyysuureista	8
C. Laskelmat kuolinsyiden mukaan	9
4. Reaaliväestö ja stationaarinen väestö	11
5. Tulosten luotettavuudesta	13
A. Kokonaiskuolleisuus	13
B. Kuolleisuus ikäryhmittäin	17
C. Vähimmäisehdot	19
6. Yleispäätelmä	20

Oheinen selostus koskee ennen muuta kuolleisuus- ja eloonjäämislukuja, joiden laskemiseksi Tilastokeskuksen väestöselvityksyksikössä on laadittu kaksivaiheinen laskentajärjestelmä. Aivan lyhyesti sanoen laskenta tapahtuu siten, että ensin tulostetaan TASSU-ohjelmilla lähtötietoina tarvittavat taulut, jotka sitten

siirretään APL:ään, missä varsinaiset laskelmat suoritetaan ja julkaisutaulut tulostetaan. Lisäksi APL:ssä on ohjelmia, joilla voidaan laskea esim. tyyppi-riivot ja kuolleisuusindeksit. Laskelmiin kuuluvien suureiden määritelmät ja laskukaavat on esitetty virallisen tilaston julkaisuissa.

1. Mitä ovat kuolleisuus- ja eloonjäämisluvut?

Kuolleisuus- ja eloonjäämisluvut muodostavat ehyen ja hyvin rajatun systeemin, mikä tarkoittaa sitä, että ko. suureet voidaan määritellä ja myös laskea toistensa avulla tietyssä järjestyksessä. Tätä järjestelmää voisi luonnehtia siten, että kysymyksessä on kulloinkin erään stationaarisen väestön kuolleisuus, ja stationaarinen väestö taas määräytyy yksinomaan kuolleisuuden mukaan. Juuri tästä syystä kuolleisuus- ja eloonjäämislukuja voi pitää "puhdistettuina" suureina, joita väestörakenne tai muut väestölliset tekijät eivät vääristä miksikään.

Järjestelmää voi toisaalta luonnehtia siten, että kysymyksessä on kohorttimallin soveltaminen. Tämä pitää paikkansa silloinkin, kun laskelmat suoritetaan jotakin periodia koskevinä. Suureiden määrittely ja tulkinta perustuvat kohorttiajatteluun. Sellainen periodilaskelma kuvaa kuolleisuutta ko. ajankohdan mukaisena, mutta toisaalta sitä voidaan käyttää myös ennusteena. Esim. keskimääräiset jäljelläolevat elinajat ikäryhmittäin tarjoavat käyttökelpoisen ennusteen, joka kuitenkin pitää paikkansa vain sillä edellytyksellä, että kuolleisuus pysyisi vastaisuudessa nykyisellä tasolla.

On aika ajoin keskusteltu siitä, eikö kuolleisuustauluja pitäisi laskea säännöllisesti myös kohorteittain. Käytännössä tähän ei ole menty. Kaikkialla virallisen tilaston kuolleisuustaulut lasketaan pe-

riodikohtaisesti, mikä sopii parhaiten useimmille käyttäjille. Kohorttikohtaisia laskelmia on sen sijaan usein tehty tutkimusten yhteydessä.

Käytäntö on vaihdellut jonkin verran sen suhteen, mitä suureita kuolleisuustauluihin sisällytetään. Nimike kuolleisuus "-taulu" on peräisin siitä, että esitettiin tietyt ikäryhmittäin lasketut suureet rinnakkain tietyssä järjestyksessä. Näinhän oikeastaan nytkin tapahtuu, mutta laskelmien laajeneminen on johtanut siihen, että jo yksittäisiä suureita koskevat tulosteet muodostavat tauluja. Tavallisesti laskelmaan sisällytetään ainakin kuolemanvaaraluvut, eloonjäämisluvut ja keskimääräiset jäljelläolevat elinajat. Täydellisempään laskelmaan tulevat äsken mainittujen lisäksi todennäköiset jäljelläolevat elinajat, kuolintodennäköisyydet ja stationaarinen väestö. Lisäksi voidaan esittää erillisinä tulosteina joitakin tunnuslukuja kuten tyyppilinen elinaika ja stationaarisen väestön keskimääräisikä. Kuolleisuusintensiteetti, joka aikaisemmin sisällytettiin varsinaiseen tauluun, jätetään nykyisin yleensä pois.

Jo varhaisimmissa kuolleisuustauluissa muuttujina olivat ikä ja sukupuoli, ja ne ovat yhä perusmuuttujia laskelmissa. Myös siviilisääty on demografisesti tärkeä muuttuja, joka kuitenkin aiheuttaa tulkinnallisia ja samalla laskennallisia ongelmia. Kaikesta huolimatta laskelma

voidaan suorittaa siviilisäädyn mukaan ainakin kuolemanvaaralukujen osalta ja myös muiden suureiden osalta soveltamalla kohorttiajattelua siviilisäätyryhmiin. Tilastokeskuksen julkaisemiin kuolleisuustauluihin on em. lisäksi sisällytetty muuttujina kuntamuoto ja lääni sekä seutukaava-alue.

Kuolleisuus- ja eloonjäämislukujen laskeminen merkitsee itse asiassa erään yleisemmän matemaattisen menetelmän soveltamista. Laskelmissa seurataan, kuinka oletettu samalla hetkellä syntyneiden joukko (kohortti) pienenee kuolleisuuden - ja vain sen - johdosta. Vertauksenomaisesti voisi sanoa, että kysymyksessä on otanta ilman palautusta, ja tällä ehdolla malli toimii moitteettomasti. Jos joukko pienenee muista syistä kuin kuolleisuuden johdosta ja samoin, jos siihen tulee uusia yksilöitä (muutto-liike, siviilisäädyn muutokset jne.), las-

kelman luotettavuus kärsii. Joka tapauksessa ns. riskiväestön muodostavat tällöin ne, jotka kulloinkin ovat vielä elossa.

Riskiväestön käsite tulee erikoisella tavalla esille, kun lasketaan kuolleisuustaulut kuolemansyiden mukaan. Tällöin nim. muodostuu laskennallisesti osajoukkoja, joista jokainen voitaisiin määrittellä kuolemansyykohtaiseksi riskiväestöksi. Toisaalta kaikkien elossaolevien joukko (kuolemansyistä riippumatta) on eräs mahdollinen riskiväestö. Viimeksi mainittu käsite näyttää parhaiten vastaavan käytännön ajattelua, jolloin oletetaan, että esim. syöpään voi kuolla kuka tahansa. Ensiksi mainittua riskiväestön käsitettä vastaa ominaiskuolemanvaara ja jälkimmäistä osittaiskuolemanvaara. Näistä on Tilastokeskuksen asianomaiseen julkaisuun sisällytetty vain jälkimmäinen, mutta APL-ohjelmassa tulostuu myös edellinen.

2. Laskennan kulku

Laskelmien kulku on suurpiirteisesti esitetty kaaviossa 1, josta käyvät selville myös suureiden nimet ja niiden symboliset merkinnät. Kaaviosta on nähtävissä, että eloonjäämistodennäköisyys l_x on määrittelyn ja laskennan kannalta keskeinen suure. Tätä suuretta ei kuitenkaan voida käytännössä ottaa laskelmien lähtökohdaksi, koska sitä ei tunneta etukäteen eikä sitä voida suoraan aineistotiedoista laskea.

