

# Arbeidsnotater

S T A T I S T I S K S E N T R A L B Y R Å

Dronningensgt. 16, Oslo-Dep., Oslo l. Tlf. 41 38 20

IO 74/41

29. oktober 1974

## GLATTING AV DØDELIGHETS RATER SOM ER BASERT PÅ DATA FRA EN TOÅRS PERIODE

av

Tor Haldorsen

### INNHold

	Side
Tabellregister .....	2
Figurregister .....	2
1. Innledning .....	3
2. Modell og teoretisk bakgrunn .....	3
3. Konstruksjon og prøving av glidende gjennomsnitt .....	5
4. Valg av metode .....	14
5. Estimer for de laveste aldrer .....	16
6. Estimer for de høyeste aldrer .....	16
Referanser .....	21

## TABELLREGISTER

	Side
3.5 Korrelasjonskoeffisienter mellom estimatorene $\hat{\xi}_t$ og $\hat{\xi}_{t+s}$ , og varians for ulike metoder .....	13
6.1 Estimater for variansen til råratene for menn 1966-1970 .....	17
6.2 Råratene for kvinner 1971-1972. Promille .....	17
6.4 Gjennomsnittlig antall gjenstående leveår ved alderen $n$ beregnet ved tre ulike metoder med data fra 1966-1970 .....	20

## FIGURREGISTER

	Side
3.1 Råratene for menn 1967-1968 glattet med tre uveide gjennomsnitt. Promille .....	7
3.2 Råratene for menn 1967-1968 glattet med et femleddet glidende gjennomsnitt som bevarer 2. og 3. gradspolynom. Promille ...	8
3.3 Råratene for menn 1967-1968 glattet med et 21-ledds glidende gjennomsnitt som bevarer 2. og 3. gradspolynom. Promille ...	10
3.4 Råratene for menn 1967-1968 glattet med Spencers 21-leddsformel. Promille .....	11
4.1 Råratene for menn i tre to-årsperioder. Promille .....	15
6.3 Råratene for de eldste aldersklassene med data fra 1961-1972. Promille .....	19

## 1. INNLEDNING

Når en regner ut dødelighetsrater for ulike aldersklasser og setter ratene inn i et diagram med alder langs den vannrette akse, vil kurven som kan trekkes, bli ujevn og hakket. En forestiller seg at den "virkelige" dødeligheten ville blitt avbildet i en glatt kurve, og en betrakter derfor de utregnede rater som råestimater for den "virkelige" dødeligheten og bruker dem til å konstruere estimater som gir en glattere kurve.

Byrået beregner dødelighetstabeller på grunnlag av data om levende og døde i fem-årsperioder. De rater som beregnes av dette materialet gir såpass regelmessige kurver at en for de fleste aldrer kan nøye seg med en meget mild form for glatting. For bedre å følge med i dødelighetsutviklingen vil en nå produsere dødelighetstabeller basert på data fra to-årsperioder. For at kurvene skal bli rimelig glatte med data fra en to-årsperiode, vil en trenge en annen glattingsmetode enn den som brukes på data fra fem-årsperioder. I dette notat redegjøres for arbeidet med å finne fram til en slik glattingsmetode.

Vi begrenset oss til å se på metoder som bygger på glidende gjennomsnitt. Om disse sier Sverdrup (1973) s. 344: "Det er dette faktum at det glidende gjennomsnitt er så robust overfor à priori spesifikasjoner som gjør det så anvendelig i praksis. I så henseende står de glidende gjennomsnitt i klasse med ikke-parametriske metoder."

I kapittel 2 gir vi teoretisk bakgrunn for konstruksjon av noen glidende gjennomsnitt, mens kapittel 3 og 4 er kommentarer til en empirisk utprøving av disse og noen andre tradisjonelt brukte glidende gjennomsnitt. I kapittel 5 og 6 gir vi forslag til løsning av en del spesielle problemer i glattingen av ratene for de laveste og høyeste aldersklassene.

## 2. MODELL OG TEORETISK BAKGRUNN

Anta at råratene er realisasjoner av stokastiske variable  $X_t$  for  $t=1, \dots, 105$ . La  $\xi_t$  være den "virkelige" dødelighet for alder  $t$  og anta at  $X_t$ -ene er uavhengige med

$$EX_t = \xi_t,$$

og var  $X_t = \sigma^2.$

Vi ønsker å estimere  $\xi_t$ -ene på grunnlag av  $X_t$ -ene og begrenser oss til å se på estimatorene av formen  $\hat{\xi}_t = \sum_{v=-a}^b r_v X_{t+v}$ , der vektene  $r_{-a}, r_{-a+1}, \dots, r_b$  er konstanter uavhengig av  $t$ . Slike estimatorene kalles glidende gjennomsnitt. Til å estimere  $\xi_t$  bruker vi altså en veid sum av  $a+b+1$  observasjoner. Foruten  $X_t$  brukes observasjonene fra  $a$  aldersgrupper før  $t$  og  $b$  aldersgrupper etter  $t$ . Vi betegner de ulike glidende gjennomsnitt ved vektene i gjennomsnittet,  $(r_{-a}, r_{-a+1}, \dots, r_0, \dots, r_{b-1}, r_b)$ . Når  $a=b$  og  $r_{-v}=r_v$  for  $v=1, \dots, a$  kan vi nøye oss med å skrive  $(r_{-a}, r_{-a+1}, \dots, r_0, \dots)$ , idet leddene til høyre for  $r_0$  da er gitt. Tallet  $a+b+1$  angir det glidende gjennomsnitts lengde og  $(a,b)$  bestemmer plasseringen. Hvis  $a=b$  så sier vi at gjennomsnittet er sentrert. Hvis  $a=b$  og  $r_v=r_{-v}$  for  $v=1, \dots, a$ , er gjennomsnittet symmetrisk.

Vårt problem er å bestemme  $a, b$  og  $r_v$ -ene på en hensiktsmessig måte. Det er vanlig i statistisk estimering å kreve at estimatorene er forventningsrette. I vår situasjon vil det være å forlange at  $E\hat{\xi}_t = \xi_t$ . Siden vi har antatt at  $EX_t = \xi_t$  er dette det samme som å kreve

$$(2.1) \quad \sum_{v=-a}^b r_v \cdot \xi_{t+v} = \xi_t.$$

Hvis vi visste at  $\xi_t$ -ene lå på en av kurvene i en kjent klasse, kan vi av ligning (2.1) utlede eksakte krav til  $r_v$ -ene for at estimatoren skal være forventningsrett. Nå er det ikke gitt at "virkelig" dødelighet lar seg uttrykke ved hjelp av en spesiell klasse funksjoner, men vi har et stort empirisk materiale som kan gi grunnlag for forslag til gode tilnærminger. Vi vil se på en klasse tilnærmingskurver og kreve at estimatorene er forventningsrette hvis  $\xi_t$ -ene ligger på en av kurvene i denne klassen. De krav vi satte til klassen av tilnærmingskurver, var at kurvene skulle gi en brukbar tilnærming og at den var lett å arbeide med.

