



Matematisk statistik
Stockholms universitet

En empirisk jämförande studie mellan
långtidssjuk- och
premiebefrielseförsäkringar avseende
en icke-parametrisk och en
parametrisk modell.

David Persson

Examensarbete 2002:6

Postadress:

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm
Sverige

Internet:

<http://www.math.su.se/matstat>



Matematisk statistik
Stockholms universitet
Examensarbete 2002:6,
<http://www.math.su.se/matstat>

En empirisk jämförande studie mellan långtidssjuk- och premiebefrielseförsäkringar avseende en icke-parametrisk och en parametrisk modell.

David Persson*

december 2001

Sammanfattning

This essay studies the portfolio of individual long-term health insurances at LFAB. The t -frequencies and the termination function are presented and estimated. These two probabilities are studied to see if there are any differences between sexes, between sick insurances and insurances to cover payment of premiums in case of incapacity or different waiting periods. The study shows that the long-term sickness has increased in later years and that the t -frequencies for women are considerably higher than the t -frequencies for men. The essay also looks at different parametric models to see if there exists a model that can explain termination from illness by recovery or by death. It turns out that the model best suited for the data used in this study is a model based on a modified Gompertz distribution.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91 Stockholm.

En empirisk jämförande studie mellan
långtidssjuk- och premiefrielseförsäkringar
avseende
en icke-parametrisk och en parametrisk modell

David Persson*

December 2001

Abstract

This essay studies the portfolio of individual long-term health insurances at LFAB. The t-frequencies and the termination function are presented and estimated. These two probabilities are studied to see if there are any differences between sexes, between sick insurances and insurances to cover payment of premiums in case of incapacity or different waiting periods.

The study shows that the long-term sickness has increased in later years and that the t-frequencies for women are considerably higher than the t-frequencies for men. The essay also looks at different parametric models to see if there exists a model that can explain termination from illness by recovery or by death. It turns out that the model best suited for the data used in this study is a model based on a modified Gompertz distribution.

* I would like to thank the people at the division for actuarial technique on LFAB who made this essay possible by allowing full access to their facilities.

1 FÖRORD	2
1.1 KORT OM LÄNSFÖRSÄKRINGSGRUPPEN OCH LÄNSFÖRSÄKRINGAR LIV	2
1.2 ALLMÄNT	2
2 INLEDNING	3
2.1 BAKGRUND	3
2.2 PROBLEMFÖRMULERING	3
2.3 SYFTE	4
2.4 DATA	4
3 TEORI	6
3.1 SJUKFÖRSÄKRINGSTEORI	6
3.2 FÖRSÄKRINGSTYPER	8
3.2.1 <i>Långtidssjukförsäkring</i>	8
3.2.2 <i>Premiefrielseförsäkring</i>	9
3.2.3 <i>Kapitalförsäkring, K-försäkring</i>	9
3.2.4 <i>Pensionsförsäkring P-försäkring</i>	9
4 T-FREKVENS	10
4.1 SKATTNING AV T-FREKVENS	10
4.2 BERÄKNING AV T-FREKVENSER	11
5 AVVECKLINGSPROCESSER	14
5.1 FILTRATION OCH MARTINGALTEORI	14
5.2 RÄKNEDPROCESSER	15
5.3 CENSURERING OCH TRUNKERING	17
6 AVVECKLINGSFUNKTIONEN	19
6.1 ICKE-PARAMETRISKA MODELLER	19
6.2 SKATTNING AV AVVECKLINGSFUNKTIONEN	21
6.2.1 <i>Konfidensintervall</i>	24
6.2.2 <i>Konfidensband</i>	26
7 ICKE-PARAMETRISKT TEST	28
7.1 LOG-RANK TEST	29
7.2 GEHAN TEST	31
8 PARAMETRISKA MODELLER FÖR AVVECKLINGSFUNKTIONEN	33
8.1 MODELLER	33
8.2 LIKELIHOOD	36
8.3 VAL AV MODELL	36
9 RESERVER	40
9.1 UTJÄMNING	40
10 SLUTSATSER	43

Appendix

- A. observerade T- frekvenser.
- B. observerade avvecklingsfunktioner.
- C. Diagram avseende Nelson-Aalen skattningen med tillhörande konfidensband/intervall.
- D. Avvecklingsfunktionen enligt Grunderna 1984.
- E. Förhållandet mellan parametrisk, ickeparametrisk skattad avvecklingsfunktion och den minstakvadrat anpassade Nelson-Aalen skattningen.
- F. Beräknade reserver

Referenser

1 Förord

1.1 Kort om Länsförsäkringsgruppen och Länsförsäkringar Liv¹

Länsförsäkringsgruppen består av 24 självständiga, kundägda och lokala länsförsäkringsbolag som äger gemensamma verksamheter.

Länsförsäkringar AB, LFAB, ägs till 96,5 procent av de 24 lokala länsförsäkringsbolagen och till 3,5 procent av EurAPCo. Wasa stiftelsen äger en aktie. Länsförsäkringar Wasa är ett moderbolag i en koncern som bedriver verksamhet inom liv- och sakförsäkring, bank, fond och kapitalförvaltning. Bolaget ger länsförsäkringsbolagen service inom dessa områden och svarar för gruppens gemensamma utvecklingsverksamhet.

Länsförsäkringar Liv är ett helägt dotterbolag till Länsförsäkringar AB. Bolaget bedriver traditionell livförsäkringsverksamhet enligt ömsesidighetsprincipen, vilket innebär att all vinst överförs till kunderna. Genom dotterbolaget Länsförsäkringar fondliv ansvarar bolaget för fondförsäkringar inom gruppen.

Sedan den 1 januari 1999 bedrivs Wasa Livs och Länsförsäkringar Livs traditionella livförsäkringsverksamhet i Länsförsäkringar Wasa Liv Försäkringsaktiebolag. Vid halvårsskiftet fusionerades de två fondförsäkringsbolagen Wasa Fondförsäkring och Länsförsäkringar Fondliv till Länsförsäkringar fondliv. Avslutningsvis förvärvade Länsförsäkringar Liv per 31 december, 1999 Länsförsäkringar Fondliv från Länsförsäkringar Wasa.

1.2 Allmänt

De resultat som framkommit i denna rapport är något modifierade för att skydda LFAB's intressen.

Beräkningsprogrammet SAS har använts vid utsökningar från F-systemet på LFAB Liv, MathCad vid beräkningar av integraler samt Microsoft Excel Problemlösaren² för att lösa de optimeringsproblem som framkommit i rapporten.

¹ Informationen är hämtad ur "Länsförsäkringsgruppen verksamhet 2001".

² Algoritmer som används av Problemlösaren:

Microsoft Excel Problemlösaren använder den icke-linjära optimeringskoden GRG³ (Generalized Reduced Gradient) som är utvecklad av Leon Lasdon, University of Texas, Austin och Allan Waren, Cleveland State University. I linjära problem och heltalsproblem används simplexmetoden, utvecklad av John Watson och Dan Fylstra, Frontline Systems, Inc. Mer information erhålls i FTN's dödlighetsundersökning 1999.

2 INLEDNING

Denna rapport är en empirisk jämförelsestudie mellan fristående sjukförsäkringar och premiefrielseförsäkringar. Den undersökta populationen indelas i ett antal delmängder för att belysa skillnader mellan försäkringstagare, d v s beroende på vem personen är, samt hur valet av försäkring som vederbörande gör, påverkar de olika sannolikheterna som är av intresse för ett försäkringsbolag.

2.1 Bakgrund

Det finns en lång tradition i Sverige inom både långtidssjuk- och premiefrielseförsäkring och så tidigt som 1891 införde livbolaget Thule de första premiefrielseförmånerna vilket snart följdes upp av flera andra livbolag. Förmånen medgavs dem som i bolaget innehade livförsäkring eller kapitalförsäkring med kapitalutbetalning fram till och med 55 års ålder. Premiefrielseförsäkringen gav med tiden upphov till viktiga erfarenhetsrön ifråga om sjukförsäkringsvillkor och år 1911 grundades det första sjukförsäkringsbolaget Eir och ett år senare bolaget Valkyrian. Det förstnämnda bolaget Eir var underställt de livbolag som fanns medan Valkyrian var bundet till de bolag som förfogade över olycksfallsförsäkringar. I början ledde detta till att sjukförsäkringar hade låg prioritet, vilket höll tillbaka utvecklingen för dessa försäkringar. Först 40 år senare startade de berörda bolagen ett samarbete vilket, 1954, ledde till att bolagen började använda gemensamma beprövade tekniska baser för beräkning av premier och sjukreserver, se Mattsson [1956]. Det naturliga blev att även samma grunder och villkor kom till användning för premiefrielseförsäkringen. Dessa tekniska baser utvecklades ytterligare 1965, 1973 se Dillner [1973] och 1984 se Källström [1990]. Att dessa grunder ändras med tiden beror dels på att sjukbilden bland de försäkrade förändras samt på externa händelser så som förändringar i pensionssystemet e t c. I dag används i stor utsträckning grundfria försäkringar som innebär att bolagen var för sig bestämmer sina premier och reserver utifrån egna erfarenheter.

2.2 Problemformulering

Modeller som ska beskriva långtidssjukligheten kan bli väldigt komplexa om hänsyn skall tas till varje enskild individs beteendemönster samt hur just denna individ reagerar för detta beteendemönster. Faktorer som kan påverka sjukligheten är exempelvis typ av arbete, kön, ålder, rökning, antalet barn i hemmet e t c. Sjuklighet beror även på externa effekter i form av konjunkturcykler, ändrade sociala skyddsnät, oroligheter i omvärlden och så vidare, vilket i stor utsträckning bidrar till att sjukligheten förändras över tiden. I och med att sjukligheten ändras från en tid till en annan, blir det svårt att prediktera och detta ger anledning till att kontinuerligt upprepa nya sjuklighetsundersökningar för att på så sätt fånga upp den sjuklighet som råder idag. Dock antas förhållandet mellan olika delbestånd ha en viss varaktighet över tiden.