Kuolemanvaarakerroin q_x sen sijaan on suure, joka voidaan arvioida käyttämällä hyväksi perustilaston tietoja. Arviointia varten on olemassa useita vaihtoehtoisia laskukaavoja. Kaavioon 1 on (kohtaan 2) merkitty kaava, jota voisi nimittää "normaalikaavaksi" ja joka ainakin 1-vuotisikäryhmitystä käytettäessä on melko luotettava. Tätä kaavaa käytettiin Tilastokeskuksessa aikaisemmin virallisen tilaston kuolleisuustauluja laskettaessa. Nykyisin käytetään "osituskaavaa", joka vastaa paremmin klassillisen todennäköisyyslaskennan periaatteita. Sitä varten tarvitaan syntyneiden lukumäärä (B), ja lisäksi tehdään kaikissa ikäryhmissä nettosiirtolaisuuskorjaus. Näiden kaavojen avulla saadut tulokset eivät sannottavasti poikkea toisistaan, ellei peräkkäisten kohorttien välillä ole huomattavan suuria eroja. Erilaisissa selvityslaskelmissa sekä tietopalvelussa (esim. kuolleisuustaulut kunnittain) on monesti edullisempaa käyttää normaalikaavaa,

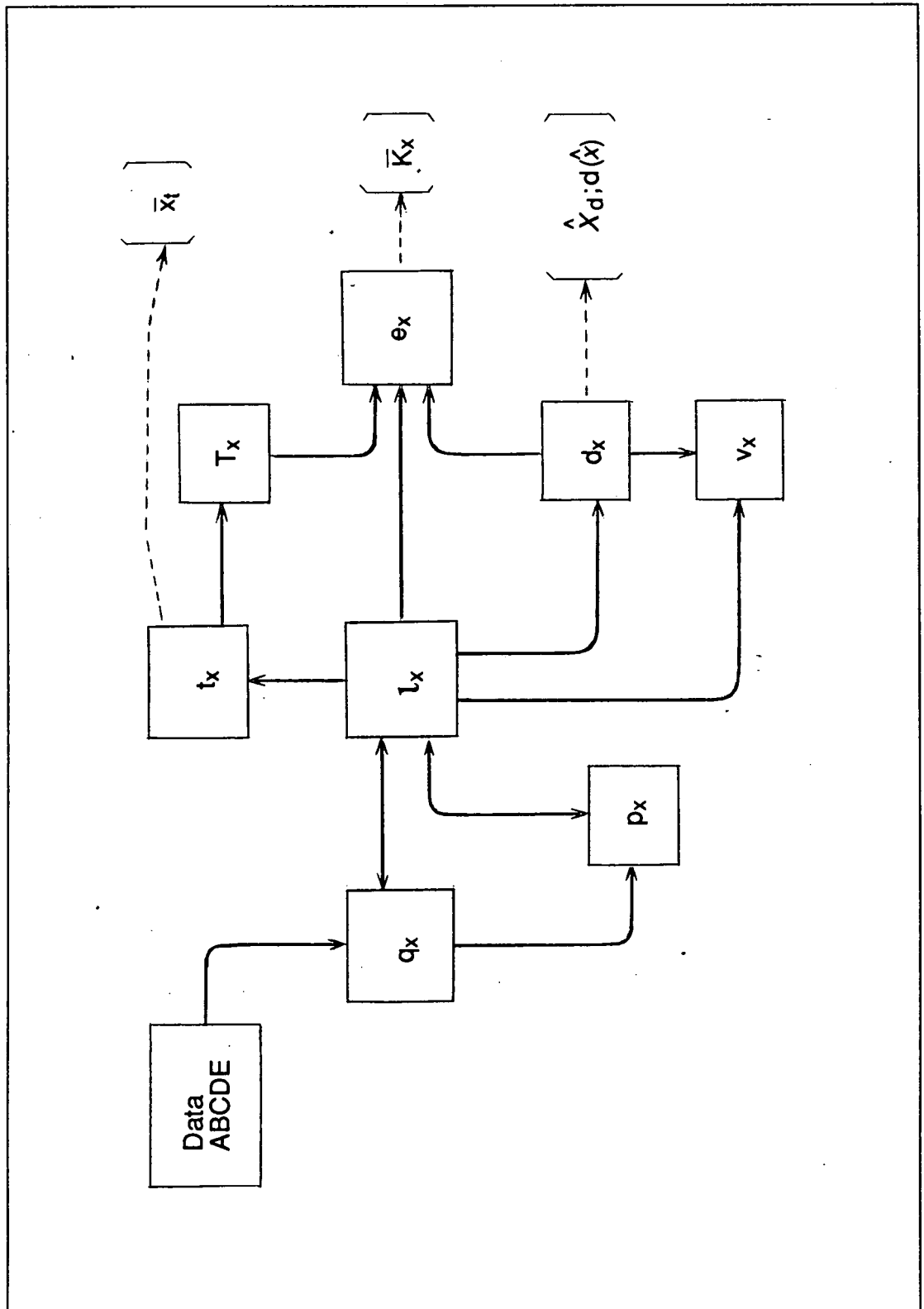
koska se on yksinkertaisempi kuin osituskaava.

Kun kuolemanvaaraluvut ovat käytettävänä, voidaan laskea eloonjäämistodennäköisyydet ja niiden avulla edelleen muut halutut suureet ennen kaikkea keskimääräinen jäljelläoleva elinaika e_x , joka saadaan useallakin eri muotoisella kaavalla.

Kun laskenta tapahtuu tietokoneella, on mahdollista menetellä useassa kohden toisin kuin aikaisemmin käsin laskettaessa. Esim. suuretta p_x ei tietokoneella laskettaessa kannata tulostaa näkyviin. Sama koskee suuretta T_x , jota kylläkin voidaan käyttää kaavan (8) osoittamalla tavalla apusuurena elinaikoja laskettaessa.

Oman kysymyksensä muodostaa ikäluokituksen valinta. Satunnaisvaihtelun vähentämiseksi käytetään usein 5-vuotiskäluokitusta 1-vuotiskäluokituksen sijasta. Väestöselvitysyksikön laskentaohjelmassa on kuitenkin sovellettu 1-vuotiskäluokitusta kauttaaltaan laskennan alkuvaiheessa, mikä tietenkin edellyttää, että aineistotiedot on laadittu vastaavalla tavalla. Näin saatujen eloonjäämislukujen avulla lasketaan kuolemanvaaraluvut 5-vuotiskäryhmittäin kaavaa (1) käyttäen, jolloin lopputulos on luotettavampi kuin, jos alun perin tyydyttäisiin karkeampaan luokitukseen.

Kaavio 1. Kuolleisuustaulujen laskeminen



Kaavio 1. (Jatk.)