Som klasse tilnærmingskurver valgte vi de som er slik at  $a+b+1$  på hverandrefølgende punkter ligger på en annengradskurve. Merk at vi ikke krever at hele  $\{\xi_t\}_{t=0}^{105}$  ligger på en annengradskurve. Med denne klasse kurver er kravet som ligning (2.1) setter til  $r_v$ -ene ekvivalent med

$$(2.2) \quad \begin{aligned} \sum_{v=-a}^b r_v &= 1 \quad \text{og} \\ \sum_{v=-a}^b v \cdot r_v &= 0 \quad \text{og} \\ \sum_{v=-a}^b v^2 \cdot r_v &= 0. \end{aligned}$$

En sier ofte at metoder som tilfredsstillter disse krav bevarer annengrads-  
polynomer eller slipper ethvert annengradspolynom uendret igjennom. Hvis  
en metode tilfredsstillter  $\sum_{v=-a}^b v^3 \cdot r_v = 0$  i tillegg til ligningene i (2.2),  
vil den bevare tredjegradspolynomer. Dette er spesielt tilfelle med symme-  
triske gjennomsnitt som tilfredsstillter (2.2).

Selv om  $a$  og  $b$  er gitt, vil det være flere metoder som slipper et-  
hvert annengradspolynom uendret igjennom. For å kunne velge blant disse  
ser vi på estimatorenes varians. Siden vi har antatt at  $X_t$ -ene er uavhengige  
og har den samme varians,  $\sigma^2$ , så blir  $\text{var}(\hat{\xi}_t) = \text{var}\left(\sum_{v=-a}^b r_v X_{t+v}\right) = \sigma^2 \cdot \sum_{v=-a}^b r_v^2$ .  
Vi søker en metode som gir en glattere kurve enn  $X_t$ -ene gir, og dette kan  
vi oppnå ved å finne en metode som demper de tilfeldige variasjonene i  
 $X_t$ -ene. For at dette skal skje må metoden være slik at summen,  $\sum_{v=-a}^b r_v^2$ , er  
mindre enn 1, og vi søker de metoder der summen er så liten som mulig.

$\sum_{v=-a}^b r_v^2$  gir uttrykk for et glidende gjennomsnitts variansreducerende evne,  
dess mindre summen er, dess større er den variansreducerende evnen. Den  
variansreducerende evne til et glidende gjennomsnitt gir et indirekte mål  
for glattheten til kurvene som metoden gir. Siden ikke noe direkte mål  
peker seg naturlig ut, bruker vi det glidende gjennomsnitts variansredu-  
serende evne som kriterium til å velge blant de metoder som bevarer annen-  
gradspolynomer.

### 3. KONSTRUKSJON OG PRØVING AV GLIDENDE GJENNOMSNIITT

La  $n$  stå for lengden av det glidende gjennomsnitt. For hvert mulig  
verdipar  $(n, a)$ , når  $n = 5, 7, \dots, 21$  og  $a = n-1, n-2, \dots, n/2-1/2$ , fant vi  
det glidende gjennomsnitt som hadde størst variansreducerende evne av dem  
som tilfredsstilte (2.2). Dette gav oss 63 ulike glidende gjennomsnitt.  
Tolv av disse ble valgt ut til videre prøving. Da vi var interesserte i  
å undersøke hvordan egenskapene til de glidende gjennomsnitt avhenger av  
lengden, hadde vi blant de tolv, sentrerte glidende gjennomsnitt av lengdene  
5, 7, ..., 21.

Det er nødvendig med en empirisk undersøkelse av egenskapene til  
de tolv glidende gjennomsnittene av flere grunner:

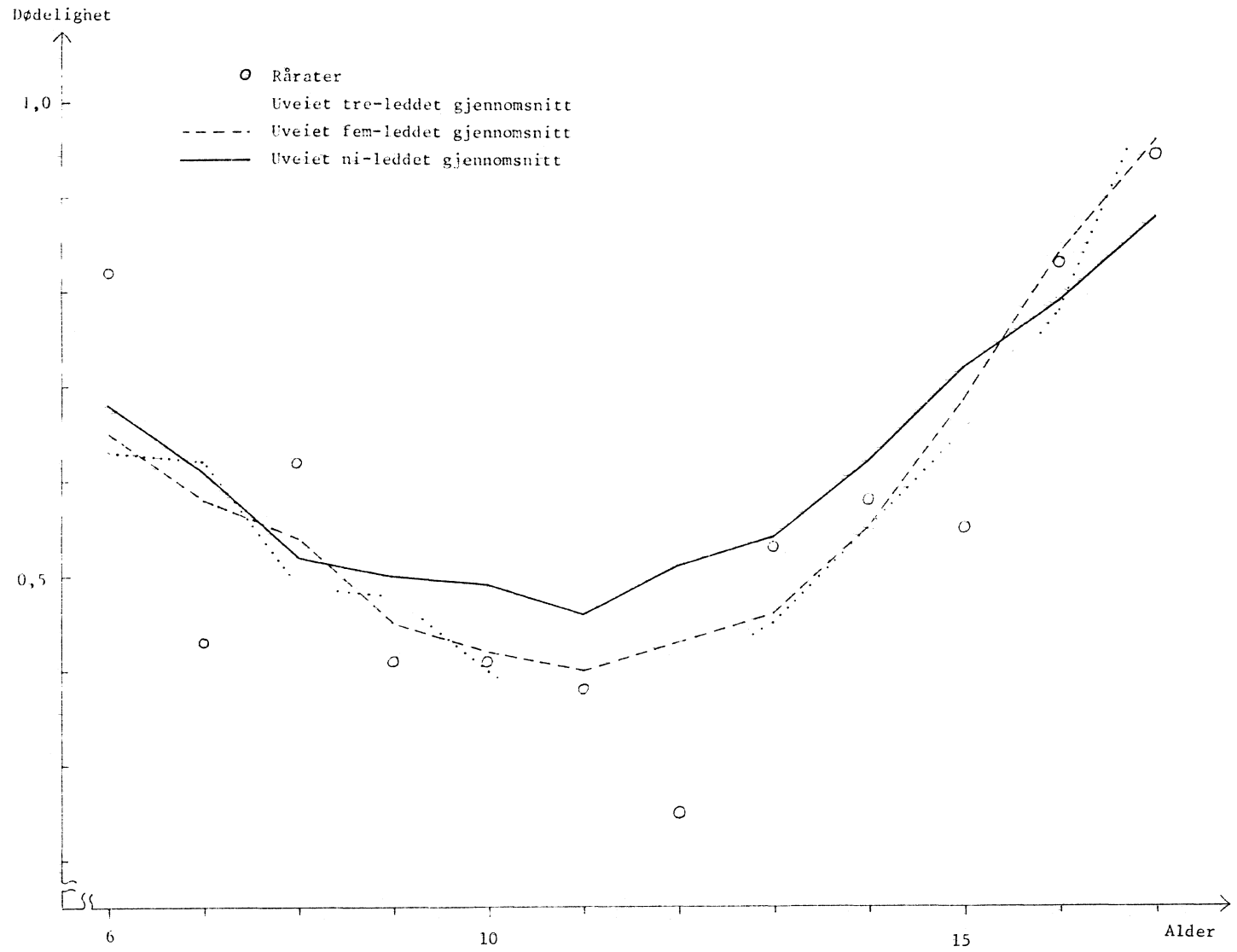
- (1) Vi vet at de har gunstige egenskaper hvis den "virkelige" dødelighetskurve er en 2.-grads evt. 3. gradskurve sett  $a+b+1$  aldrer om gangen. Men den "virkelige" dødelighetskurve er høyst tilnærmet en slik kurve.
- (2) Det er ikke gitt at den metode som har størst variansreducerende evne, gir den kurve som er glattest og best etter optisk bedømming.