Svårigheter att estimerar flertalet variabler som kan vara av intresse och bristen på information, leder till att tidigare undersökningar, och även denna, gjort uppdelning avseende på kön, ålder, karens och på försäkringstyp.

Den allmänna erfarenheten säger att premiefrielseförsäkringar har en lägre utnyttjandegrad än långtidssjukförsäkringar. Detta leder till att kostnaden för premiefrielseförsäkring på ett visst belopp är lägre än för en sjukförsäkring med

motsvarande belopp och karenstid, trots att de två försäkringarna tekniskt sett är lika. En fråga är då om försäkringstagare på senare tid i större utsträckning utnyttjar en premiefrielseförsäkring, vilket i sin tur leder till att kostnaderna för bolaget rörande denna försäkring ökar. Denna ökning i utnyttjandegraden för premiefrielseförsäkringar borde i så fall visa sig i att sannolikheten att insjukna liknar motsvarande sannolikhet för långtidssjukförsäkringar.

En fråga är också om skillnad förekommer i sannolikhet att avgå från tillståndet sjuk/skadad för premiefrielse- respektive sjukförsäkringar. I dag antas denna övergångssannolikhet vara densamma för de två försäkringarna, eftersom skillnad härrör från utnyttjandegraden, vilket enbart bör synas i frekvensen att insjukna och inte i sannolikheten att tillfriskna. Det är dock möjligt att en viss skillnad finns i denna sannolikhet då urvalet för de två försäkringarna kan vara olika. Detta kan komma av att i fall där försäkringen avslutats, har det framkommit att den försäkrade varit berättigad till premiefrielse, vilket lett till att den försäkrade fått en retroaktiv ersättning. Detta kan innebära att försäkrade personer med en relativ kort sjuktid med större sannolikhet än personer med lång sjuktid inte upptäcker att de är berättigade till en premiefrielseförsäkring. Om detta gäller, bör ett bestånd med premiefrielseförsäkringar i förhållande till sjukförsäkringar innehålla en större del långtids sjuka vilket genererar skillnader i de skattade sannolikheterna att tillfriskna från sjukdom eller skada.

2.3 Syfte

Syftet med denna undersökning är att empiriskt studera förhållandet mellan långtidssjuk- och premiefrielseförsäkringar, mellan kvinnor och män och mellan R- och tre månadskarens, för att se om olika indelningsgrupper bör behandlas på olika sätt. Det andra syftet är att se hur sjukligheten har förändrats i förhållande till tidigare sjuklighetsundersökningar för att se om de antaganden som bolaget baserar sina beräkningar på idag bör förändras. Avslutningsvis undersöks hur väl en parametrisk fördelning passar till data i förhållande till den vedertagna icke-parametriska modellen som i stor utsträckning används idag för att skatta sannolikheten att tillfriskna från sjukdom eller skada.

2.4 Data

Informationen är administrerad i F-systemet på Länsförsäkringar Liv AB och består av ögonblicksbilder på sjukfall från årskiftena 1996-12-31, I(97), fram till årskiftet 2000-12-31, I(01), samt data från aktuellt datum 2001-08-31, I(Ab), vilket innebär att information från sex skilda tidpunkter används. Materialet omfattar således sjukfall som dels inträffat fr o m 1 januari 1997, dels sådana sjukfall som inträffat före 1 januari 1997 och som pågick 1 januari 1997. För pågående sjukfall har längden på sjuktiden beräknats per 31 augusti 2001. Låt I beteckna det studerade intervallet.

Låt I(97), I(01), I(Ab) o s v beteckna informationen vid respektive tidpunkt. Vid beräkning av avvecklingsfunktionen för sjuk respektive premiefrielseförsäkringar har de sex bestånden länkats samman på så sätt att beståndet vid aktuellt datum summerats med individer som inte finns i det aktuella beståndet, enligt:

$$Avv = (Ab) \cup (I(97) \cap I(98)) \cup (I(98) \cap I(99)) \cup \dots (I(01) \cap I(Ab)).$$

För långtidssjukförsäkringar där data indikerar att avgång från sjukdom berott på återköp³, saknar värde eller obetald, antas i denna studie härröra från fel i instansningsförfarandet, således kommer denna typ av avveckling att behandlas som tillfrisknad från sjukdom. De avvecklingar som inträffar strax före, eller strax efter, att försäkringen avslutas, behandlas på samma sätt som ovan. Vidare gäller för sjukförsäkringen att de två indelningarna avseende karenstyp, R- och 3- månadskarens, är nästintill identiska eftersom en och samma försäkring vanligtvis innehåller flera olika karens.

Data baserade på premiefrielseförsäkringar behandlas på samma sätt som ovan. Notera dock att variabel tillfrisknad för dessa försäkringar förklarar en förväntad eller faktisk avvecklingstidpunkt. Tillfriskning som således sker i framtiden antas vara högercensurerad. Dessutom gäller för denna försäkring att indelningen avseende karenstyp är disjunkt.

Mängden 3- månadskarens kommer även att inkludera 1- månadskarens och uppdelning avseende ålder och karens blir givetvis per definition disjunkta.

Studien baseras på ett relativt kort tidsintervall I , detta för att försöka modellera den sjuklighet som gäller i dag, dock antas urvalet vara av den storleksordningen att skattningarna kan anses vara stabila.

³ Dessa typer av avvecklingar från tillståndet sjuk/skadad är tekniskt sätt inte möjlig, eftersom en person ej kan köpa tillbaks en "riskförsäkring". Likväl kan inte en person avgå från tillståndet sjuk p g a att försäkringen varit obetald eller saknar värd. Detta föranleder antagandet om att personen först har tillfrisknat varefter vederbörande har avslutat sin försäkring.

3 Teori

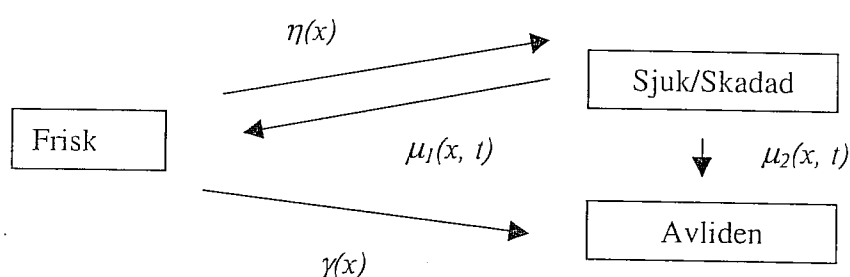
I detta kapitel kommer en övergripande beskrivning av de teorier som ligger till grund för denna sjuklighetsundersökning. Den enkla modell och de funktioner och övergångsannolikheter som är av intresse i denna rapport beskrivs i Avsnitt 3.1. För mer utförlig beskrivning av dessa teorier hänvisas läsaren till Klein och Moeschberger [1997] samt Ajne och Ohlin [1994]. Vidare ges en kort beskrivning i Avsnitt 3.2 av de försäkringstyper som förekommer i undersökningen, Lewander [2001].

3.1 Sjukförsäkringsteori

De två storheter som är av intresse vid en sjuklighetsundersökning, baserad på den s.k. Svenska modellen, är intensiteten att insjukna och överlevnadsfunktionen. För att anknyta till försäkringstekniska termer inom sjuklighetsundersökningar, kommer överlevnadsfunktionen att benämnas avvecklingsfunktionen.

För att illustrera relationen mellan dessa funktioner antas en enkel modell över sjukligheten. Modellen består av tre tillstånd: frisk, sjuk/skadad och tillståndet avliden se *Figur 1*.

Figur 1.



Figur 1 innehåller således tillstånden frisk, denna inkluderar även de som varit sjuka eller skadade, tillståndet sjuk/skadad och avslutningsvis tillståndet avliden. De övergångsintensiteter som återfinns i *Figur 1* ovan definieras som:

- $\eta(x)$ = Intensiteten att bli sjuk, givet att stå i tillstånd frisk.
- $\gamma(x)$ = Intensiteten att avlida givet att stå i tillstånd frisk.
- $\mu_1(x, t)$ = Intensiteten att tillfriskna från tillståndet sjuk/skadad
- $\mu_2(x, t)$ = Intensiteten att avlida från tillståndet sjuk/skadad

$X = x$ betecknar åldern för övergång från tillstånd frisk till tillstånd sjuk/skadad och $T = t$ står för tiden en person tillbringar i tillstånd sjuk/skadad, innan övergång till tillstånd frisk eller avliden sker. Den stokastiska variabeln T antas vara oberoende, icke-negativ och kontinuerlig och kommer att benämnas durationstiden.