- q_x = kuolemanvaara ikävälillä $x, x + 1$
- p_x = eloonjäämistodennäköisyys ikävälillä $x, x + 1$
- l_x = eloonjäämistodennäköisyys iässä x
- t_x = stationaarinen väestö ikävälillä $x, x + 1$
- T_x = stationaarinen väestö iästä x ylöspäin
- d_x = kuolintodennäköisyys ikävälillä $x, x + 1$
- e_x = keskimääräinen jäljelläoleva elinaika iässä x
- \bar{K}_x = x vuotta täyttäneiden kuolleisuus vuotta kohti
- v_x = todennäköinen jäljelläoleva elinaika iässä x
- \hat{x}_d = tyypillinen elinaika
- $d(\hat{x})$ = edellistä vastaava kuolintodennäköisyys
- \bar{x}_t = stationaarisen väestön keskimääräisikä

1) $q_x = d_x / l_x = (l_x - l_{x+1}) / l_x = 1 - p_x$

2) $(ABCDE) \rightarrow q_x \cong D_x / (\bar{A}_x + E_x) = \dots (*)$

3) $p_x = l_{x+1} / l_x = 1 - q_x$

4) $l_x = l_0 \cdot p_0 \cdot p_1 \cdots p_{x-1} \quad (l_0 = 1)$
 $= \sum d_x \quad (x = x, x+1, \dots, \omega)$

5) $t_x \cong (l_x + l_{x+1}) / 2$

6) $T_x = \sum t_x = \sum l_x - l_x / 2 \quad (x = x, \dots, \omega)$

7) $d_x = l_x - l_{x+1} = q_x \cdot l_x$

8) $e_x = T_x / l_x = \sum l_x / l_x - .5 \quad (x = x, \dots, \omega)$
 $= \bar{X}_d = \sum (x + .5) \cdot d_x / \sum d_x - x \quad (x = x, \dots, \omega)$

9) $\bar{K}_x = l_x / T_x = 1 / e_x$

10) $v_x = x_i - x$, kun $l_i = l_x / 2$

- *) A = väkiluku vuoden lopussa
- \bar{A} = keskiikä
- B = syntyneet
- C = alle 1-vuotiaina kuolleet
- D = kuolleet
- E = synt.päiv. jälkeen kuolleet
- x = ikä (ω = korkein todettu ikä)

3. Suureiden tulkinnasta

Kun tehdään kuolleisuustasoa koskevia päätelmiä ja vertailuja, voi tarkastelun kohteena olla kokonaiskuolleisuus tai kuolleisuus ikäryhmittäin.

A. Kokonaiskuolleisuus

Kokonaiskuolleisuutta kuvaavista suureista on tärkein yleinen kuolleisuusluku

$$\bar{K}_0 = 1/e_0$$

Varsinaisia kokonaiskuolleisuuden mittalukuja ei muita olekaan, vaikka kuolleisuustason ollessa kysymyksessä esitetään tavallisesti elävänä syntyneiden keskimääräinen jäljelläoleva elinaika (e_0) tai vastaava todennäköinen jäljelläoleva elinaika (v_0) eli "mediaani-ikä", kuten sitä voisi myös nimittää.

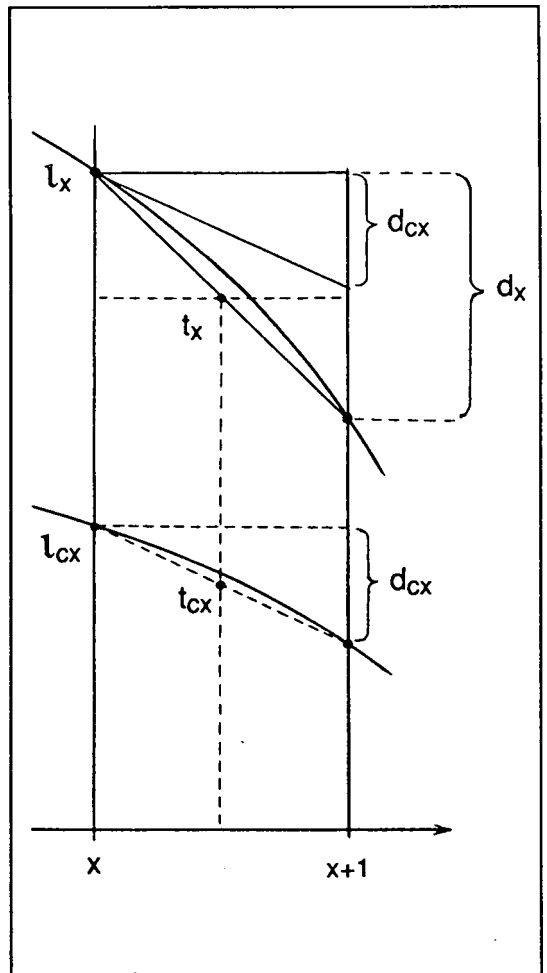
Saattaa herätä kysymys, miksi "aikaa ilmaiseva suure" voi olla kuolleisuustason mittaluku. Selityksenä tähän on matemaattinen sattuma. Koska $T_x = e_x \cdot l_x$, minkä voi nähdä kaavasta (8), niin kaavassa (9) suure l_x supistuu pois ja \bar{K}_x :n arvoksi tulee $1/e_x$ eli \bar{K}_0 :n arvoksi $1/e_0$, kuten jo todettiin. Kokonaiskuolleisuuden taso on siis yksinkertaisesti kääntäen verrannollinen elinaikaan nähden, ja sen vuoksi voidaan päätelmät tehdä elinajan nojalla laskematta varsinaista kuolleisuuskertoimta näkyviin.

B. Todennäköisyysuureista

Edellä olevaan liittyen voisi vielä kysyä, onko \bar{K}_x todennäköisyysuure tai sellaiseen verrattava suhteellinen frekvenssi, jonka arvo pysyy nollan ja yhden välillä. Jos näin olisi, niin e_x :n arvo olisi aina suurempi kuin yksi. Ikävuoteen 100 asti laskettujen taulukkoarvojen mukaan e_x jää ykköstä suuremmaksi, mutta on

osoitettavissa, että kyllin suurilla x :n arvoilla (ja q_x :n kasvaessa yli tietyn rajan) toteutuu $e_x < 1$ ja samalla $\bar{K}_x > 1$, joten siis \bar{K}_x ei ole todennäköisyysluku.

Kuolemanvaaraluku (q_x) sitä vastoin on todennäköisyysluku, kuten voi päätellä jo sen määritelmästä (kaava 1), jonka mukaan kysymyksessä on eloonjäämisfunktion suhteellista vähenemistä kuvaava kerroin. Tähän sopii myös em. vertaus otannasta ilman palautusta. Joukosta ei voida ottaa pois enempää, kuin siinä on (vrt. kuvio 1).



Kuvio 1. Kuolemanvaarakertoimen määrittäminen

Myös suure p_x on todennäköisyysluku, ja sama pitää paikkansa eloonjäämisluvusta l_x , josta käytetäänkin nimitystä "eloonjäämistodennäköisyys" silloin, kun sitä ei ole kerrottu millään 10:n potenssilla.

On huomattava, että esim. kuolemanvaara ei tule ykköstä suuremmaksi, vaikka ikäintervalliksi otettaisiin hyvinkin pitkä aikaväli. Ääritapaus on loppupäästä avoimen luokan eli x vuotta täyttäneiden kuolemanvaara, joka on aina yksi. Muutenhan kuolemanvaara on sitä suurempi, mitä pitempään ikävähliin se kohdistuu, joskaan q_x :n arvo ei ole tarkasti suhteessa intervallin pituuteen.

Lisäksi on syytä huomauttaa, että käsite "riskiaika" määritellään oppikirjoissa samaksi kuin suure t_x , joka tarkasti ilmaistuna on l_x -funktion integraali välillä $x : x + 1$, mutta joka käytännössä lasketaan kaavan (5) mukaisena likiarvona. Tämä suure ei ole todennäköisyysluku, sillä käytettäessä vuotta pitempää ikäintervallia saadaan myös (ainakin osaksi) ykköstä suurempia t_x -arvoja. Riskiaika eli "ikäryhmässä elettyjen vuosien määrä" riippuu siis kylläkin ikäintervallin pituudesta, mutta ei ole siihen tarkasti verrannollinen.