Vi ville også prøve fire uveide glidende gjennomsnitt og fire glidende gjennomsnitt som erfaringsmessig har vist seg egnet til glatting av dødelighetskurver. De fire sistnevnte vil vi kalle "aktualformler" når de omtales. Vi hadde således tjue formler til videre prøving. Vi anvendte disse på råratene for menn og for kvinner for hver av de tre 2-årsperiodene 1967-1968, 1969-1970 og 1971-1972. Vi kunne dermed observere hvordan hver av metodene virket på seks ulike sett råratere. En del av resultatene ble tegnet inn i diagram sammen med råratene for nærmere undersøkelse.

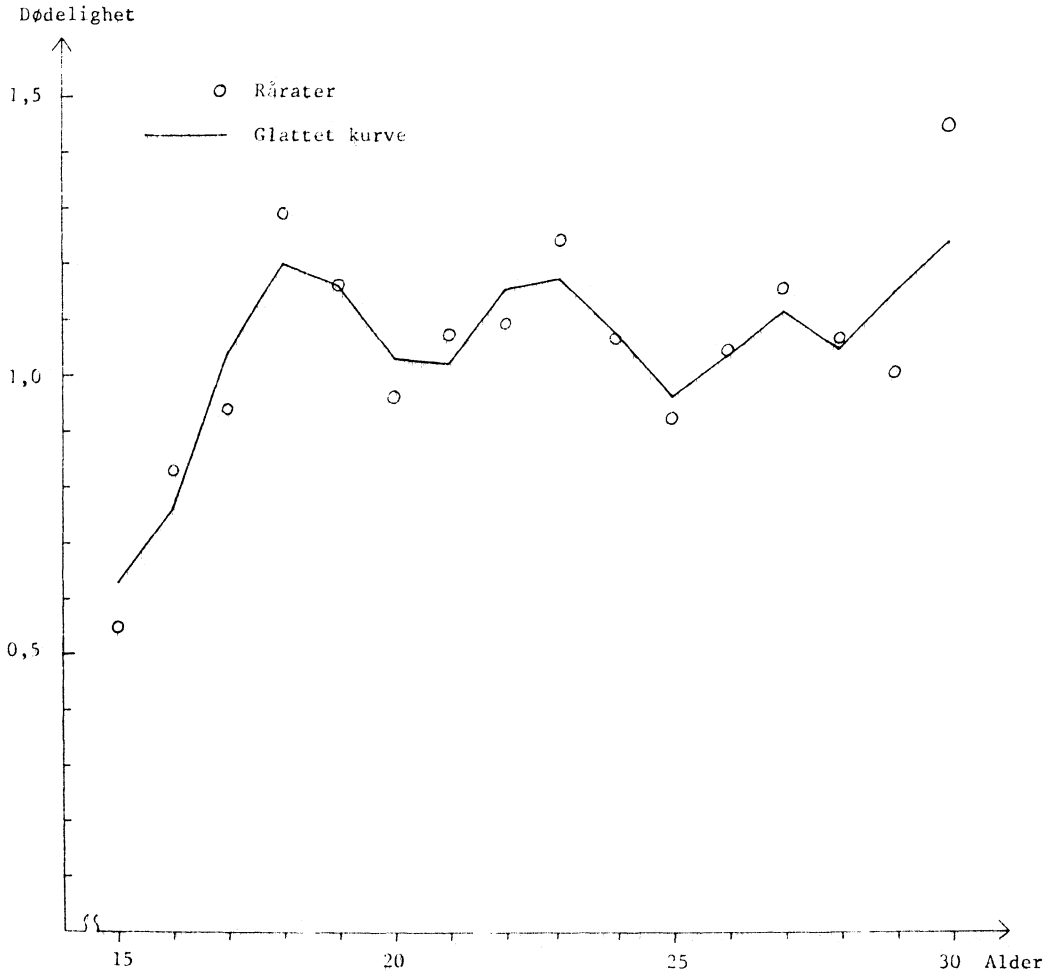
I Byråets beregninger til dødelighetstabeller basert på data om døde fra 5-årsperioder, bruker en det glidende gjennomsnitt  $(1/3, \underline{1/3}, 1/3)$  for å glatte ratene for aldersklassene f.o.m. 2 år t.o.m. 94 år. Dette er et 3-leddet uveid gjennomsnitt, og det kunne tenkes at et 5, 7 eller 9-leddet uveid gjennomsnitt ville gi gode resultater på 2-årsdata. Det gjør de imidlertid ikke. Uveide gjennomsnitt slipper førstegradspolynomer (rette linjer), men ikke annengradspolynomer, uendret igjennom. Den "virkelige" dødelighetskurven kan med en viss rimelighet sies å være en rett linje når man ser på tre aldrer om gangen, men når syv eller ni aldrer sees under ett, må en i hvert fall ha et annengradspolynom som tilnærmingskurve. Figur 3.1 viser ulike uveide glidende gjennomsnitt anvendt på råratere.