Sannolikheten att en person vid x års ålder insjuknar eller får en skada i det lilla tidsintervallet dx är $\eta(x) \cdot dx$. Den andra funktionen, avvecklingsfunktionen, definieras som

sannolikheten att en person som insjuknar vid åldern x år fortfarande är sjuk t år senare, denna sannolikhet kan skrivas som:

$$\lambda(x, t) = P(T \geq t | X = x). \quad (3.1.1)$$

En avvecklingsfunktion är en icke växande funktion som antar värdet ett då $t = 0$ samt antar värdet noll för $t = \infty$. Enligt definition är (3.1.1) komplementet för fördelningsfunktionen, $F_T(t | x) = P(T < t | X = x)$, vilket innebär att $\lambda(x, t) = 1 - F_T(t | x)$. Förhållandet mellan täthets- och avvecklingsfunktionen ges då av:

$$\lambda(x, t) = \int_t^{\infty} f(s | x) ds, \quad (3.1.2)$$

detta kan uttryckas enligt:

$$\frac{d\lambda(x, t)}{dt} = -f(t | x). \quad (3.1.3)$$

Vid sjuklighetsundersökningar summeras intensiteten att tillfriskna $\mu_1(x, t)$, med intensiteten att avlida $\mu_2(x, t)$. Denna summering är tillåten då intensiteterna är oberoende samt att övergång från tillstånd sjuk/skadad till något av de två tillstånden, leder i båda fallen till att pågående ersättning från bolaget gällande en sjuk eller premiefrielseförsäkring avslutas innan en försäkringstid tar slut. Detta skrivs som:

$$\mu(x, t) = \mu_1(x, t) + \mu_2(x, t). \quad (3.1.4)$$

Vidare kan intensitetsfunktionen (3.1.4) definieras som gränsvärdet att tillfriskna i det lilla tidsintervallet dt , givet att en person som insjuknade vid x års ålder fortfarande är sjuk t år senare, enligt:

$$\mu(x, t) = \lim_{dt \rightarrow 0} \frac{P(T \leq t + dt | T \geq t, X = x)}{dt}, \quad (3.1.5)$$

här gäller att $\mu(x, t)$ är ickenegativ d v s, $\mu(x, t) \geq 0$. Som tidigare antagits är den stokastiska variabeln T kontinuerlig, vilket gör att uttrycket (3.1.5) kan skrivas som:

$$\mu(x, t) = \frac{f(t | x)}{\lambda(x, t)} = -\frac{d \log(\lambda(x, t))}{dt}. \quad (3.1.6)$$

Relationen mellan intensitetsfunktionen, $\mu(x, t)$, och avvecklingsfunktionen, $\lambda(x, t)$, kan enligt (3.1.6) skrivas om till:

$$\lambda(x, t) = \exp\left(-\int_0^t \mu(x, s) ds\right). \quad (3.1.7)$$

Genom att kombinera sannolikheten att insjukna och fortfarande vara sjuk erhålls den kombinerade sannolikheten

$$v(x, t) \cdot dt = \eta(x) \cdot \lambda(x, t) \cdot dt \quad (3.1.8)$$

Storheten $v(x, t)$ i (3.1.8) kallas för t-frekvensen och kommer att behandlas i Kapitel 4. De två funktioner som studeras i en sjuklighetsundersökning är således t-frekvensen (3.1.8), och avvecklingsfunktionen (3.1.1). Enligt den Svenska modellen används två funktioner vid beräkning av premier för långtidssjuk- respektive premiefrielseförsäkringar och för beräkning av sjukreserven används avvecklingsfunktionen allena.

En enkel sjukpremie beräknad enligt den Svenska modellen, diskuteras utförligt i Dillner [1969], [1973], definieras för en årlig krona enligt:

$$E(k, x, z - x) = \int_0^{z-x-k} e^{-\delta \cdot s} \cdot \frac{l_{x+s}}{l_x} \cdot v(x + s, k) \cdot \int_k^{z-x-s} \frac{\lambda(x + s, u)}{\lambda(x + s, k)} \cdot e^{-\delta \cdot u} du ds$$

Här står k för karens tiden, z betecknar slut åldern för en den försäkrade, kvoten l_{x+s}/l_x för sannolikheten att en x åring lever vid $x + s$ års ålder och δ betecknar intensiteten från räntan, $\delta = \ln(1-i)$ där i betecknar räntesatsen. Beteckningen $v(x + s, k)$ innebär att sannolikheten att insjukna och fortfarande vara sjuk påverkas av valet av karens tid k , dvs försäkringstagarens benägenhet att insjukna antas till en viss mån fångas upp i individens val av karens. Vidare är det avvecklingsfunktionen som styr storleken på sjukreserven, även denna beräkning diskuteras i Dillner [1969], [1973] och beräknas enligt den Svenska modellen som:

$$a(x, t, z - x - t) = \int_t^{z-x} \frac{\lambda(x, u)}{\lambda(x, t)} \cdot e^{-\delta(u-t)} du, \quad t \geq k,$$

här är δ och k som ovan. Sjukreserven behandlas mer utförligt i Kapitel 9.

3.2 Försäkringstyper

De två försäkringar som ingår i arbetet är som tidigare nämnts sjuk- och premiefrielseförsäkringar, dock är premiefrielseförsäkringen en tilläggsförsäkring som ingår i exempelvis en kapital eller en pensionsförsäkring. För att ge en vidare bild av vad en sjukundersökning innebär, kommer nedan en kort beskrivning av respektive försäkring.

3.2.1 Långtidssjukförsäkring

En långtidssjukförsäkring ger den försäkrade rätten till en månadsvis ersättning så länge som den nedsatta arbetskapaciteten, orsakad av sjukdom eller av en olycka, är total eller åtminstone 50 procent. Ersättningen påbörjas då karenstiden tar slut och fortsätter fram till

dess försäkring avslutas. I de fall där den nedsatta arbetskapaciteten inte är total, utbetalas en ersättning som motsvarar graden på skadan eller sjukdomen/arbetsoförmågan. En praxis för försäkringsbolagen är att arbetsoförmågan antas vara antingen 50, 75 eller 100 procent.

3.2.2 Premiefrielseförsäkring

Tekniskt sett utgör premiefrielseförsäkringen, Pb- försäkringen, en långtidssjukförsäkring. Detta innebär att vid fortsatt arbetsoförmåga efter karenstidens slut finansieras en pensions- eller en kapitalförsäkring, kallade P- respektive K-försäkringar, så länge arbetsoförmågan kvarstår. Skillnaden mellan en långtidssjuk- och en Pb-försäkring är således att en sjukförsäkring ger den försäkrade rätten till en inkomst i "handen" medan en Pb-försäkring ger rätten till en finansiering av en utgift.

För P- försäkringar gäller karenstiden tre månader med sammanläggning av perioder med arbetsoförmåga om minst en månad inom en 12-månadsperiod. Vidare gäller rörlig karens, R, för K-försäkringar, d v s när den försäkrade på grund av sjukdom eller olycksfallskada blir berättigad till sjukbidrag eller förtidspensionering träder Pb-försäkringen i kraft.

3.2.3 Kapitalförsäkring, K-försäkring

Med K-försäkring menas i inkomstskattelagen en försäkring där utfallande belopp är fritt från inkomstskatt och att en premie som erläggs för försäkringen, ej är avdragsgill. En kapitalförsäkring kan utbetalas med ett engångsbelopp eller med periodisk utbetalning. I vissa fall kan utfallande belopp vara befriade från arvs- och gåvoskatt.

3.2.4 Pensionsförsäkring P-försäkring

P-försäkringar i inkomstskattelagens mening kännetecknas av att utfallande belopp inkomstbeskattas samt att utbetalningen sker periodiskt där det utbetalade beloppet inte får sjunka utan måste vara konstant eller stiga. Vidare gäller att ålderspensionen får tidigast utbetalas från 55 års ålder och att återköp eller belåning av pensionsförsäkringen inte är tillåten.

4 T-frekvens

Intensiteten att en person blir sjuk vid x års ålder och kvarstår som sjuk t år senare betecknas $\nu(x, t)$. I termer av sjuklighetsintensiteten $\eta(x)$ och avvecklingsfunktionen $\lambda(x, t)$ kan t-frekvensen, enligt Formel (3.1.8), skrivas som $\nu(x, t) = \eta(x) \cdot \lambda(x, t)$, se Arne [1990] och Dillner[1973].

4.1 Skattning av t-frekvens

De två funktioner som används i den Svenska modellen för att skatta t-frekvensen är, som tidigare nämnts, insjuknandeintensiteten⁴ för en x -åring $\eta(x)$ och avvecklingsfunktionen $\lambda(x, t)$, som är sannolikheten att en individ som insjuknar vid x års ålder, kvarstår som sjuk t år senare. T-frekvensen blir då de kombinerade sannolikheterna att insjukna i x års ålder och att kvarstå som sjuk t år senare, vilket kan skrivas som:

$$\nu(x, t) = \eta(x) \cdot \lambda(x, t). \quad (4.1.1)$$

Funktionen $\nu(x, t)$ beräknas för värden på t som är möjliga att observera, dvs för $t \geq k$, där k betecknar karenstiden. Skattningen av t-frekvensen (4.1.1) fås genom att beräkna kvoten mellan antalet sjukfall vid åldern x med åtminstone en duration på t år och det totala försäkringsbeståndet, både aktiva och sjuka, vid samma ålder. För höga värden på t studeras enbart avvecklingsfunktionen istället för ovanstående frekvens.

För att skatta t-frekvensen definieras följande storheter:

$M(x, t, I) =$ Antalet sjukfall som inträffat bland de försäkrade i åldern x år, under observationsperiod I som studeras, med en sjukduration av åtminstone t år.

$N(x, I) =$ Antal försäkringsår för personer i åldern x år i beståndet under observationsperioden I .