Mainittakoon tässä yhteydessä ikäryhmittäinen keskikuolleisuusluku, joka saadaan lasketuksi käyttämällä kaavaa

$$k_x = d_x/t_x \equiv D_x / \bar{A}_x$$

Tätä suuretta ei yleensä sisällytetä kuolleisuustauluihin, mutta sitä käytetään joskus apusuurena kuolemanvaaralukujen laskettaessa. Näiden kahden suureen tulkinta on usealla tavalla erilainen.

Voidaan osoittaa, että k_x ei ole todennäköisyysluku, sillä sen arvo tulee ikäasteikon loppupäässä (kun $q_x > 2/3$) ykköstä suuremmaksi. Näin on siitä huolimatta, että kysymyksessä olevan kuolleisuusriskin vaikutusajaksi määritellään

yksi vuosi silloinkin, kun käytetään vuotta pitempää ikäintervallia. Tämä pätee myös em. suureen \bar{K}_x suhteen.

Suureen q_x kohdalla riskin vaikutusaikasen sijaan on sama kuin ikäintervallin pituus. Kun q_x :n vaikutus kohorttiin l_x on verrattavissa otantaan ilman palautusta, niin k_x :n vaikutus väestöön t_x vastaa lähinnä otantaa palautuksella, jolloin väestöstä poistuneet eivät palaa siihen, vaan tilalle tulee uusia yksilöitä.

C. Laskelmat kuolinsyiden mukaan

Kuolleisuustaulujen laskeminen kuolinsyiden mukaan merkitsee sitä, että normaaliin tapaan laskettuja kuolleisuustauluja jatkovaiheena laajennetaan ottamalla mukaan uusi kvalitatiivinen dimensio (joukko kuolinsyitä). Tällä tavoin saadun tauluston suhteen pitävät suurimalta osin paikkansa samat periaatteet ja tulkinnat kuin normaalitauluston suhteen, mutta eräissä kohden on eroja, joihin on aiheellista viitata.

Edellä kohdassa 1. mainittiin ne kaksi erilaista kuolemanvaaraluvun määritelmää, jotka voivat tulla kysymykseen (Ks. myös kuvio 1). Osittaiskuolemanvaara kuolinsyyn c suhteen on

$$q_{cx} = d_{cx}/l_x$$

missä nimittäjänä (riskiväestönä) ovat kaikki elossaolevat kuolinsyistä riippumatta. Ominaiskuolemanvaara taasen saadaan kaavasta

$$\bar{q}_{cx} = d_{cx}/l_{cx}$$

jolloin riskiväestönä on kuolinsyyn c määrittämä elossaolevien osajoukko (syntymähetkellä l_{co}). Vaikka etukäteen ei tiedetä, minkä syyn johdosta kukin kuolee, on jälkikäteen (laskennallisesti) mahdollista määritellä kuolinsyiskohtaisia riskiväestöjä.

Väestöselvitysyksikön laskentaohjelmassa on asetettu etusijalle osittaiskuolemanvaarakerroin paitsi sen tulkinnallisen realistisuuden vuoksi myös siksi, että muut laskettavat suureet muodostavat juuri sen kanssa parhaiten yhteen sopivan järjestelmän. Tämä suure (q_{cx}) on siten tarkoituksenmukaisin, kun halutaan selvittää kuolleisuuden jakautumista ikäryhmittäin kuolinsyiden mukaan.

Muista suureista mainittakoon eloonjäämisluku l_{co} , yleinen kuolleisuusluku K_c ja elinaika e_{co} , jotka esitetään julkaisu-
tauluissa q_{cx} :n ohella.

Kokonaiskuolleisuuden tasoa sinänsä yllä mainituista kuvaa vain \bar{K}_c (tai \bar{K}_{co} analogisesti suureen \bar{K}_0 kanssa merkittävä). Tämä suure on additiivinen ts. \bar{K} on kaikkien osakertoimien \bar{K}_c summa.

Suure l_{co} ilmaisee, kuinka suuri suhteellinen osuus kokonaiskuolleisuudesta tulee kuolemansyyn c osalle, ja on käyttökelpoinen mittaluku kuolinsyiden kes-

kinäisiä vahvuuksia arvioitaessa. Kokonaiskuolleisuuden mittaluvuksi l_{co} ei kuitenkaan sovellu. Saattaa nim. tapahtua, että l_{co} kohoaa, vaikka kokonaiskuolleisuuden taso kaikkien kuolinsyiden mukaan laskee, jolloin myös kuolinsyyn c aiheuttama kuolleisuus ehkä tosiasiallisesti alenee.

Ei myöskään suure e_{co} mittaa kuolinsyyn c osalta kokonaiskuolleisuuden tasoa, ja se eroaa siis tässä suhteessa em. suureesta e_0 , jota voidaan epäsuorasti käyttää kuolleisuustason indikaattorina. On muistettava, että suure e_{co} on määritelmän mukaisesti "ominais"-luku, tarkemmin sanoen osakohortista l_{co} kuolleiden keskimääräisikä. Jos halutaan tarkastella jotakin tällaista osajoukkoa esim. syöpään kuolleita tai itsemurhan tekijöitä erikseen muista kuolinsyistä riippumatta, niin silloin e_{co} (tai $1/e_{co}$) kylläkin kuvaa kokonaiskuolleisuuden tasoa. Sellaisenaan e_{co} ilmaisee, missä iässä keskimäärin minkin kuolinsyyn johdosta kuolla.

4. Reaaliväestö ja stationaarinen väestö

Kuolleisuustaulut koskevat itse asiassa vain stationaarista väestöä, mutta samantapaisia laskelmia voidaan tietysti vaurausin tehdä myös reaaliväestön suhteen, ja väestörakennetta koskevissa tarkasteluissa on kiinnostavaa verrata näitä väestöjä keskenään.

Yleensä reaaliväestö ei ole stationaarinen, mutta saattaa tulla sellaiseksi, jos syntyvyys ja kuolleisuus pysyvät pitkään aikaa samansuuruisina eikä muutto- liike vaikuta väestörakenteeseen. Jos kummankin väestön ikäryhmittäiset luvut ovat käytettävissä, voidaan vastinlu-

kuja vertailemalla päätellä, poikkeako reaaliväestö oleellisesti stationaarista. Lisäksi voidaan tekemällä vertailuja eri ajankohdilta arvioida, onko reaaliväestö kehittyvässä stationaariseksi.

Ei ole mitään sellaista mittalukua, jonka avulla voitaisiin todeta, onko jokin reaaliväestö stationaarinen vai ei. On kuitenkin mahdollista tehdä asiaa koskevia päätelmiä käyttämällä elementaarisia tilastollisia tunnuslukuja.