Vi må altså bruke glidende gjennomsnitt som i hvert fall slipper annengradspolynomer uendret igjennom. Når slike anvendes på råratene, avsløres et annet problem. Metoder som bruker for å få observasjoner om gangen, vil gi estimater som følger råratene for godt, dvs. de eliminerer for lite av det vi må kunne anta er tilfeldige variasjoner i råratene. Dette forholdet er illustrert i figur 3.2. Det glidende gjennomsnitt som der er anvendt, er  $(-\frac{3}{35}, \frac{12}{35}, \frac{17}{35}, \dots)$ . Dette har størst variansreducerende evne blant de med  $a=2$  og  $a+b+1=5$  som slipper annengradspolynom uendret igjennom. Siden  $r_{-2}=r_2$  og  $r_{-1}=r_1$ , vil det også slippe tredjegrads polynomer uendret igjennom.

Figur 3.1. Rårater for menn 1967-1968 glattet med tre uveide gjennomsnitt. Promille



Figur 3.2. Råater for menn 1967-1968 glattet med et femleddet glidende gjennomsnitt som bevarer 2. og 3. gradspolynom. Promille





Av de tolv glidende gjennomsnitt vi konstruerte ser det også ut som de med mange (17-21) ledd gir fortegninger. I figur 3.3 gjengir vi resultatet av å anvende det glidende gjennomsnitt  $\frac{1}{3059}$  (-171, -76, 9, 84, 149, 204, 249, 284, 309, 324, 329, ...) på et sett rårater. Det har størst variansreducerende evne blant de med  $a+b+1=21$  og  $a=10$  som slipper annengradspolymer uendret igjennom. Det synes som dette glidende gjennomsnitt gir en utflating av kurven som vi mener ikke er reell.

Det fins en rekke glidende gjennomsnitt som har vært anvendt til å glatte dødelighetskurver. Av disse prøvde vi følgende fire:

Spencers 17 ledds-formel:

$$\frac{1}{175} (-4, -3, -1, 2, 6, 19, 23, 30, \underline{31}, \dots),$$

Spencers 21-ledds-formel:

$$\frac{1}{350} (-1, -3, -5, -5, -2, 6, 18, 33, 47, 57, \underline{60}, \dots),$$

Woolhouses formel:

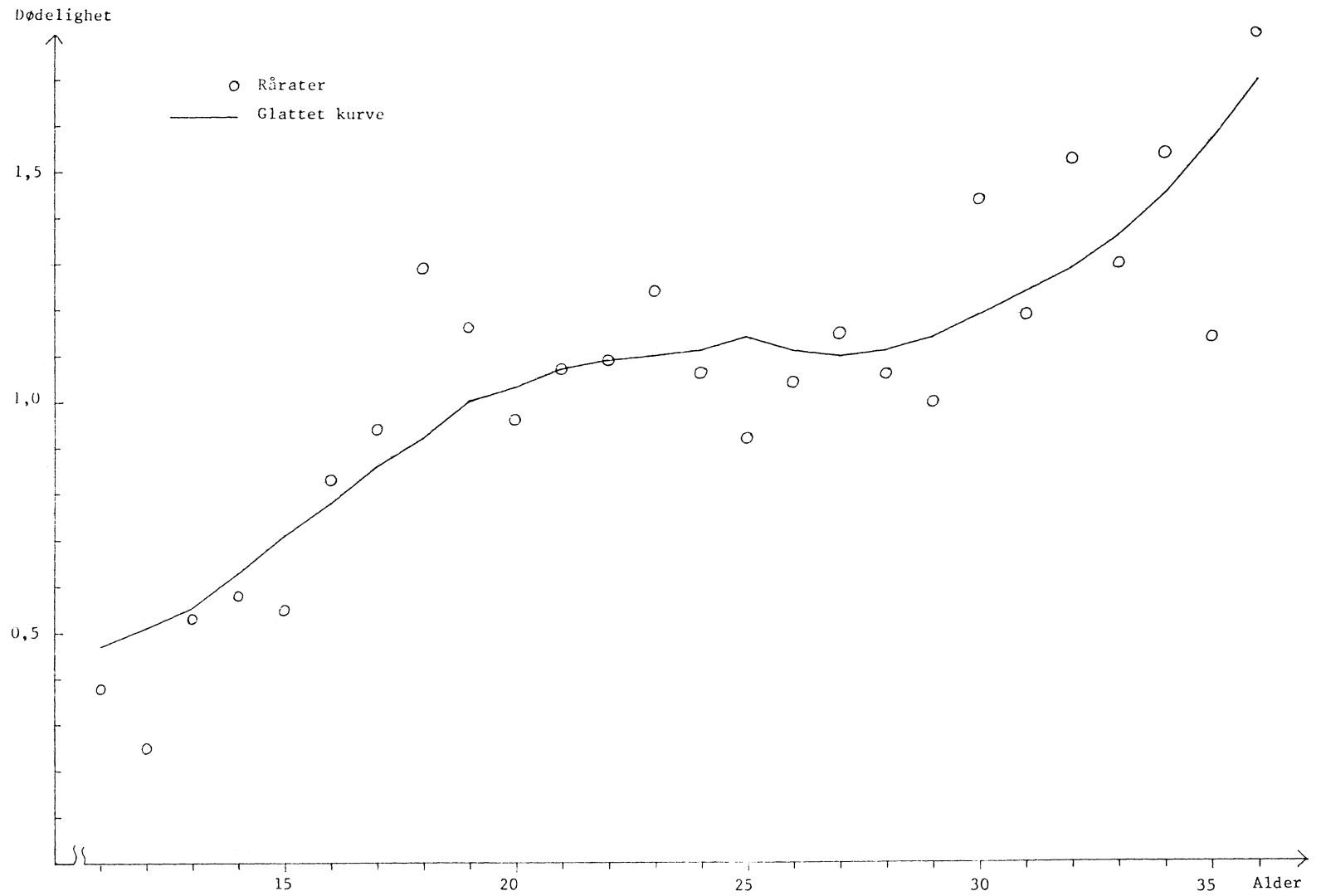
$$\frac{1}{125} (-3, -2, 0, 3, 7, 21, 24, \underline{25}, \dots),$$