Låt nu $n_x(1997)$ beteckna antalet försäkrade personer i åldern x år den 1 januari år 1997. Om vi bortser från eventuella säsongsvariationer i insjuknandefrekvensen så är medelvärdet av $n_x(1997)$ och $n_x(1998)$ en rimlig skattning av den sammanlagda

⁴ Kvantiteten $\nu(x, t)$ ska inte ses som en korrekt skattning, enligt teorier om stokastiska processer, av övergångssannolikheten att gå från tillstånd frisk till tillstånd sjuk. Skattningen är systematiskt för låg, eftersom måttet refererar till antalet sjukfall i relation till det totala antalet friska plus de som redan är sjuka. Intensiteten skall därför inte ses som ett precist mått på insjuknandefrekvensen utan, tillsammans med avvecklingsfunktionen, ge en korrekt skattning av det förväntade antalet, vid x års ålder, som insjuknar med en duration, $t \geq k$. Det finns en viss svaghet i att skatta t-frekvensen på detta sätt eftersom olika bestånd kan inkludera olika stora andelar personer som är sjuka. Exempelvis kan två portföljer av försäkrade jämföras där den ena enbart består av nytecknade försäkringar, vilket innebär att alla i portföljen initialt är friska i kontrast mot en avtagande portfölj som, under vissa extrema betingelser, till stor del består av sjuka personer. I den sistnämnda portföljen skulle nya insjuknandefall vara begränsade i förhållande till den förstnämnda portföljen vilket i sin tur kan ge stora skillnader i den skattade t-frekvensen.

exponeringstiden för långtidssjuk- och premiefrielseförsäkrade individer under år 1997. Genom att summera de skattade risktiderna från intilliggande år erhålls den totala risktiden $N(I)$, enligt:

$$N(1997,2000) = \frac{1}{2}n(1997) + n(1998) + \dots + \frac{1}{2}n(2001)$$

Då försäkringsbestånden är relativt små kommer beräkningar att utföras i åldersklasser istället för enskilda år.

4.2 Beräkning av t-frekvenser

För beräkning av t-frekvenser har försäkringsbeståndet delats upp i delmängder. Dessa mängder är uppdelade efter kön, åldersklass, karens och på sjuk- respektive premiefrielseförsäkring. Åldrarna är grupperade i klasserna: 25-29, 30-34, ..., 60-64 och benämns 27, 32, ...62. Karensen är uppdelad i två grupper, rörlig karens respektive 3 månadskarens. Premiefrielseförsäkringen delas upp i K och P skattade försäkringar där den sistnämnde antas ha tre månadskarens och den K skattade R-karens.

Samtliga mängder är beräknade med avseende på 3 durationsmått, ≥ 3 månader, ≥ 6 månader och ≥ 1 år. För de två första durationsmåttarna är observationsperioden från den 1 januari 1997 till 31 december 2000 vilket innebär en observationsperiod på 4 år.

Durations måttet för $t \geq 1$ år kan enbart observeras under 3 år, eftersom en person som insjuknar i slutet av 2000 inte kan observeras under ett helt år, vilket leder till att observationsperioden i detta fall är mellan den 1 januari 1997 och 31 december, 1999.

I *Tabell 4.2.1* och *4.2.2* nedan jämförs de i undersökningen framräknade t-frekvenserna för långtidssjukförsäkringar med FTN's sjukundersökning 1993 samt med 1984 års grunder. Dessa undersökningar har en karensindelning avseende exakt en månad, exakt 3 månader och längre än 3 månader vilket innebär att gruppen R-karens jämförs med längre än 3 månader och tre månadskarens jämförs med 3 månaders. Pb-försäkringen jämförs med FTN's premiefrielseundersökning 1974. Efter denna undersökning har inga fler undersökningar för Pb-försäkringar blivit gjorda. För resultat se *Tabell 4.2.3*. Vidare visas i *Tabell 4.2.4* och *4.2.5* förhållandet i utnyttjandegraden mellan sjuk- respektive Pb-försäkringar och relationen mellan t-frekvenser för män och kvinnor och avslutningsvis förhållandet mellan de två karensmåttarna ovan.

I Bilaga A jämförs de observerade t-frekvenserna för långtidförsäkringar med FTN sjuklighetsundersökning 1993 för durationstiderna 6 månader och 1 år. Motsvarande resultat för Pb-försäkringar har tidigare inte blivit genomförda.

Tabell 4.2.1 och Bilaga A tyder på att betydligt fler personer insjuknar under det studerade intervallet än vad som framgår av både G84 och FTN, vilket tyder på att försäkrade personer oftare insjuknar i långtidssjukdomar på senare tid än vad tidigare undersökningar visar. Detta stöder en del av antagandet i problemformuleringen 2.2, om att löpande genomföra sjuklighetsundersökningar.

De estimerade t-frekvenserna för Pb-försäkringar i denna studie visar på liknande resultat som för FTN's rapport 1974, se *Tabell 4.2.3*. Tidsskillnaden mellan dessa två undersökningar är dock så betydande att inga direkta slutsatser bör grundas på detta resultat

Tabell 4.2.4 och *4.2.5* indikerar att det förekommer stora avvikelser i insjuknandefrekvensen för långtidssjukdomar mellan sjuk- och Pb-försäkringar för samtliga

indelningar. Detta bekräftar antagandet om att utnyttjandegraden är lägre för Pb- i jämförelse med sjukförsäkringar. Gällande sjukförsäkringar pekar studien på att skillnad finns mellan män och kvinnors observerade t-frekvenser både för rörlig och för 3 månadskarens, däremot finns igen direkt skillnad mellan kön vad det gäller Pb-försäkringar. Avslutningsvis framträder det för Pb-försäkringar att avvikelser finns mellan de två karenstiderna, denna skillnad är dock inte lika stark för sjukförsäkringar.

Tabell 4.2.1, *Observerade och beräknade t-frekvenser för sjukförsäkringar jämförs med FTN's sjukundersökning och 84 års grunder.*

1000*t-frekvensen för män och kvinnor med karenstiden tre månader.

Duration, $t \geq 3$ månader.

Ålder	Obs. t-fr.		G84		FTN 93	
	Kvinnor	Män	Kvinnor	Män	Kvinnor	Män
27	9,6	4,9			6,6	2,7
30			7,3	6,5		
32	10,9	5,8			7,1	2,8
37	17,3	8,1			6	3
40			10,5	9,4		
42	17,5	9,0			8,4	4,2
47	22,6	11,0	14,5	12,9	12,2	5,8
52	25,7	16,0	18,9	16,9	14,8	9,5
57	36,8	22,1	26,1	23,2	26,1	15,6
62	48,6	27,0	38,3	33,8	15,1	27,4

Tabell 4.2.2, *Observerade och beräknade t-frekvenser för sjukförsäkringar jämförs med FTN's sjukundersökning.*

1000*t-frekvensen för män och kvinnor med R-karens.

Duration, $t \geq 6$ månader.

Ålder	Obs. t-fr.		FTN 93	
	Kvinnor	Män	Kvinnor	Män
27	6,8	3,0	1,3	0,8
30				
32	5,7	4,2	1,4	1,4
37	11,0	6,6	2,3	1,9
40				
42	12,0	7,0	5,1	2,9
47	15,1	8,9	6,4	3,5
52	13,5	13,1	10,4	7,7
57	26,8	17,8	19,6	18,9
62	8,8	25,1	20,8	25,2

Tabell 4.2.3, Observerade och beräknade t-frekvenser för Pb-försäkringar jämförs med FTN's sjukundersökning 1974. 1000*t-frekvensen för män och kvinnor med karenstiden tre månader. Duration, $t \geq 3$ månader.

Ålder	Obs. t-fr.		FTN 74	
	Kvinnor	Män	Kvinnor	Män
27	1,9	1,9	0,5	0,9
32	2,1	2,2	0,9	1,4
37	3,5	2,7	1,3	2,3
42	4,4	2,9	2,1	3,1
47	6,0	4,7	3,4	4,6
52	7,8	6,5	4,4	7,2
57	11,5	10,7	5,7	11,4
62	8,7	11,7	7,5	15,2

Tabell 4.2.4 observerade t-frekvenser gällande 1997-2000. 1000*t-frekvenser, Utnyttjandegraden för R- karens, man och kvinna. Duration, $t \geq 3$ månader.

Ålder \ Kön	Pb R		Sj R	
	Kvinnor	Män	Kvinnor	Män
27	0,6	0,0	11,9	5,0
32	0,0	1,2	7,9	6,2
37	1,8	0,5	12,8	8,5
42	0,4	0,9	14,2	9,0
47	1,5	2,3	18,0	11,0
52	1,9	2,0	17,5	16,4
57	4,4	5,1	30,7	21,4
62	2,7	3,9	20,4	27,3

Tabell 4.2.5 observerade t-frekvenser gällande 1997-2000. 1000*t-frekvenser, Utnyttjandegraden för karenstiden tre månader, man och kvinna. Duration, $t \geq 3$ månader.

Ålder \ Kön	Pb 13		Sj 13	
	Kvinnor	Män	Kvinnor	Män
27	1,9	1,9	9,6	4,9
32	2,1	2,1	10,9	5,8
37	3,4	2,6	17,3	8,1
42	4,4	2,8	17,5	9,0
47	5,8	4,6	22,6	11,0
52	7,6	6,4	25,7	16,0
57	11,3	10,5	36,8	22,1
62	8,5	11,5	48,6	27,0

5 Avvecklingsprocesser

Detta kapitel behandlar i korthet de teorier och definitioner som ligger till grund för den icke-parametriska skattningen avseende avvecklingsfunktionen med tillhörande konfidensband/intervall. Informationen som den observerade datamängden genererar definieras i Avsnitt 5.1. Den process som ska förklara avvecklingar från sjukdom behandlas i 5.2. Avslutningsvis diskuteras i 5.3 vilken form av censurering eller trunkering som den observerade datamängden kan vara utsatt för.