Kaaviossa 2 on esitetty joukko tunnuslukuja, jotka kohdistuvat toisaalta väkilu-

Kaavio 2. Reaaliväestö ja stationaarinen väestö 1988

Väkiluku, Kuolleet Tunnuslukuja/ikä	Reaaliväestö		Stationaarinen väestö	
	Miehet	Naiset	Miehet	Naiset
Väkiluku	2 401 368	2 552 991	7 065 518	7 868 903
Keskiarvo	35,2	39,0	37,2	40,6
Mediaani	34,4	37,6	35,9	39,8
Hajonta	20,8	23,0	22,7	24,3
Vinous	0,119	0,183	0,167	0,103
Kuolleet	24 702	24 388	100 000	100 000
Keskiarvo	66,8	76,0	70,7	78,7
Mediaani	70,6	79,2	73,7	81,5
Hajonta	17,3	14,8	16,2	14,1
Vinous	- 0,665	- 0,645	- 0,572	- 0,590
Tyyppiarvo	77,0	81,4	77,5	83,8
Huippuordin.	0,0326	0,0435	0,0322	0,0416
Yleinen kuolleisuusluku	0,0103	0,0096	0,0142	0,0127

kujakaumaan ja toisaalta kuolleiden jakaumaan. Useat näistä luvuista tulivat jo edellä puheeksi, mutta todettakoon vielä, että esim. keskiarvo on stationaarisen väestön kuolleita koskevana sama kuin e_0 ja mediaani sama kuin todennäköinen jäljelläoleva elinaika (v_0). Yleinen kuolleisuusluku on stationaarista väestöä koskevana $1/e_0$ mutta reaali väestöä koskevana D/\bar{A} , joskin kaaviossa 2 on nimittäjänä yksinkertaisuuden vuoksi käytetty ko. vuoden (1988) väkilukua (A).

Koska väkilukujakauma ja kuolleiden jakauma ovat eri suuntiin vinoja (mikä näkyy vinousluvuihin), on niiden keskiarvoilla ja mediaaneilla vastakkainen sijaintijärjestys. Väkilukujakauman keskiarvo on mediaania suurempi, mutta kuolleiden jakaumassa mediaani on keskiarvoa suurempi ja lisäksi tyyppiarvo näitä kumpaakin suurempi.

Äsken mainitut kolme tunnuslukua ovat stationaarisen väestön kohdalla suurempia kuin reaali väestön kohdalla, mutta useita näistä eroista voisi luonnehtia pieniksi. Sama koskee kaavion muitakin lukuja, ja niinpä voidaan kysyä, eikö vuoden 1988 reaali väestö ole rakenteeltaan lähellä stationaarista.

On muistettava, että ko. tunnusluvut eivät anna mitään suoranaista vastausta yllä esitettyyn kysymykseen. Aikaisemmilta vuosilta ja 5-vuotisperiodeilta tehdyt laskelmat kuitenkin antavat aiheen olettaa, että reaali väestö on muuttumassa stationaarisen suuntaan. Toisaalta ikäryhmittäinen vertailu osoittaa, että eroa on vielä huomattavasti. Mm. sodanjälkeisten suurten ikäluokkien aiheuttama poikkeavuus reaali väestön ikärakenteessa näkyy yhä selvästi tilastoissa.

5. Tulosten luotettavuudesta

Kysymys laskelmien luotettavuudesta ei väestöselvityksissä ole yhtä keskeinen kuin eräillä muilla sektoreilla, missä joudutaan käyttämään otantamenetelmiä. Maamme väestötilastot muodostavat hyvin luotettavan tietolähteen, ja sikäli kuin niissä on virheellisyyksiä, ne ovat suhteellisesti siksi vähäisiä, ettei kuolleisuustaulujen yhteydessä ole pidetty tarpeellisena esittää luotettavuuslaskelmia.

Kaikesta huolimatta tulisi olla tietoinen julkaistujen tulosten luotettavuustasosta. Varsinkin silloin, kun laskelmat kohdistuvat osaväestöihin (lääni, kunta jne.) tai kun kohdeajanjakso on lyhyt (yksi vuosi), on otettava vakavasti huomioon satunnaisvaihtelun vaikutus tuloksiin. Myös käytetyt luokitukset vaikuttavat asiaan.

Lienee makuasia, puhutaanko "vaihtelusta" vaiko "virheestä". On todettavissa, että satunnaisvaihtelu käyttäytyy samaan tapaan kuin satunnaisvirhe, joten siihen voidaan soveltaa tilastollisia virhearvioinnin menetelmiä.

A. Kokonaiskuolleisuus

Käytännössä joudutaan tavallisesti jo ennen laskelmien suorittamista ottamaan kantaa kysymykseen, kuinka suuri kohderyhmän (sen kokonaisväestön) tulisi olla, jotta tulokset olisivat riittävän luotettavia. Vähimmäisvaatimuksena voidaan pitää, että ainakin kokonaiskuolleisuuden taso saadaan kohtuullisen luotettavasti määritetyksi. Kokonaiskuolleisuutta kuvaa epäsuorasti elinaika e_0 , joka on d_x -jakauman keskiarvo iän suhteen. Luotettavuuspäätelmissä voidaan sen vuoksi ottaa mittapuuksi keskiarvon keskivirhe

$$s(\bar{x}) = s_d / \sqrt{n}$$

missä s_d on d_x -jakauman keskihajonta ja n tapausten lukumäärä (otoskoko).

Kaaviossa 3 on esitetty joitakin suureen $s(\bar{x})$ ohjearvoja ja niitä vastaavia ryhmäkokoja. Suhteellinen keskivirhe, joka on myös merkitty näkyviin, saadaan jakamalla absoluuttinen keskivirhe itse kes-

Kaavio 3. Keskiarvon keskivirhe

Ryhmäkoko	Keskivirhe vuosina	Suhteellinen keskivirhe %
1 125 000	0,2	0,27
180 000	0,5	0,67
45 000	1,0	1,34
11 250	2,0	2,68
2 813	4,0	5,36

Olettamus: $\bar{x}_d = 74,6$ $s_d = 15,0$ $K_0 = 0,01$

kiarvolla. Kaaviossa on oletettu hajonan arvoksi 15,0 ja keskiarvoksi 74,6 vuotta, jotka edustavat vuoden 1988 arvoja molempien sukupuolten osalta. Yleisen kuolleisuusluvun (K_0) arvoksi on oletettu 0,01, minkä mukaisesti on ilmaistu vertailulukuna käytetty ryhmäko (väkiluku).