Karups formel:

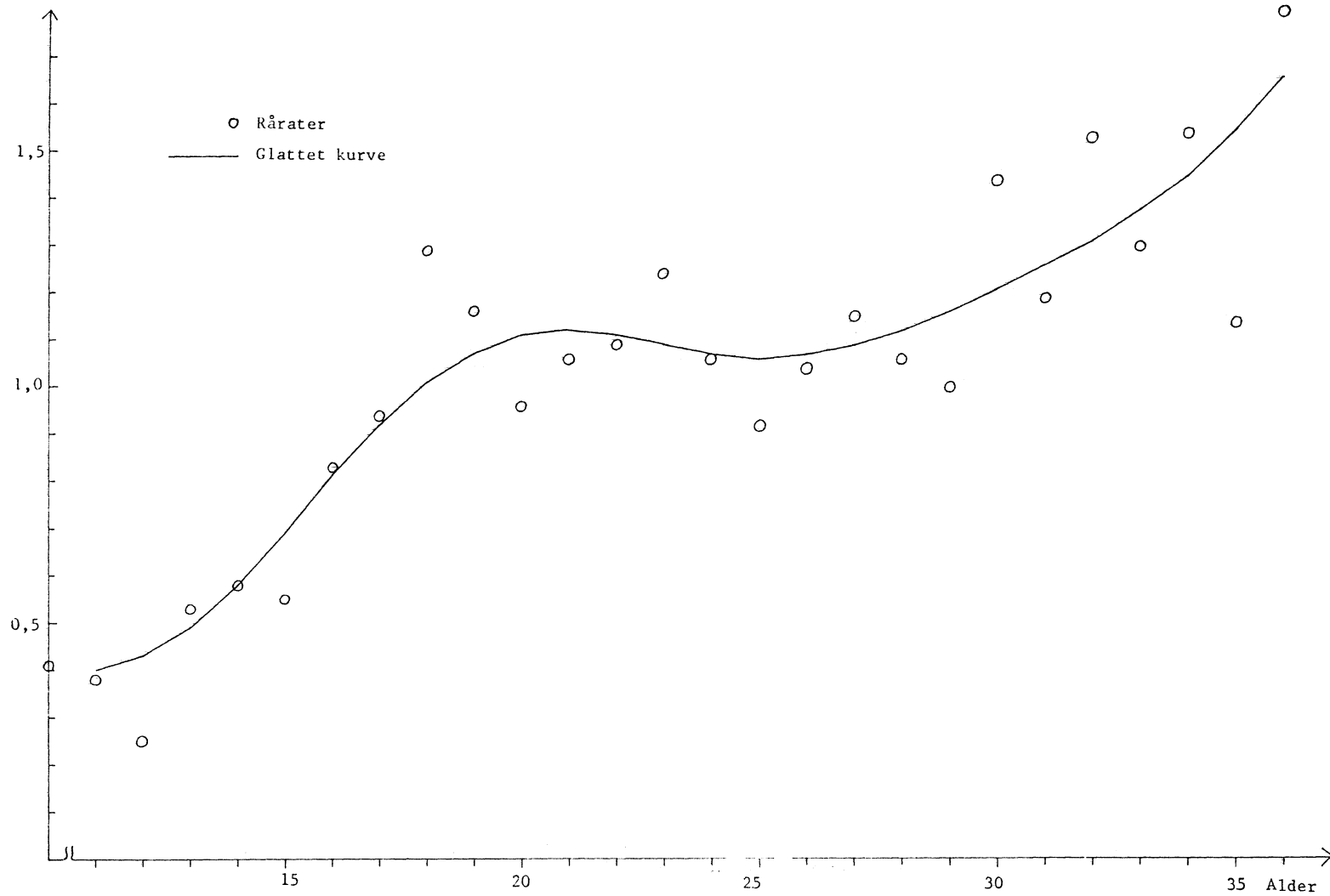
$$\frac{1}{625} (-2, -6, -9, -8, 0, 21, 53, 87, 114, \underline{125}, \dots).$$

De gav glatte kurver uten vesentlig fortegning. I figur 3.4 har vi tegnet inn resultatet av å anvende Spencers 21-leddsformel på de samme rårater som vi glattet ved en annen 21-leddsformel i figur 3.3. Spencers 21-leddsformel har minst variansreducerende evne, men vi holder det for sikkert at den gir den riktigste og kanskje også den glatteste kurve.

Figur 3.3. Rårater for menn 1967-1968 glattet med 21-leds glidende gjennomsnitt som bevarer 2. og 3. grads polynomer. Promille



Figur 3.4. Rårater for menn 1967-1968 glattet med Spencers 21-ledds formel. Promille



Vi sammenliknet også de andre tre "aktuarformlene" med de sentrerte av samme lengde som hadde størst variansreduserende evne, og vi fant at de tre "aktuarformlene" gav de glatteste kurver.

Dette viser at glattheten på den glattede kurve ikke bare kan avhenge av hvilken dempning av de tilfeldige variasjoner vi får for den enkelte alder. En kan også si at glattheten vil avhenge av hvilken sammenheng det er mellom estimatene for aldrer som er nær hverandre. En slik sammenheng kan f.eks. måles ved korrelasjonskoeffisienten mellom  $\hat{\xi}_t$  og  $\hat{\xi}_{t+s}$ , estimatorene en metode gir for aldrer  $t$  og  $t+s$ . Størst betydning for glattheten har det at koeffisienten er stor for  $s=1$ . I tabell 3.5 finner en korrelasjonskoeffisientene for ulike  $s$  for de fire "aktuarformlene" og for de fire glidende gjennomsnitt de ble sammenliknet med.

Vi ser at "aktuarformlene" har større korrelasjonskoeffisient for  $s=1$  enn de som de sammenliknes med. Det kan være forklaringen på at de gir glattere kurver. Et felles trekk ved "aktuarformlene" er at korrelasjonskoeffisienten avtar raskt når  $s$  øker, og det kan være årsak til at de gir mindre utflating av topper på dødelighetskurven enn de glidende gjennomsnitt de sammenliknes med.