5.1 Filtration och martingalteori

De definitioner som följer nedan är givna i Øksendal [2000] och Björk [1998]. Låt Ω beteckna utfallsrummet för den observerade datamängden, då definieras en σ -algebra \mathcal{F} på Ω som den familj av delmängder på Ω som har följande egenskaper:

- (i) $\emptyset \in \mathcal{F}$.
- (ii) $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$, där F^c är komplementet till F .
- (iii) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

En filtration $\underline{\mathcal{F}} = \{\mathcal{F}_t; t \in I\}$ är en växande följd av σ -algebror på sannolikhetsrummet (Ω, \mathcal{F}, P) , där P är ett sannolikhetsmått. En filtration har följande egenskaper:

- (i) $\mathcal{F}_t \subseteq \underline{\mathcal{F}}, \forall t \in I$.
- (ii) $s < t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t$.

Detta innebär att \mathcal{F}_t är högerkontinuerlig och växande.

En stokastisk process X_t är adapterad till en filtration $\underline{\mathcal{F}}$, om X_t är \mathcal{F}_t -mätbar $\forall t \in I$. Vidare gäller att en process som är adapterad med vänsterkontinuerliga inkrement är predikterbar och lokalt begränsad.

En sekvens av stokastiska variabler kallas för en martingal med avseende på filtrationen $\underline{\mathcal{F}}$, om för alla $t \geq 0$ gäller att

- (i) $E(|X(t)|) < \infty$;
- (ii) $E(X(t)|\mathcal{F}_s) = X(s), \quad \forall s \leq t$.

Avslutningsvis definieras en stokastisk process $\{X(t), t \geq 0\}$ som en "Brownsk rörelse" eller Wienerprocess om följande gäller:

- (i) $X(0) = 0$;
- (ii) För $0 \leq u < v \leq s < t$, så är de två stokastiska variablerna $X(v) - X(u)$ och $X(t) - X(s)$ oberoende.
- (iii) För $t > s$, $X(t) - X(s) \sim N(0, \sqrt{s-t})$.
- (iv) X består av kontinuerliga inkrement.

5.2 Räkneprocesser

Avvecklingar från sjukdom kan definieras utifrån en räkneprocess $\{N(t), t \geq 0\}$ som är en stokastisk process som räknar antalet förekomster av en viss händelse, då tiden t , fortskrider. För en mer utförlig framställning av räkneprocesser hänvisas till Gill [1984] och Aalen [1978]. Låt t variera i intervallet $[0, 1]$. En multivariat räkneprocess definieras då som

Definition 5.1 Multivariat räkneprocess

- (i) Varje delprocess N_i , $i = 1, 2, \dots, n$, är en högerkontinuerlig stegfunktion med värdet 0 vid $t = 0$, och antar ett ändligt antal steg, där varje steg är positivt med värdet 1.
- (ii) Två delprocesser N_i och N_j ($i \neq j$) tar inga steg samtidigt.

En räkneprocess kan således i denna studie betraktas som en stokastisk process med n komponenter som räknar antalet förekomster av tillfrisknad för n olika sjuka eller skadade individer där händelserna antas inträffa en och en. Varje komponent $N_i(t)$, kan ses som en stegfunktion av tiden t , som antar heltals värdena +1 eller noll. Funktionen antar värdet noll vid tiden noll och är högerkontinuerlig.

Låt dt beteckna ett litet tidsintervall intill t . Vidare, låter $\alpha_i(t)$ vara intensitetsprocessen för N_i , då kan under regularitetsvillkor en intensitetsprocess definieras enligt:

$$\alpha_i(t) \cdot dt = P[dN_i(t) = 1 | \mathcal{G}_{t-}], \quad (5.2.1)$$

där \mathcal{G}_t betecknar filtrationen för händelser fram till och med tiden t . I och med att $dN(t)$ antar värdet 0 eller 1, i det lilla tidsintervallet dt , gäller även att $E[dN(t) | \mathcal{G}_{t-}] = \alpha(t) \cdot dt$.

För varje person observeras nu den totala durationen, dvs längden på sjukdomen $T_i = t_i$ eller durationen upp till en viss förbestämd tidpunkt c_i . Antag följande räkneprocess för individ i :

$$N_i(t) = I\{T_i \leq t, T_i \leq c_i\} = I\{T_i \leq t\} \cdot I\{T_i \leq c_i\}, \quad (5.2.2)$$

där $I\{\cdot\}$ betecknar en indikatorvariabel. N_i anger således händelserna att individen i har avvecklats från sjukdom innan avgång p g a höger eller stokastisk censurering har inträffat. Det innebär att N_i endast kan ta ett steg då $T_i \leq c_i$ och $T_i \leq t_i$ vilket är detsamma som att individ i tillfrisknar eller avlider inom det observerade tidsintervallet. Vad kan då sägas om intensiteten $\alpha_i(t)$ i detta fall? Givet vad som hänt före tidsintervallet dt , är det möjligt att observera om individ i har avvecklats vid tiden $T_i \leq t$ eller om personen har censurerats vid tiden $c_i \leq t$ eller till sist om individen kvarstår i tillståndet sjuk/skadad och är ocensurerad. I de första två fallen är sannolikheten för att N_i ska ta ett steg i intervallet dt , noll. I det tredje fallet är antingen $T_i \in [t, t + dt]$ eller $T_i > t$. Enligt (3.1.5) är således sannolikheten för ett steg i intervallet dt , $\mu_i(t) \cdot dt$, där $\mu_i(t)$ är intensiteten för en sann avvecklingstidpunkt T_i , oberoende av i . Definiera nu:

$$R_i(t) = I\{T_i \geq t, c_i \geq t\}, \quad i = 1, 2, \dots, n \quad (5.2.3)$$

vilket leder till följande sannolikhet:

$$P[dN_i(t) = 1 | \underline{\Omega}_{t-}] = \mu_i(t) \cdot R_i(t) \cdot dt \quad (5.2.4)$$

Här indikerar således $R_i(t)$ om individ i står under risk för avveckling just före tiden t . Från (5.2.1) och (5.2.4) ses nu att intensiteten $\alpha_i(t)$ för den multivariata räkneprocessen, definierad enligt 5.1, kan skrivas som:

$$\alpha_i(t) = \mu_i(t) \cdot R_i(t), \quad i = 1, 2, \dots, n \quad (5.2.5)$$

Som tidigare antagits är intensitetsprocessen $\mu(t)$, densamma för samtliga individer, vilket gör det möjligt att aggregera de individuella processerna (5.2.2), på så sätt erhålls:

$$N(t) = \sum_{i=1}^n N_i(t) \quad (5.2.6)$$

Denna process beräknar det totala antalet observationer i intervallet $[0, t]$. Intensitetsprocessen för (5.2.6) kan, enligt (5.2.5) och genom de faktum att två avvecklingar inte sker samtidigt, skrivas som:

$$\alpha(t) = \mu(t) \cdot R(t) \quad (5.2.7)$$

där:

$$R(t) = \sum_{i=1}^n R_i(t), \quad (5.2.8)$$

här anger $R(t)$ det totala antalet personer som står under risk just före tiden t . $R(t)$ antas vara en icke-negativ stokastisk variabel som går att observera. En process med dessa egenskaper sägs vara predikterbar. $R(t)$ har även en begränsad variation. Vidare gäller

att en intensitet som är definierad enligt (5.2.7) har fördelen att riskfunktionen $R(t)$, godtyckligt kan förändras under den observerade tidsperioden. Om förändringarna av $R(t)$ beror på externa effekter kan detta modelleras med en växande filtration $\{\mathcal{F}_t\}$. Vidare antas $\mu(t)$ vara en icke-negativ deterministisk funktion och tolkas som den individuella intensiteten för en avveckling.

Nu kan en stokastisk process $M(t)$, definieras i det lilla intervallet dt enligt

$$dM(t) = dN(t) - \alpha(t)dt, \quad (5.2.9)$$

vars väntevärde betingat av all information fram till och med tiden just före t antar värdet noll, $E[dM(t) | \mathcal{F}_t] = 0$. På de sätt som process (5.2.9) är definierad ger Avsnitt 5.1 att denna process är en martingal. Genom att integrera (5.2.9) erhålls

$$M(t) = N(t) - \int_0^t \alpha(s)ds = N(t) - \beta(t), \quad (5.2.10)$$

där $\beta(t)$ betecknar den integrerade intensiteten att avgå från tillståndet sjuk. Även processen (5.2.10) är en martingal och används vid framtagande av konfidensgränser för en ickeparametrisk skattning baserad på Nelson-Aalen's ansats av avvecklingsfunktionen. Se avsnitt (6.1).

5.3 Censurering och Trunkering

För att utföra analys av data som behandlar tid till en viss händelse, i denna studie tid till tillståndet tillfrisknad/avliden från tillståndet sjuk/skadad, bör först specificeras vad för slags data som används. De observationer som i denna studie ska förklara ovan nämnda information kan vara representerad på ett sådant sätt att vissa problem uppstår. Den vanligaste typen av problem benämns censurering och är vanligt förekommande när sjukfallen i ett bestånd endast kan studeras under en relativt kort tidsperiod. Detta innebär att personer som fortfarande är sjuka vid observationsperiodens slut ger upphov till högercensurerad information. Ett annat fall som kan inträffa är trunkering, vilket innebär att vissa observationer utelämnas ur studien p.g.a. att beståndet är valt på ett sådant sätt att vissa individer inte har möjlighet att ingå i undersökningen. Ett tredje fall av censurering som kallas stokastisk censurering kan också förekomma. Stokastisk censurering kommer av att individer avgår från tillståndet sjuk p.g.a. andra orsaker än just tillfrisknad eller avliden, orsakerna kan vara varierande och i vissa fall härröra från fel i instansningsförfarandet. Denna typ av censurering kommer att behandlas som en högercensurering. För mer utförlig beskrivning av trunkering och censurering hänvisas till Klein och Moeschberger [1997].