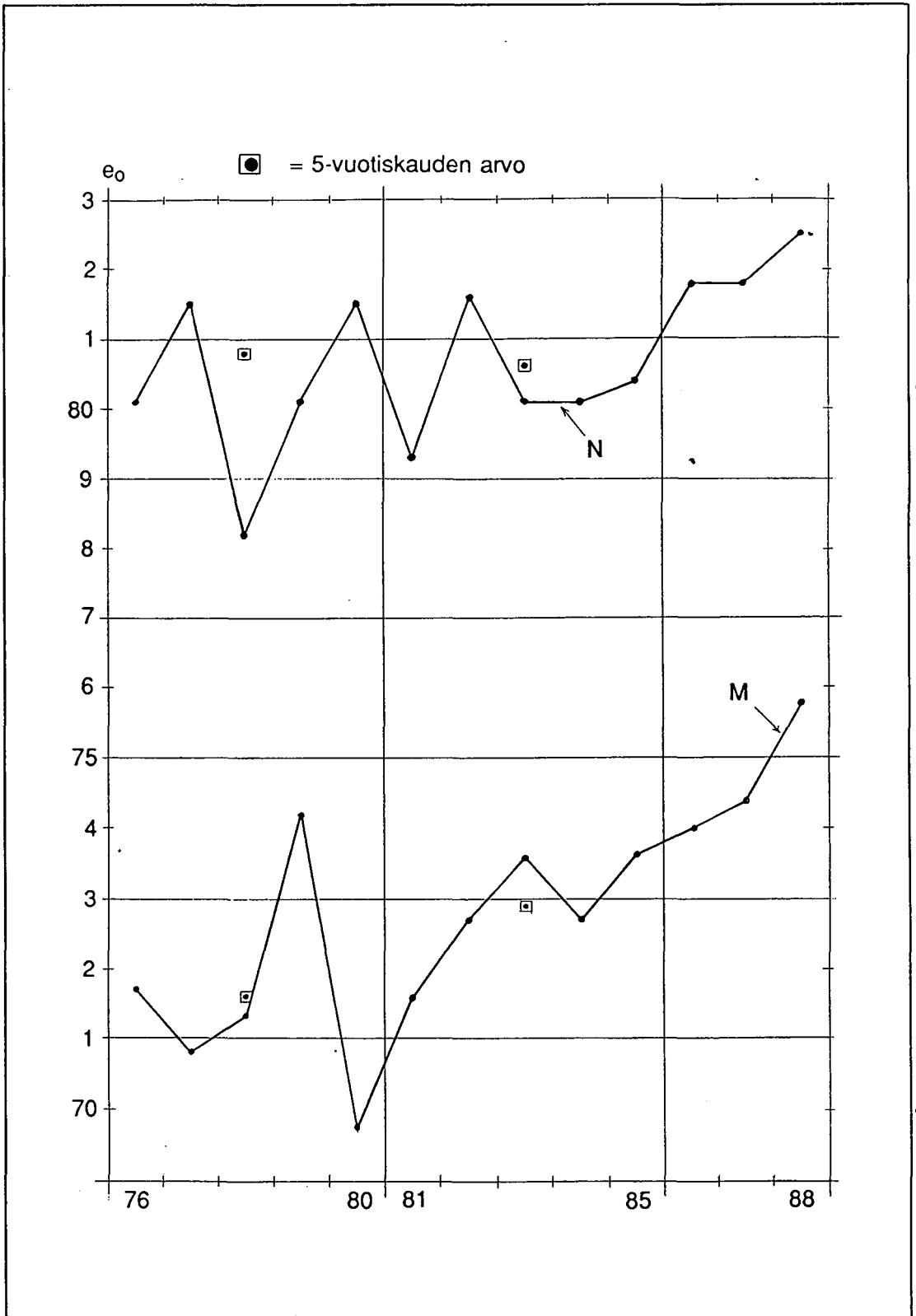
Jos elinajan e_0 keskivirheeksi sallitaan 2 vuotta, niin ryhmäkoon tulisi olla 11 250, mikä soveltuu Ahvenanmaan väestöön (24 000 v. 1988), kun miehet ja naiset otetaan kohderyhmäksi erikseen. Suhteellinen keskivirhe on tällöin 2,7 %. Kuten kuvioista 2 nähdään, on satunnaisvaihtelu Ahvenanmaan keskiarvoissa ollut sanotun suuruista vuosina 1976-88, joskin on todettava, että viimeisten 5 vuoden aikana vaihtelu on jostakin syystä ollut vähäisempää.

Kun luotettavuuspäätelyä sovelletaan maan koko väestöön, niin e_0 :n absoluuttinen keskivirhe jää kaavion arvoa 0,2 pienemmäksi sekä miehillä että naisilla. Kuviossa 3 on esitetty e_0 :n arvot vuosilta 1976-88, ja diagrammoista on nähtävissä, että satunnaisvaihtelu sinänsä pysyy arvion mukaisissa puitteissa. Tavallista suuremmat heilahtelut 1980-luvun alkupuolella johtuvat muista syistä. Esim. vuoden 1985 poikkeuksellisen kylmä talvi kohotti kuolleisuutta joissakin väestöryhmissä, ja sen johdosta e_0 aleni. Tarkemmin laskien ja käyttäen hyväksi kaavion 2 suureita vuodelta 1988 saadaan keskiarvon \bar{x}_d keskivirheeksi

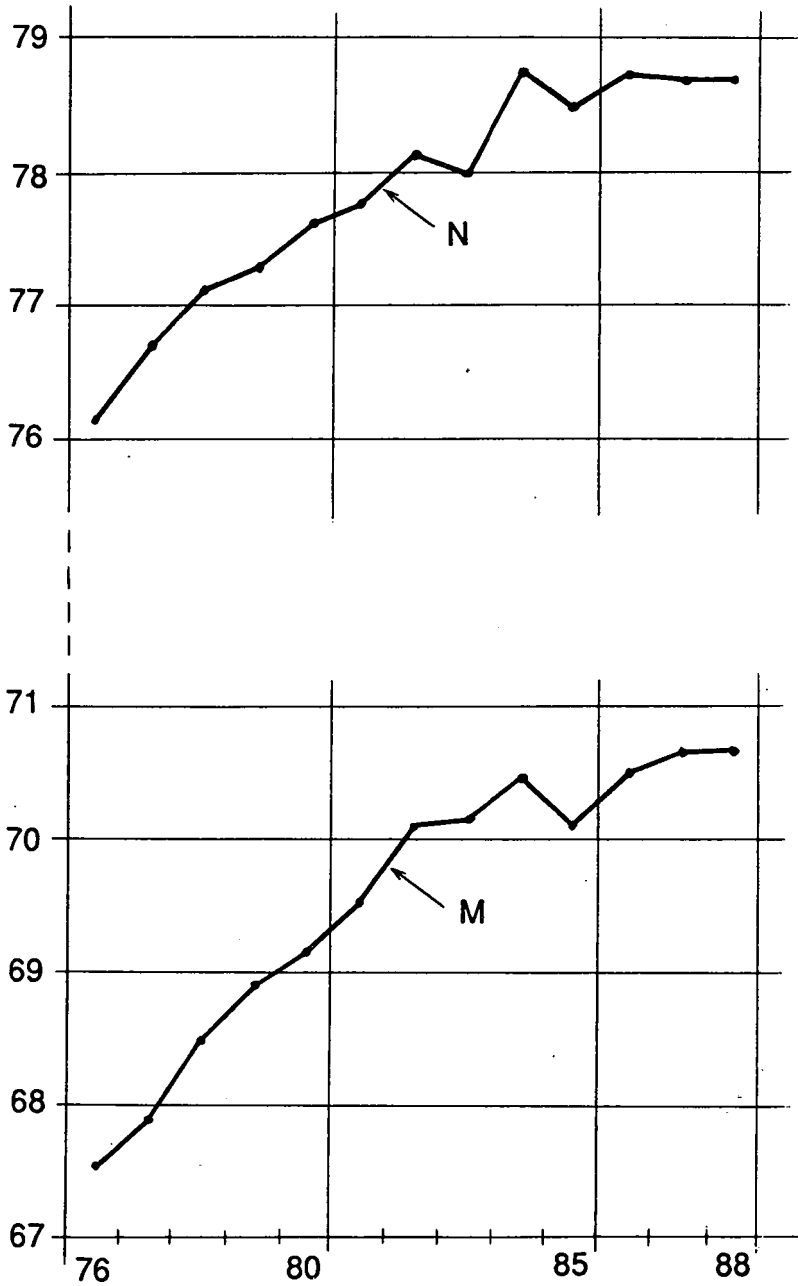
0,11 vuotta miehillä ja 0,09 vuotta naisilla.