Tabell 3.5. Korrelasjonskoeffisienter mellom estimatorene  $\hat{\xi}_t$  og  $\hat{\xi}_{t+s}$ , og varians for ulike metoder

	s=										Varians
	1	2	3	4	5	6	7	8	9	10	
Woolhouses formel .....	0,913	0,741	0,520	0,286	0,077	-0,025	-0,073	-0,078	-0,055	-0,019	$0,179x\sigma^2$
15-ledds sentrert variansminimerende ....	0,905	0,769	0,608	0,439	0,276	0,133	0,018	-0,061	-0,102	-0,106	$0,151x\sigma^2$
Spencers 17-leddsformel	0,940	0,814	0,625	0,418	0,216	0,071	-0,036	-0,077	-0,084	-0,065	$0,153x\sigma^2$
17-ledds sentrert variansminimerende ....	0,919	0,804	0,667	0,519	0,372	0,233	0,111	0,014	-0,056	-0,097	$0,133x\sigma^2$
Karups formel .....	0,944	0,790	0,577	0,351	0,156	0,018	-0,056	-0,078	-0,067	-0,044	$0,163x\sigma^2$
19-ledds sentrert variansminimerende ....	0,930	0,831	0,713	0,583	0,450	0,320	0,200	0,095	0,010	-0,053	$0,119x\sigma^2$
Spencers 21-leddsformel	0,957	0,836	0,660	0,461	0,269	0,111	0,000	-0,061	-0,082	-0,074	$0,143x\sigma^2$
21-ledds sentrert variansminimerende ....	0,938	0,852	0,748	0,634	0,514	0,394	0,279	0,174	0,083	0,007	$0,108x\sigma^2$

#### 4. VALG AV METODE

Valget av glattingsformel blant dem vi betrakter, må bli et forsøk på å finne den gyldne middelvei. En må finne en formel som er av en viss lengde, slik at den ikke følger tilfeldighetene for godt, samtidig som den ikke er så lang at den "retter ut" reelle variasjoner i dødelighetskurven. Vi valgte det 9-leddete sentrerte glidende gjennomsnitt som har størst variansreducerende evne blant dem av sin lengde og plassering som slipper annengradspolynomer uendret igjennom. Siden dette er symmetrisk, vil det også bevare tredjegradspolynomer. Formelen vi bruker er  $\frac{1}{231}$  (-21, 14, 39, 54, 59, ...). Denne formelen vil bli brukt for aldrene 5,6..., 87 og 88 år. Senere skal vi finne egne formler for de yngste og eldste aldersklassene.

Denne formel gir kurver som er omtrent like glatte som de en får av 5-årsdata når en bruker 3-leddet gjennomsnitt.

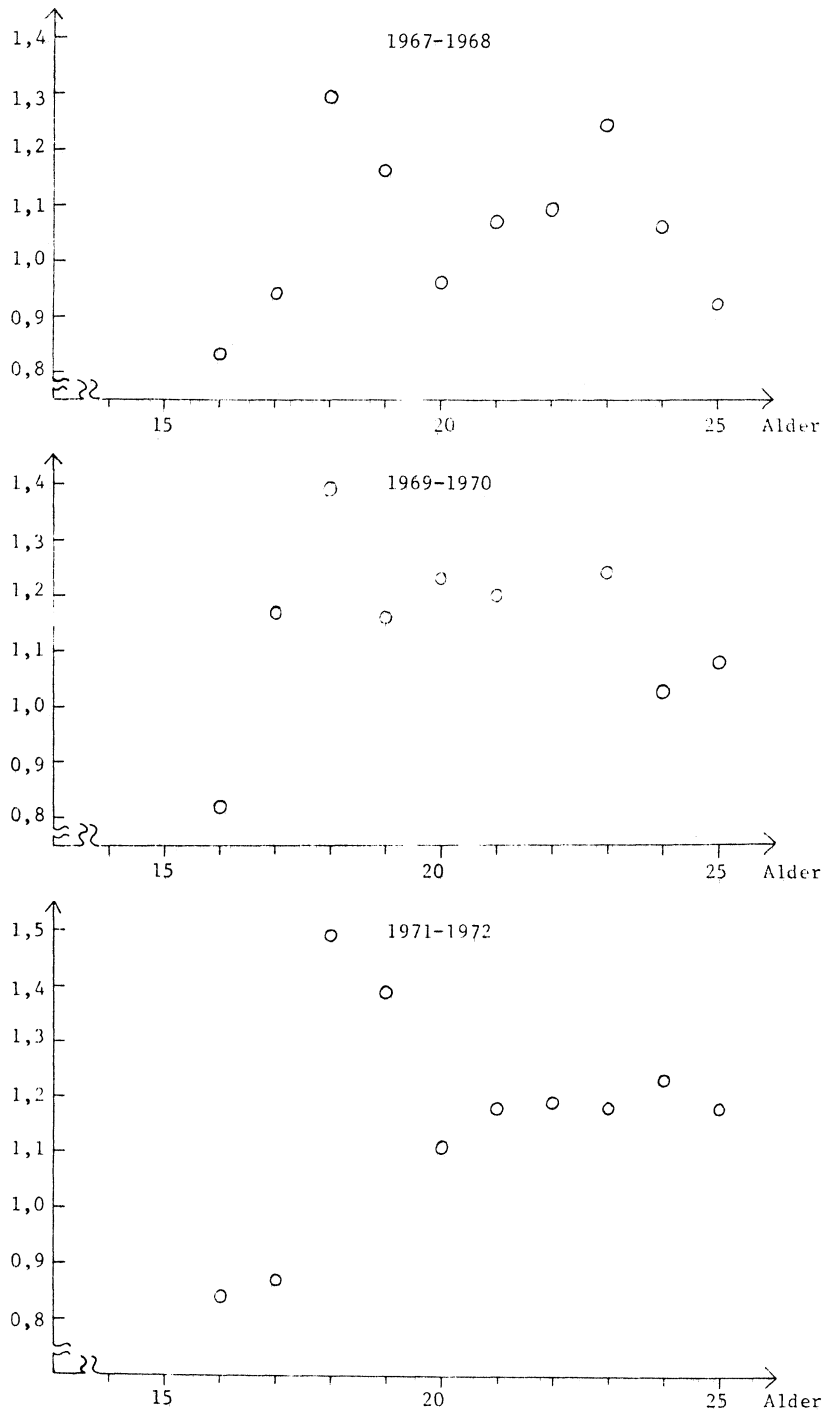
Hvis de råratene vi skulle glatte hadde vært mer uregelmessige, ville vi antakelig ha foretrukket å bruke en av de fire "aktuarformlene".