Denna undersökning studerar tillfrisknad som sker efter 1 januari, 1997 vilket innebär att tillfrisknad som sker före detta datum utgår från studien, vilket gör att observationer med lång durationstid med större sannolikhet ingår i urvalet, detta kallas för vänstertrunkering. Denna trunkering kommer dock att förbises eftersom denna studie är intresserad av hur sjukligheten betar sig i det studerade intervallet I och kommer således att utesluta information om hur sjukligheten såg ut tidigare år. Detta eftersom sjukligheten erfarenhetsmässigt förändras över tiden. Ytterligare ett fall av vänstertrunkering ingår, vilket uppkommer av att det är endast durationstider längre än

en karenstid k , som studeras. Resultaten kommer således att underskatta sannolikheten för en avveckling och bör därför kompenseras. Vilka åtgärder som krävs för att ta hänsyn till högercensurering, vänstertrunkering och stokastisk censurering för den parametriska respektive ickeparametriska modellen beskrivs mer utförligt i Avsnitt 6.2 och 8.2 nedan.

6 Avvecklingsfunktionen

Avvecklingsfunktionen som även kallas överlevnadsfunktionen, definieras som den betingade sannolikheten att en person som insjuknar vid x års ålder kvarstår som sjuk t år senare. Låt T_i stå för den återstående sjuklighetstiden för individ i , $i = 1, 2, \dots, n$, d v s durationen för person i . T_i är oberoende lika fördelade icke-negativa stokastiska variabler med fördelningsfunktionen: $P(T_i \leq t | X = x) = F(t | x)$ och $f(t | x) = F'(t | x)$. Vidare kan således avvecklingsfunktionen specificeras som ett minus fördelningsfunktionen $F(t | x)$, se Kapitel 3.1. För enkelhetens skull kommer $\lambda(t, x)$ skrivas som $\lambda(t)$.

Detta kapitel kommer att skatta avvecklingsfunktionen under ett icke-parametriskt antagande, vilken kommer att jämföras med en parametrisk modell som behandlas i Kapitel 8. Till skillnad från en parametrisk modell kräver inte en icke-parametrisk modell att avvecklingstidpunkterna, vid närvaro av censureringar, behöver följa en viss specifik fördelning vilket gör den mer flexibel. Modellen är således användbar för att grafiskt kunna anpassa komplicerade modeller. Oberoende av vilken framställning som används är det intressant att studera intensitetsfunktionen $\mu(t)$, d v s intensiteten att en person tillfrisknar eller avlider i det lilla tidsintervallet $[t, t + dt]$ givet att personen är sjuk vid tiden t . Avsnitt 6.1 behandlar icke-parametriska modeller för att i 6.2 skatta den valda modellen med tillhörande konfidensband/intervall.

6.1 Icke-parametriska Modeller

Den modell som på senare tid används för att skatta avvecklingsfunktionen $\lambda(t)$, bygger på att estimeras den integrerade intensiteten $\beta(t)$ att gå från tillståndet sjuk/skadad med hjälp av Aalens ansats. Denna ansats antar kontinuerliga tidpunkter för avvecklingarna. Modellen grundar sig på räkneprocesser och martingalteorier, se Kapitel 5. Förutom att få punktskattningar ger metodiken även möjlighet att bilda approximativa konfidensintervall och konfidensband. Även grova skattningar av intensiteten att gå från tillståndet sjuk kan fås genom att studera lutningen på den skattade avvecklingsfunktionen, dessa skattningar är dock svåra att utföra och ger oprecisa värden. För att få en relativt god skattning av intensitetsfunktionen, baserad på en empiriskt skattad avvecklingsfunktion, kan utjämnande kärnfunktioner användas. För en mer detaljerad diskussion om denna teknik se Ramlau-Hansen [1983].

Det bör noteras att det även finns en diskret framställning för att modellera avvecklingsfunktionen, denna benämns Kaplan-Meiers skattning eller "Product-limit estimator", se Cox and Oakes [1984], och definieras som:

$$\lambda_{KM}(t) = \prod_{t_i < t} \left(1 - \frac{d_i}{r_i} \right) \quad (6.1.1)$$

där d_i är antalet avvecklingar i tidpunkten t_i , r_i är antalet under risk vid samma tidpunkt och d_i / r_i är maximum likelihood skattningen av intensiteten att avgå från tillståndet sjuk.

Avvecklingsfunktionen kommer i denna studie baseras på Aalens ansats. För att erhålla en skattning av funktionen $\lambda(t)$, studeras således först den integrerade intensiteten:

$$\beta(t) = \int_0^t \mu(s) ds. \quad (6.1.2)$$

Genom att finna en skattning för $\beta(t)$ fås även en skattning av avvecklingsfunktionen, $\lambda(t)$,

Intensiteten att avgå från tillståndet sjuk $\mu(t)$, kan skrivas om i termer av räkneprocesser och riskfunktioner genom att använda martingal processen $dM(t) = dN(t) - \alpha(t) \cdot dt$ och relationen $\alpha(t) = \mu(t) \cdot R(t)$, definierad i Kapitel 5, av detta följer att:

$$dN(t) = \mu(t)R(t)dt + brus \quad (6.1.3)$$

eftersom $E[dM(t) | \mathcal{G}_t] = 0$. Efter lite omskrivning går det nu att uttrycka skattningen $\beta^*(t)$ som:

$$\beta^*(t) = \int_0^t \frac{J(s)}{R(s)} dN(s), \quad (6.1.4)$$

här står $J(t)$ för indikatorfunktionen $1\{R(t) > 1\}$. Funktionen är till för att undvika möjligheten att antalet under risk just före tiden t är lika med noll. Skattningen (6.1.4) introducerades av Aalen (1978) som en generalisering av den empiriska kumulativa intensitetsfunktion som föreslogs av Nelson (1969). Skattningen (6.1.4) benämns därför Nelson-Aalen skattningen.

Nelson-Aalen skattningen kan nu representeras av en enkel summa. Antag nu att avvecklingar sker i distinkta tidpunkter $t_1 < t_2 < \dots$. För varje tidpunkt t_i antar funktionen $dN(t_i)$ värdet ett d v s stegfunktionen $N(t_i)$ tar ett steg, annars gäller att $dN(t_i) = 0$. Detta innebär att (6.1.4) kan skrivas som en summa enligt:

$$\beta^*(t) = \sum_{t_i < t} \frac{1}{R(t_i)} \quad (6.1.5)$$

detta är en icke avtagande, högerkontinuerlig stegfunktion, där $R(t_i)$ är antalet under risk just före tiden t_i .

Det går nu att skapa en punktvis skattning av avvecklingsfunktionen genom att kombinera (3.1.7) och (6.1.5), skattningen blir då:

$$\lambda^*(t) = \exp\left(-\sum_{i:t_i < t} \frac{1}{R(t_i)}\right). \quad (6.1.5)$$

6.2 Skattning av avvecklingsfunktionen

Observationsperioden startar den 1 januari 1997 och avslutas den 31 december 2000 men för pågående sjukfall beräknas durationen per 31 augusti 2001. Notera att det är endast längden på durationen samt insjuknande ålder som är av intresse för en sjukundersökning, inte när personen insjuknade, så länge en individ varit observerad sjuk med en durationstid $t > k$ i det undersökta intervallet.

Inträdestidpunkten definieras av:

k	om insjuknandetidpunkten inträffar från och med 1997-01-01
$1997-01-01 - \text{insjuknandetidpunkten} + k$	om insjuknandetidpunkten inträffar före 1997-01-01

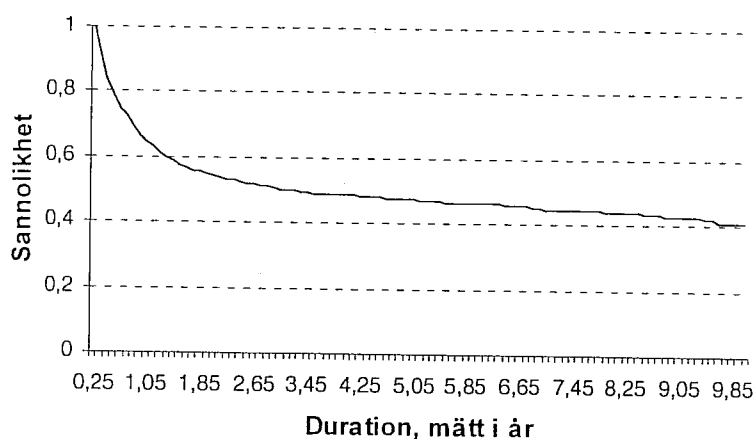
Då antalet sjukfall som inträffat i en viss ålder x , är ganska litet, indelas insjuknandeåldrarna i klasserna, - 24, 25 -34, 35-44,, 55-64 vilka betecknas med 20, 30 osv. Data materialet är, som i fallet med skattningen av t -frekvensen, uppdelat efter kön, sjuk- eller Pb-försäkring. De två karenstiderna som använts är som tidigare R respektive tremånadskarens

Notera att risksumman $R(t)$ är framtagen på så sätt att observationerna är vänstertrunkerade, detta leder till att Nelson-Aalen skattningen för avvecklingsfunktionen vid tiden t , definieras som sannolikheten att fortfarande vara sjuk vid t , betingad av att någon gång vara sjuk längre än tre månader, dvs i detta fall karenstiden.