Näiden arvioiden nojalla voidaan sanoa, että maan koko väestöä koskevissa laskelmissa e_0 saadaan riittävän luotettavana myös yksittäisiltä tilastovuosilta ja kummankin sukupuolen osalta erikseen. Läänikohtaisissa laskelmissa virhe on jo 0,5 vuotta ja enemmänkin, mutta suhteellinen virhe pysyy kuitenkin suunnilleen yhden prosentin rajoissa, jos Ahvenanmaata ei oteta lukuun. Kuolleisuustaulut on Tilastokeskuksessa laskettu vuodesta 1971 alkaen vuosittain, ja lääneittäin on vuodesta 1974 lähtien tulostettu ns. lyhennetyt taulut. Näitä tuloksia tarkasteltaessa ei tulisi rajoittua pelkästään yhden tilastovuoden lukuihin, vaan päätelmiä tehtäessä tulisi nojautua useamman vuoden tuloksiin. Siten saadaan myös käsitys satunnaisvaihtelun suuruudesta.

Laskelmien luotettavuutta voidaan parantaa valitsemalla kohdeperiodi useamman vuoden pituiseksi. Hyvin tavallinen on 5-vuotisperiodi, jota on käytetty 1930-luvulta lähtien virallisen tilaston kuolleisuustauluja laskettaessa. Tällöin tulee muutostapausten so. kuolleiden lukumäärä suunnilleen 5-kertaiseksi ja keskivirhe yli puolta pienemmäksi verrattuna yhden tilastovuoden lukemiin. Toisaalta vielä pitempi periodi (esim. 10 vuotta) ei ehkä ole yhtä hyvä siitä syystä, että sellaisen periodin sisällä saattaa esiintyä trendi tai jokin muu ilmiö, jonka toteaminen on tärkeää.



Kuvio 2. Elinlaka e_0 Ahvenanmaalla 1976-88



Kuvio 3. Elinluka e_0 , koko maa 1976-88

B. Kuolleisuus ikäryhmittäin

Ikäryhmittäisen kuolleisuuden mittaluku on kuolleisuustauluissa ennen muuta kuolemanvaarakerroin (q_x), jonka keskivirhe on

$$s(q_x) \equiv \sqrt{q_x \cdot (1 - q_x) / \bar{A}_x}$$

missä \bar{A}_x on ikäryhmittäinen keskiväkiluku. Kun $s(q_x)$ jaetaan itse kertoimella q_x , saadaan suhteellinen keskivirhe.

Samoin kuin edellä kokonaiskuolleisuuden kohdalla, voidaan myös ikäryhmittäisiä kuolleisuuslukuja laskettaessa määritellä etukäteen ryhmäkoossa siten, että saavutetaan riittävä luotettavuustaso. Olisi kuitenkin työlästä suorittaa arviointi kutakin ikäryhmää varten erikseen. On käytännöllisempää tarkistaa, onko ikäryhmittäinen keskiväkiluku (\bar{A}_x) riittävän suuri niissä ikäryhmissä, joissa keskivirhe on suurin. Samassa yhteydessä voidaan tietenkin myös päätellä, mikä olisi luotettavuuden kannalta riittävän suuri kokonaisväkiluku.

On selvää, että riittävän suuri kokonaisväkiluku on käytännössä yleensä mahdollisuus, jos ikäluokitus on yksityiskohtainen ja luotettavuudelle asetetaan vähänkin tiukempia vaatimuksia. Esim. 1-vuotisikäryhmitystä käytettäessä maamme kokonaisväestö jakaantuu siksi pieniin osaryhmiin, että suhteellinen keskivirhe on useita prosentteja, jopa kymmeniä prosentteja. Läänitasolla virhe tulee siihen verrattuna moninkertaiseksi. Virhettä voidaan tällöin pienentää käyttämällä laskelmissa 5-vuotisperiodia.

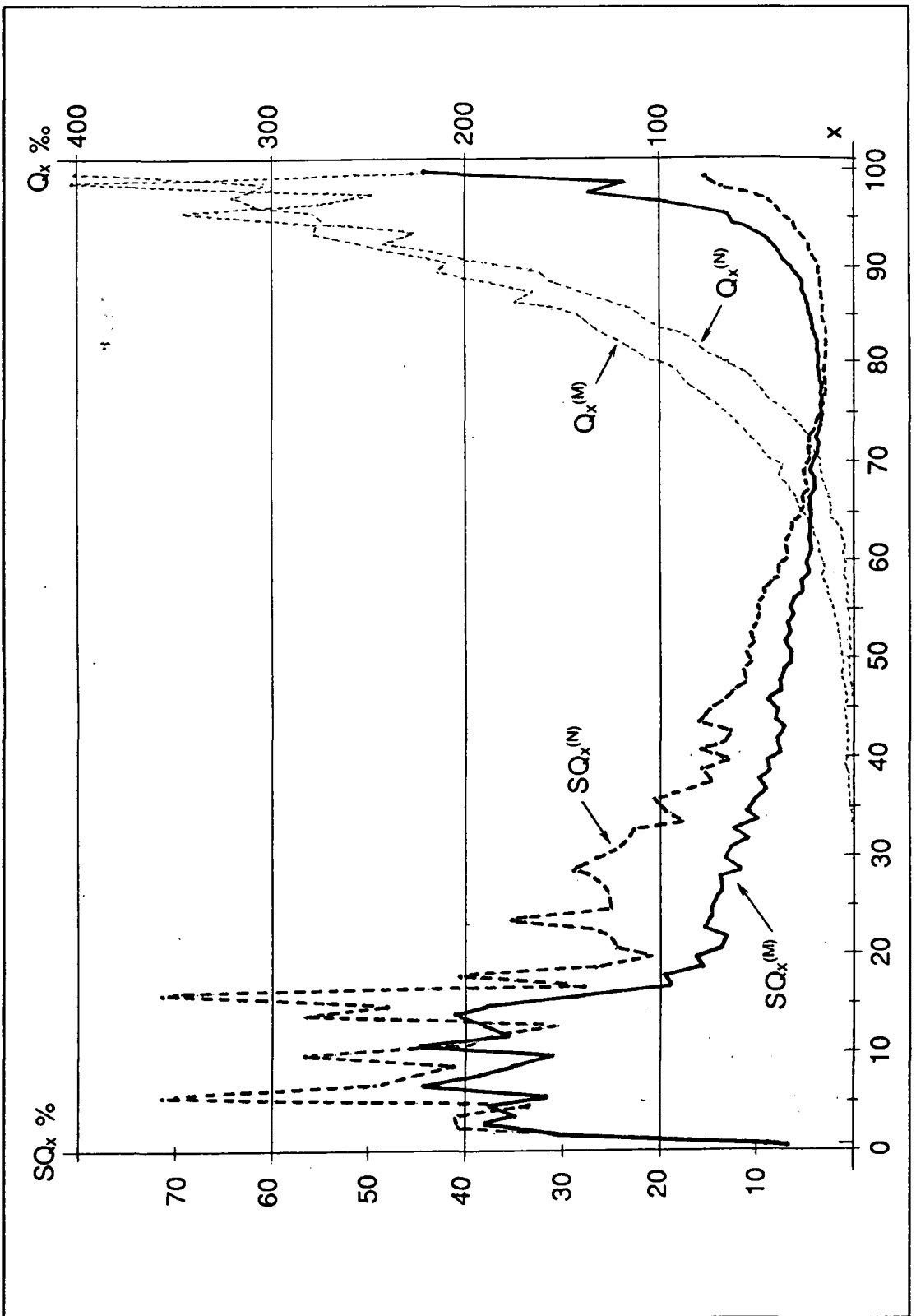
Kaaviossa 4 on esitetty suhteellisen keskivirheen arvoja erälle q_x :n arvoille ja ikäryhmittäisille keskiväkiluvuille. Jos käytetään 1-vuotisikäryhmitystä, niin q_x :n vaihteluväli on 0,0001-0,5, jolloin esim. ryhmäkoossa ollessa 100 000 suhteellinen virhe pienimmillä q_x :n arvoilla on yli 10 %. Kun todelliset väkiluvut miesten ja naisten osalta erikseen ovat huomattavasti äsken mainittua pienempiä, on sanottava, että tilastovuositain laadituissa kuolleisuustauluissa kuole-