Vi tror at biologisk sett er det intet som skulle gi brå uregelmessigheter i "virkelig" dødelighet, men det kan være sosiale forhold som bevirker slike. Et slikt forhold mener vi gjør seg gjeldende for 18-årige menn. Disse har, når en ser på råratene for periodene 1967-1968, 1969-1970 og 1971-1972 en markert høyere dødelighet enn aldersgruppene ved siden av. Av figur 4.1 kan vi se at dette forhold er typisk for alle de tre periodene, og da antar vi at det ikke kan bero på tilfeldigheter. En forklaring kan være 18-årsgrensen for førerkort for bil.

Det viste seg vanskelig å bevare dette forhold ved de glattingsmetodene vi anvendte. Samtlige metoder forskjøv toppunktet på kurven mot aldersklassene 20-21 år og gav en verdi for 18-åringer som antakelig er for lav. Valget av glattingsformel ble delvis bestemt ved at den foreslåtte formel er blant dem som gir en tilfredsstillende glatt kurve, og som gir minst fortegning av det nevnte forhold.

En bør være oppmerksom på at det nevnte hopp i dødeligheten er på 0,2-0,3 promille, og tilsvarer bare 6-9 døde personer i et årskull.

Figur 4.1. Rårater for menn i tre to-årsperioder. Promille



## 5. ESTIMATER FOR DE LAVESTE ALDRER

Med den formelen vi valgte, kan vi gi glattede estimater for aldrene ned t.o.m. 5 år når vi har råestimater for aldrene 1, 2, ... (0-åringer har spesiell høy dødelighet og vi ser bort fra denne aldersklassen). Vi vil også glatte ratene for aldrene 2, 3 og 4 år og må finne en metode for dette.

Foruten å gi en rimelig glatting av ratene må metoden være slik at vi får en jamn overgang mellom ratene for 4- og 5-åringer, som er funnet ved hver sin formel. For de tre aldrene brukte vi det glidende gjennomsnitt  $\frac{1}{330}$  (126, 92, 63, 39, 20, 6, -3, -7, -6) som har størst variansreducerende evne blant dem med sin plassering som bevarer annengradspolynomer. Korrelasjonskoeffisienten mellom estimatorene vi nytter for 4- og 5-åringer blir 0,818. Denne er stor nok sammenliknet med 0,685 for aldrer ved siden av hverandre med formelen for 2, 3 og 4 åringer, og 0,814 for formelem for 5, 6, ... åringer.

Utprøving på materialet gav ingen hopp eller uregelmessigheter mellom estimatene for 4- og 5-åringer.

## 6. ESTIMATER FOR DE HØYESTE ALDRER

Ved glatting av ratene for de eldste aldersklassene har vi et annet problem. Vi har i beregningene til nå antatt at variansen til råratene som inngår i beregningene, har vært den samme. Denne antakelsen blir helt urealistisk når vi f.eks. ser på aldrene 91-99. Hvis vi ser på råratene som estimat for suksessansynligheten i en binomisk situasjon, kan vi av data fra 1966-1970 for menn finne estimater for variansene til råratene. Disse er satt opp i tabell 6.1.



Tabell 6.1. Estimerer for variansen til råratene for menn 1966-1970

Alder	Estimert varians	Alder	Estimert varians	Alder	Estimert varians
80 år	$3 \times 10^{-6}$	90 år	$35 \times 10^{-6}$	100 år	$29 \times 10^{-4}$
81 "	$4 \times 10^{-6}$	91 "	$50 \times 10^{-6}$	101 "	$56 \times 10^{-4}$
82 "	$5 \times 10^{-6}$	92 "	$71 \times 10^{-6}$	102 "	$112 \times 10^{-4}$
83 "	$6 \times 10^{-6}$	93 "	$113 \times 10^{-6}$	103 "	$182 \times 10^{-4}$
84 "	$7 \times 10^{-6}$	94 "	$160 \times 10^{-6}$	104 "	$292 \times 10^{-4}$
85 "	$9 \times 10^{-6}$	95 "	$246 \times 10^{-6}$	105 "	$400 \times 10^{-4}$
86 "	$12 \times 10^{-6}$	96 "	$399 \times 10^{-6}$		
87 "	$15 \times 10^{-6}$	97 "	$596 \times 10^{-6}$		
88 "	$20 \times 10^{-6}$	98 "	$985 \times 10^{-6}$		
89 "	$26 \times 10^{-6}$	99	$1645 \times 10^{-6}$		

Når forholdet mellom variansene til råestimatene er som i tabell 6.1, vil de glattingsmetoder vi har sett på til nå, være mindre egnet. For aldrene 88, 93 og 98 år konstruerte vi derfor, under forutsetning av at forholdet mellom variansene til råratene er som i tabell 6,1, for ulike lengder og plasseringer de glidende gjennomsnitt som har størst varians-reducerende evne blant dem som bevarer annengradspolynomer. Vi brukte noen av disse på de seks sett råratene vi hadde, og de gav tilfredsstillende resultater, bortsett fra noen ujamne estimerer for de eldste aldrer på data for kvinner fra årene 1971-1972. Disse uregelmessighetene er lette å forstå når en ser på råratene vi prøvde å glatte.