Ovan nämnda resonemang gäller för karenstiden exakt 3 månader, dock bör den rörliga karensen R , behandlas något annorlunda då tidslängden på denna karens är olika för olika individer och är i regel betydligt längre än 3 månader. Ett sätt att studera R -karensen vore att sätta durationstidpunkten till noll vid påbörjad förtidspension oavsett längden på sjuktiden t . Två modeller kan då vara av intresse att studera, sannolikheten att avvecklas från tillståndet förtidspensionerad samt en funktion baserad på ålder och tidslängd fram till att den försäkrade blir berättigad till förtidspension. Dock gör bristen på information hos LFAB en sådan analys omöjlig. Detta gör att beräkningar avseende R -karensen i denna rapport kommer att genomföras på samma sätt som för karensen 3 månader vilket för långtidssjukförsäkringar innebär att sannolikheten att avgå från tillståndet sjuk/skadad kommer att överskattas eftersom beståndet rörlig karens till stor del även innehåller försäkringar med 3 månaderskarens. Avvecklingsfunktionen avseende premiefrielseförsäkringar kommer dock att underskattas för låga värden på t eftersom risksumman kommer att överskattas under ovan nämnda beräknings antaganden.

För att illustrera utseendet på den icke-parametriska Nelson-Aalen skattningen av avvecklingsfunktionen, uppritas i *Diagram 6.1* kurvan för gruppen kvinnor i åldersklassen 50 år med 3 månaderskarens för beståndet sjukförsäkringar. För att kunna jämföra olika skattningar av avvecklingsfunktionen ansätts normeringen $\lambda(t)/\lambda(k)$, där k är karenstiden, enligt definition innebär detta att $\lambda(k) = 1$.

Diagram 6.1 *Illustration på sannolikheten att avvecklas från sjukdom.*



I Tabell 6.1 till 6.4 nedan, jämförs de erhållna skattningarna med 1984 års grunder, FTN sjukundersökning 1993 och med Håkan Österbergs sjukundersökning 1990. Estimerade avvecklings sannolikheter för Pb-försäkringar jämförs med FTN premiefrielseundersökning 1974 och som nämndes tidigare finns ingen senare undersökning gjord för denna typ av försäkring. I likhet med skattningen av t-frekvensen har dessa undersökningar en karensindelning med avseende på exakt en månad, exakt 3 månader och längre än tre månader vilket innebär att gruppen R-karens jämförs med längre än tre månader och tre månaderskarens jämförs med tre månader. Notera att analysen avseende den rörliga karensen R, är något bristfällig p g a ofullständigt beräkningsunderlag.

Resultaten nedan visar män och kvinnor i åldersklass 50 år, sannolikheterna är multiplicerade med 1000 för att skattningarna ska bli mer överskådliga. Övriga resultat visas i Bilaga B.

Skattningarna som förklarar utseendet på avvecklingsfunktionen antyder i denna studie, att sannolikheten att avgå från tillståndet sjuk till något av tillstånden frisk eller avliden, för både långtidssjuk- och Pb-försäkringar, är något lägre i jämförelse med tidigare studier. Resultaten visar även att det är stor skillnad mellan avvecklingsfrekvensen mellan sjuk- och Pb-försäkringar. Detta kan tyda på, som nämndes i inledningen, att de personer med Pb-försäkring som anmäler, eller i vissa fall snarare upptäcker att de har en sådan försäkring, i medel har en allvarligare sjukdom eller skada än vad fallet är för personer med sjukförsäkringar. Detta resultat motsäger antagandet om att det inte är någon skillnad mellan avvecklingsfunktionen för de två försäkringstyperna.

Tabell 6.1, *Estimerade värden på avvecklingsfunktionen för Män med 3 mån. karens. $1000*\lambda(t)$.*

Duration	Observerad skada		Sjuk			Pb
	Pb	Sjuk	G84	FTN 93	Österberg	FTN 74
0,25	1000	1000	1000	1000	1000	1000
0,5	939	815	673	763	717	790
1	870	655	407	627	506	484
2	792	545	374	551	406	322
3	742	502	360	528	364	267
5	694	472	339	493	355	200
10	637	409	304	423	299	173

Tabell 6.2, *Estimerade värden på avvecklingsfunktionen för Män med R-karens. $1000*\lambda(t)$.*

Duration	Observerad skada		Sjuk			Pb
	Pb	Sjuk	G84	FTN 93	Österberg	FTN 74
0,25	1000	1000	1000	1000	1000	-
0,5	983	820	673	744	776	-
1	966	682	336	551	611	-
2	957	589	163	401	488	-
3	901	552	155	351	349	-
5	857	520	142	310	349	-
10	760	489	118	281	349	-

Tabell 6.3, *Estimerade värden på avvecklingsfunktionen för Kvinnor med 3 mån. karens. $1000*\lambda(t)$.*

Duration	Observerad skada		Sjuk			Pb
	Pb	Sjuk	G84	FTN 93	Österberg	FTN 74
0,25	1000	1000	1000	1000	1000	1000
0,5	967	823	673	763	745	821
1	974	679	407	627	570	496
2	861	589	374	551	452	342
3	767	549	360	528	413	293
5	693	520	339	493	373	230
10	623	491	304	423	295	200

Tabell 6.4, *Estimerade värden på avvecklingsfunktionen för Kvinnor med R-karens. $1000*\lambda(t)$.*

Duration	Observerad skada		Sjuk			Pb
	Pb	Sjuk	G84	FTN 93	Österberg	FTN 74
0,25	1000	1000	1000	1000	1000	-
0,5	997	820	673	883	800	-
1	969	682	407	732	740	-
2	959	589	374	639	704	-
3	923	552	360	620	704	-
5	885	520	339	-	602	-
10	791	489	304	-	602	-

För att klarlägga om de framtagna skattningarna i studien avviker signifikant från varandra, kommer skillnader undersökas med konfidensintervall/band samt med ett icke-parametriskt test.

6.2.1 Konfidensintervall

Till skillnad från skattningen av t-frekvensen förekommer en viss osäkerhet vid skattningar av avvecklingsfunktionen. Nedan framställs i korta drag ett konfidensintervall för avvecklingsfunktionen baserad på Nelson-Aalens ansats, för en mer utförlig härledning av konfidensintervallen hänvisas till Arne Sandström⁵ [1990]

För att finna ett uttryck för denna osäkerhet undersöks det asymptotiska utseendet på:

$$U_n(t) = \sqrt{n}[\lambda(t) - \lambda^*(t)], \quad (6.2.1.1)$$

Detta uttryck kan dock med Taylors formel skrivas om i termer av den integrerade intensiteten $\beta(t)$, för att på så sätt återknyta till Aalens ansats med tillhörande teorier. Uttrycket (6.2.1.1) kan skrivas som:

$$\lambda(t)\sqrt{n}[\beta^*(t) - \beta(t)] + \text{brus} \quad (6.2.1.2)$$

Det asymptotiska utseendet på (6.2.1.1) är det samma som för (6.2.1.2), varför ett uttryck för den senare studeras.

Teorin ger att skattningen $\beta^*(t)$ är en likformig skattning av $\beta(t)$, vilket innebär att även $\lambda^*(t)$ är en likformig skattning av $\lambda(t)$, i synnerhet kommer $\lambda^*(t)$ enligt Centrala Gränsvärdessatsen för martingaler då n blir stort, för theorem se Andersen Borgan [1985], konvergera mot en normalfördelad variabel med väntevärde $\lambda(t)$ och med en varians som kan definieras som:

$$C^*(t, t) = \lambda^{*2}(t)a^*(t) \quad (6.2.1.3)$$

Där:

$$a^*(t) = \sum_{i: t_i \leq t} \frac{n}{R(t_i)^2}, \quad (6.2.1.4)$$

där $a^*(t)$ följer av Centrala Gränsvärdessatsen för martingaler då n blir stort, avseende uttrycket (6.2.1.2) där $\lambda(t)$ är borttaget.

⁵ Ursprungligen kommer härledningen från Andersen Borgan [1985].

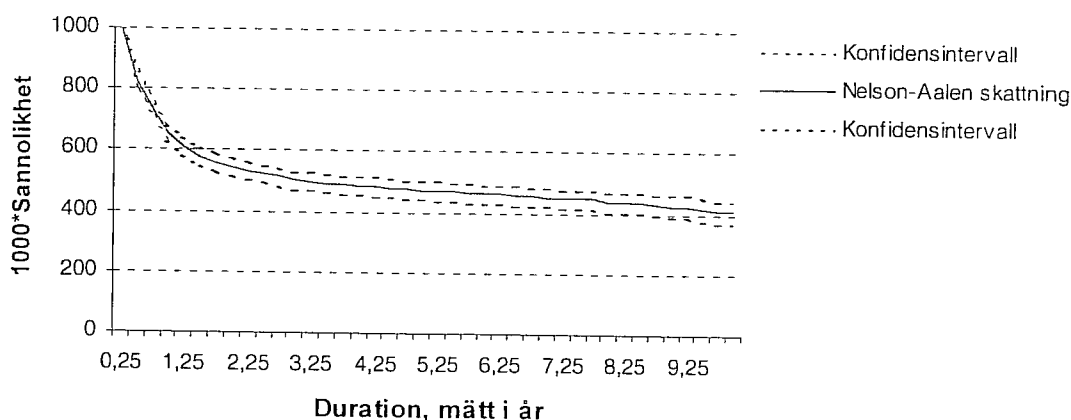
Ett punktvis konfidensintervall för $\lambda(t)$ för något t är då det asymptotiska konfidensintervallet

$$\lambda(t) \in \left[\lambda^*(t) \pm z_\alpha \lambda^*(t) \left\{ \frac{a^*(t)}{n} \right\}^{1/2} \right], \quad (6.2.1.5)$$

där $\Phi_s(z) = 1/2 (1 - \alpha)$, Φ_s är en standardnormal fördelningsfunktion och z_α α -kvantilen.