Kaavio 4. Kuolemanvaarakertoimen suhteellinen keskivirhe %

q \ n	100	1 000	10 000	100 000	1 000 000
.0001	1 000	316	100	31,6	10
.001	316	100	31,6	10	3,2
.01	99,5	31,5	9,9	3,1	1
.1	30	9,5	3	0,9	0,3
.5	10	3,2	1	0,3	0,1
.9	3,3	1,1	0,3	0,1	0,03

n = keskiväkiluku

q = kuolemanvaarakerroin



Kuvio 4. Kuolemanvaaraluvun suhteellinen keskiarvo % v. 1988

manvaaralukujen luotettavuus ei ole erityisen hyvä. Sen vuoksi olisi ikäryhmittäisissä tarkasteluissa käytettävä mieluummin 5-vuotisperiodiin kohdistuvia tauluja.

Kuviosta 4 käy selville kuolemanvaarakertoimen suhteellinen keskivirhe vuoden 1988 tilastojen mukaan. Kuvioon on merkitty myös kuolemanvaarakertoimen arvot promilleina. Virheen jakaantumisesta iän mukaan todettakoon, että suhteellinen keskivirhe saa suurimmat arvonsa toisaalta hyvin pienillä ja toisaalta hyvin suurilla q_x -arvoilla. Suurimmillaan virhe on kuvion mukaan n. 40 % miehillä ja 60-70 % naisilla. Luultavasti virhe tulee tätäkin suuremmaksi yli 100-vuotiailla q_x :n edelleen kasvaessa. Pienimmillään virhe on niissä ikäryhmissä, joissa kuolleiden lukumäärät ovat suuria eikä väkilukukaan ole aivan pieni.

C. Vähimmäisehdot

Edellä todettiin, että tulosten luotettavuus riippuu ehdoista ja siitä, mikä mitaluku on kysymyksessä. Elinaika e_0 saadaan määritetyksi luotettavammin kuin kuolemanvaaraluvut q_x , jos ryhmäkoko on sama. On ilmeistä, että elinaika e_x , joka on kertymäsuure, tulee melko luotettavasti määritetyksi muillakin x :n arvoilla kuin nolla. Sama pätee suureen l_x suhteen. Tästä huolimatta kuolleisuustaulujen laskemiseksi tarvitaan yleensä verrattain suuri ryhmäkoko.

Otantatutkimuksissa tyydytään usein pieneen otoskookoon. Tätä perustellaan tavallisimmin sillä, että suuri otoskoko on mahdottomuus tai tulee liian kalliiksi. Perusteluna voidaan esittää myös se virheopillinen näkökohta, että jos otoskoko

ylittää tietyn vähimmäisrajan, niin sen jälkeen luotettavuus paranee koon kasvaessa enää hyvin hitaasti, eikä siis kannata pyrkiä hyvin suureen otoskookoon. Selityksenä tähän on otoskoosta (N) riippuvan virhefunktion kaarevuus. Niinpä esim. keskiarvon keskivirhe vähenee nopeasti pienillä N -arvoilla, mutta sen jälkeen (kun $N > 50$) hitaasti.

Kohderyhmän pienenus on ongelma, joka tavallisimmin tulee esille kuntakohtaisissa laskelmissa. Silloin joudutaan kysymään, mikä on ryhmäkoolelle asetettava vähimmäisraja. Olettaen, että $K_0 \cong 0,01$ ja että riittää, kun saadaan e_0 jonkinlaisella luotettavuudella lasketuksi, niin yllä mainittu otoskoko 50 edellyttää 5 000 suuruista väkilukua miehillä ja naisilla eli 10 000 yhteensä. Tällöin on ikäryhmittäisten lukujen luotettavuus jo siinä määrin heikko, että ne voidaan jättää vaille humiota.

Myös e_0 :n keskivirhe on ko. rajatapauksessa siksi suuri, että tavanomaisen luotettavuusajattelun mukaan tulokset jouduttaisiin hylkäämään. Ei kuitenkaan mikään estä käyttämästä tuloksia hyväksi nojautumalla toisenlaiseen tulkintaan. Voidaan nim. katsoa, että jokainen perustilaston tiedoista laskettu e_0 joltakin ajanjaksolta on "oikea", vaikka se vaihtelee ajankohdasta toiseen. Tämän vaihtelun taustalla oleva keskimääräinen "kuolleisuustaso" on se loppujen lopuksi oikea suure, mitä laskelmien avulla pyritään arvioimaan. Sen vuoksi olisi ehkä tarkoituksenmukaista laskea e_0 kunnittain jatkuvasti ja koota tulokset aikasarjamuotoiseksi tiedostoksi. Silloin voitaisiin seurata kuolleisuuden kehitystä kunnittain, ja samalla olisi käytettävänä aineisto tietopalvelua varten.

6. Yleispäätelmä

Se tosiasia, että kuolleisuustaulujen laskeminen merkitsee erään yleisemmän matemaattisen menetelmän soveltamista, antaa mahdollisuuden suorittaa samantapaisia laskelmia muidenkin väestöllisten muutostapahtumien suhteen. Menetelmä soveltuu esim. avioituvuuden selvittämiseen. Toisaalta myös tulosten luotettavuutta voidaan eri tapauksissa arvioida samanlaisilla menetelmillä.

Luotettavuusselvityksissä tulee ehkä useimmin kysymykseen suhteellisen frekvenssin keskivirhe, jota edelläolevassa käytettiin kuolemanvaarakertoimen keskivirheen arvioimiseen. Sitä voidaan käyttää myös syntyvyys- ja muuttoliikekertoimien luotettavuutta sel-

vitettäessä. Tilastotieteen tarjoamasta laajasta testivalikoimasta saattaa lisäksi löytyä tarkoitukseen sopivia riippuen kulloinkin esillä olevasta tapauksesta.

Mitä tulee vielä kuolleisuustaulujen laskentamenetelmään, on muistettava, että erilaisia menetelmiä (tai muunnelmia) on kehitetty lukuisasti. Esim. kuolemanvaaraluvun laskemisesta voisi kirjoittaa vaikkapa pienen kirjan. Väestöselvitysryhmän APL-ohjelmaan sisältyviä laskentavaihtoehtoja voinee ainakin tulosten luotettavuuden kannalta pitää riittävän tarkkoina. Järjestelmän kehittämistä ajatellen voisi pitää suositeltavana, että tietoaineistojen ja ohjelmien käyttö tehdään mahdollisimman vaivattomaksi.



Opaste kuolleisuus- ja eloonjäämislukujen tulkintaan

Tiedustelut:

Risto Kolari
puh (90) 17 341