Tabell 6.2. Råratene for kvinner 1971-1972. Promille

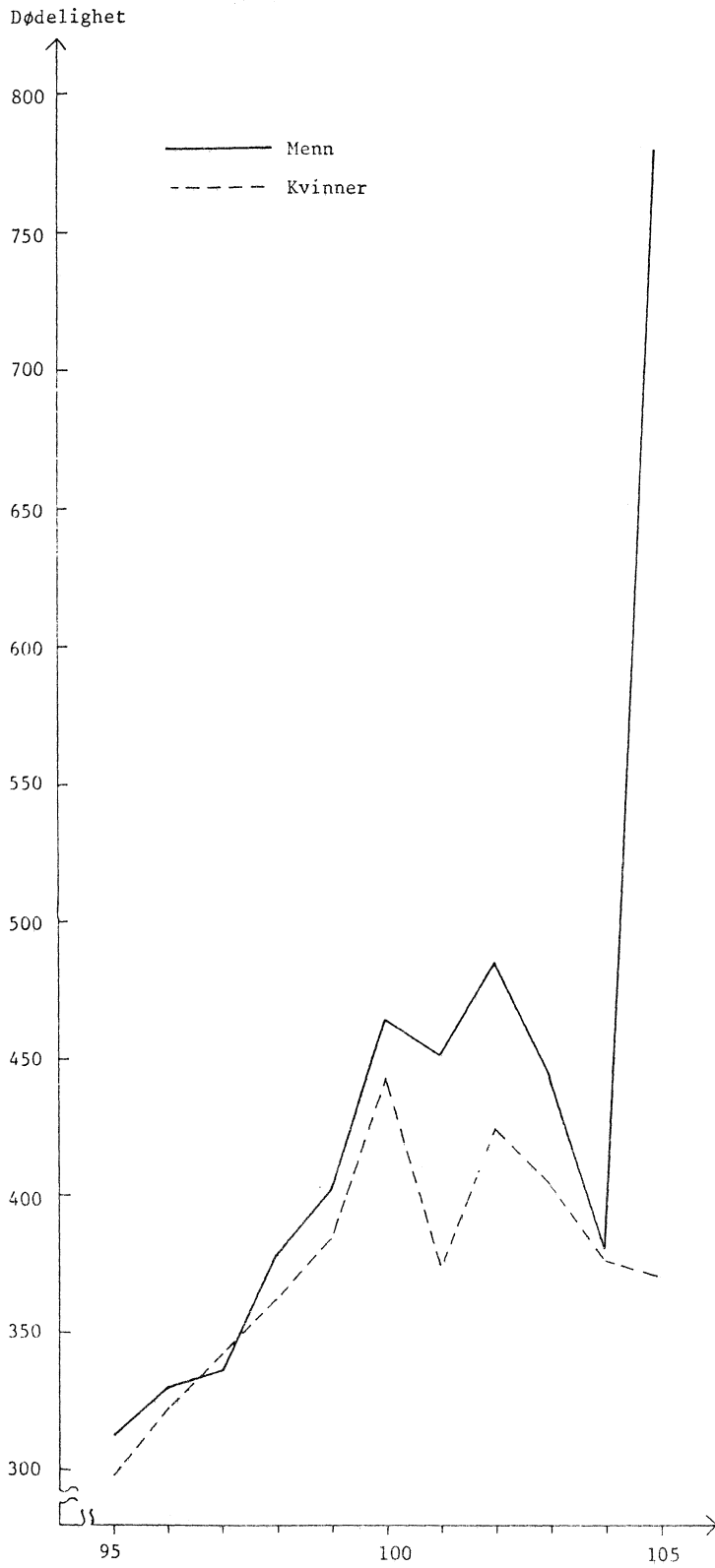
Alder	Dødssannsynlighet	Alder	Dødssannsynlighet
95 år	273	100 år	375
96 "	291	101 "	304
97 "	363	102 "	280
98 "	348	103 "	308
99 "	345	104 "	143
		105 "	0

Når datagrunnlaget blir så spinkelt som det er for aldrene 101, ..., 105, fant vi ingen rimelig måte å glatte råratene på ved hjelp av glidende gjennomsnitt. Med data for 5-årsperioder har en for aldrene 96-105 som regel foretatt en skjønnsmessig glatting, dvs. en har med utgangspunkt i råratene trukket opp en jamnt stigende kurve og latt denne gå mot 1,0 når alderen blir 106 år.

Siden vi gjerne ville ha en metode der all glatting kunne foregå maskinelt, måtte vi prøve å finne en annen fornuftig metode.

Vi så på data fra en lengre periode for å finne om noe kunne sies om nivå og form på dødelighetskurven for de aller eldste aldersklassene. Råestimer ble beregnet på grunnlag av data fra 1961-1972. Disse gjengis i figur 6.3.

Figur 6.3. Rårater for de eldste aldersklassene med data fra 1961-1972. Promille



For dem over 100 år er fremdeles datagrunnlaget spinkelt, men det synes klart at dødeligheten ikke fortsetter å stige med alderen i den takt den gjør mellom 95 og 100 år. En kan gjette på at en har en markert utflating i den virkelige dødelighetskurven for personer 100 år og eldre.

Vi valgte derfor å foreta en "preglating". For aldrene 100, ..., 105 setter vi rårate (og glattet rate) lik 0,450 for menn og 0,425 for kvinner. Så anvendes for aldrene 89, ..., 99 glattingsformelen  $(-0.20410, 0.02062, 0.14495, 0.20022, 0.20259, 0.18498, 0.14177, 0.11743, 0.08716, 0.06009, 0.04428)$ . Blant dem med samme plassering som bevarer annengradspolynomer har dette glidende gjennomsnitt stor variansreducerende evne når variansene til råratene er som i tabell 6.1.

Vi ser at det blir en viss vilkårlighet i bestemmelsen av dødeligheten for de over 99 år. For å sikre oss mot at dette ikke skal ha innflytelse på ofte brukte indikatorer, som forventet gjenstående levetid for 0, 1, 20 og 50 åringer, foretok vi noen alternative beregninger av disse. Vi brukte data om døde fra 1966-1970. Metodene var:

- skjønnsmessig glatting av ratene for 95-106 åringer, der en lar kurven stige jamt til den når 1,0 for 106-åringer.
- som i a) for 95-99 åringer og 106-åringer, men med dødsrate 0,450 for menn og 0,425 for kvinner i aldrene 100-105 år.
- som i a) for 95-99 åringer, men med dødsrate lik 1,0 for 100-åringer.

Som en ser av tabell 6.4, er de ulike beregningsmåtene uten innflytelse på de sentrale indikatorene.

Tabell 6.4. Gjennomsnittlig antall gjenstående leveår ved alderen  $n$ , beregnet ved tre ulike metoder med data fra 1966-1970

n	Glattingsmetode					
	a		b		c	
	Menn	Kvinner	Menn	Kvinner	Menn	Kvinner
0 år .....	71,09	76,83	71,09	76,83	71,09	76,83
20 " .....	53,08	58,28	53,08	58,28	53,08	58,28
50 " .....	25,32	29,47	25,32	29,47	25,32	29,46
70 " .....	10,87	12,83	10,87	12,83	10,87	12,82
90 " .....	3,16	3,59	3,16	3,60	3,14	3,55
95 " .....	2,21	2,58	2,24	2,60	2,14	2,42
98 " .....	1,67	2,09	1,78	2,14	1,41	1,60
100 " .....	1,30	1,64	1,66	1,77	0,5	0,5

## REFERANSE

- [1] Sverdrup, Erling (1973): "Lov og tilfeldighet" , bind I, 2. utgave.  
Universitetsforlaget, Oslo.