Exempel på de estimerade punktvisa konfidensintervallen ges i *Diagram 6.2*. De skattade intervallgränserna är uppritade med tillhörande Nelson-Aalen skattning för som ovan gruppen kvinnor i åldersklassen 50 år med 3 månadskarens för beståndet långtidssjukförsäkringar. Övriga grupperns konfidensintervall återfinns i Bilaga C representerade tillsammans med motsvarande konfidensband och Nelson-Aalen skattning. För att förtydliga förhållandet mellan den empiriska skattningen och dess konfidensgränser har kurvorna i Bilaga C valts att ritas med differentierade värden på y-axeln.

Diagram 6.2 Nelson-Aalen skattningen med tillhörande konfidensintervall



Med stöd av 95 % konfidensintervallen kan signifikant skillnad mellan sjuk- och Pb-försäkringar bekräftas för samtliga grupper, förutom för åldersklassen 50 år med 3 månadskarens.

De kritiska gränserna gällande Pb-försäkringar visar att det förekommer signifikanta skillnader för både karenstyp och kön.

För sjukförsäkringar indikerar skattningarna av konfidensintervallen att ingen skillnad finns mellan de två karenstyperna, vilket inte var förvånande eftersom dessa två bestod till stor del bestod av identiska försäkringstagare, se avsnitt Data 2.4. Liknande resultat gäller för förhållandet mellan män och kvinnor, dock avviker åldersklassen 40 år från dessa resultat.

Notera dock att det är just i avvecklingstidpunkterna som det existerar ett konfidensintervall och inte mellan dessa tidpunkter. Detta föranleder att även se på konfidensband till skattningen av avvecklingsfunktionen.

6.2.2 Konfidensband

För att vara säker på att avvecklingsfunktionen även mellan tidpunkterna för avgångar ligger inom ett visst osäkerhetsintervall måste ett konfidensband beräknas. Framställning återfinns i Arne Sandström [1990]⁶ och Klein och Moeschberger [1997]. Vid framtagandet av ett konfidensband för $\lambda(t)$, antas det existera en stopptid⁷ T^* , sådan att $0 \leq t < T^* < \infty$ och en fördelning $F(T^*) < 1$. Vidare definieras processen $Z_n(t) = \sqrt{n} \sqrt{\lambda(t) - \lambda^*(t)}$, vars gränsvärde följer en normalfördelad martingal Z^* , där $Z^*(0) = 0$.

Notera att fördelningen för en normalfördelad process bestäms helt av sitt väntevärde och sin varians, detta gör att processen $Z(t)$ kan skrivas om till $Z(t) / K(t)$ där $K(t) = \lambda(t)[1 + a(t)]$. Processen $Z(t) / K(t)$ har då samma fördelning som:

$$W^0 \left\{ \frac{a(t)}{1 + a(t)} \right\}, \quad \text{för } 0 \leq t < T^* \quad (6.2.2.1)$$

där W^0 är en "Brownsk brygga" process⁸. Det faktum att $Z(t) / \lambda(t)[1 + a(t)]$ och $W^0\{;\}$ har samma normalfördelning följer av att båda har väntevärde lika med noll samt samma kovariansstruktur.

Nu kan ett $100^*(1 - \alpha)$ procent konfidensband för $\lambda(t)$ skrivas som:

$$\lambda(t) \in \left[\lambda^*(t) \pm c_\alpha \lambda^*(t) \frac{[1 + a^*(t)]}{\sqrt{n}} \right], \quad (6.2.2.2)$$

där c_α är den övre α -kvantilen i fördelningen för $\sup_{t \in [0, T^*]} |W^0(t)|$. För att erhålla värden på c_α

observeras först tidpunkten t_u som är större eller lika med den minsta avvecklingstidpunkten och t_θ som är mindre eller lika med den största observerade avvecklingstidpunkten. På basis av de två tidpunkterna t_u och t_θ , definieras två mått:

$$K_u = \frac{n\sigma_\lambda^2(t_u)}{1 + n\sigma_\lambda^2(t_u)}$$

och

$$K_\theta = \frac{n\sigma_\lambda^2(t_\theta)}{1 + n\sigma_\lambda^2(t_\theta)} \quad (6.2.2.3)$$

⁶ Ursprungligen kommer härledningen från Andersen Borgan [1985] och i Klein och Moeschberger från Hall och Wellner

⁷ En stokastisk variabel T^* , som antar värden inom $[0, \infty)$ är en stopptid för filtrationen Γ om för varje $t \geq 0$ gäller att, definitionen återfinns i Andersson och Djehichle [1997]:

$$\{T^* = t\} \in \Gamma_t$$

⁸ En "Brownsk brygga" process, definition hämtad ur Steal [2001], är en normalfördelad process $\{X_t\}$ definierad på intervallet $[0, 1]$ sådan att $\text{Kov}(X_s, X_t) = s \cdot (1 - t)$ för $0 \leq s \leq t$. Denna process kan även representeras i termer av en standard wienerprocess⁹, enligt:

$$X_t = B_t - t \cdot B_1 \quad \text{för } 0 \leq t \leq 1$$

En "Brownsk brygga" process är således en wienerprocess som är begränsad att återgå till 0 vid tiden 1.

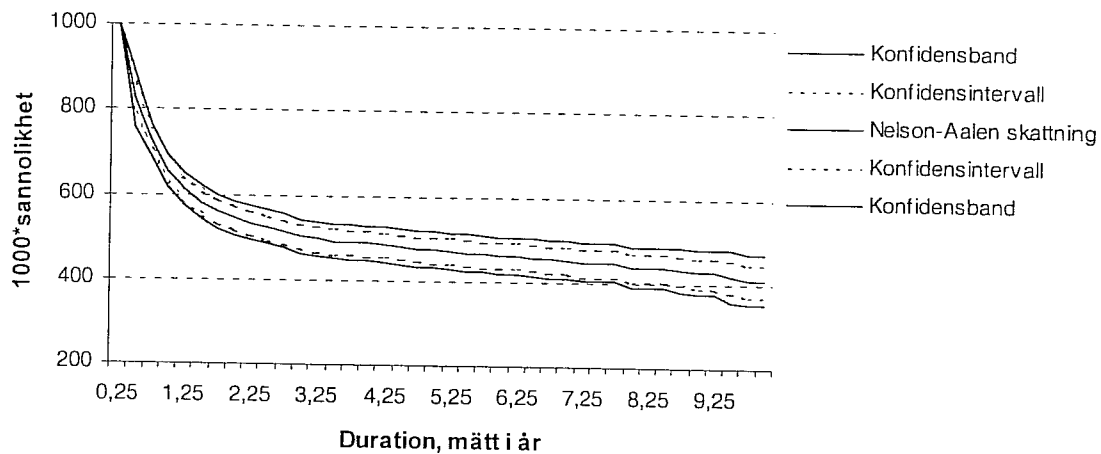
där n är storleken på det observerade beståndet. Värden som (6.2.2.3) genererar används för att finna kritiska värdena på c_α

De estimerade konfidensbanden illustreras tillsammans med motsvarande avvecklingsfunktion och konfidensintervall i *Diagram 6.3* nedan. Skattningar för de resterande grupperna återges i Bilaga C, tillsammans med motsvarande konfidensintervall och ickeparametrisk skattning.

Diagram 6.3 visar att de kritiska gränserna för konfidensbanden är något bredare än för motsvarande konfidensintervall, vilket är självklart. Dessa band visar inte lika otvetydigt som konfidensintervallen, att signifikant skillnad råder mellan sjuk- och Pb- försäkringen, trots det visar även konfidensbanden att beräkningar baserade avseende avvecklingsfunktionen bör skilja mellan långtidssjuk- och Pb-försäkring.

Till skillnad från de skattade konfidensintervallen visar konfidensbanden, gällande Pb-försäkringar, att signifikant skillnad endast finns mellan de två karenstyperna för durationstider större än två år. Vidare liknar resultaten för konfidensbanden de resultat som framkom av konfidensintervallen.

Diagram 6.3 Nelson-Aalen skattningen med tillhörande konfidensintervall och konfidensband



För att bekräfta eller dementera osäkerheten i dessa skattningar genomförs i Kapitel 7 nedan, icke-parametriska test för att undersöka om det förekommer signifikant skillnad mellan två bestämda urval.

7 ICKE-PARAMETRISKT TEST

Detta avsnitt kommer att undersöka om det förekommer skillnader mellan två grupper genom att ställa upp en nollhypotes, som istället för att direkt jämföra två resultat, jämför viktade skillnader mellan observerade och förväntade intensitetsvärden. Olika vikter kan presenteras på ett sådant sätt att testet är mer känsligt för korta eller långa durationstider vilket kan vara av intresse då de flesta avvecklingar från sjukfall inträffar för små durationstider. En mer utförlig beskrivning återfinns i Klein och Moeschberger [1997].

Testet kommer att jämföra två grupper åt gången och test hypotesen ges av:

$$H_0: \mu_1(t) = \mu_2(t), \quad \text{för } t < \tau,$$

Testet kommer således att jämföra intensitetsvärden i alla tidpunkter lägre än τ , där τ är den största observerade tidpunkten för en avveckling i de två undersökta mängderna. Den alternativa hypotesen är global i den mening att nollhypotesen förkastas om det i någon tidpunkt föreligger skillnad.

Datamängden som studeras i studien innehåller både högercensurerade och vänstertrunkerade observationer, detta är dock inget hinder. Låt $t_1 < t_2 < \dots < t_D$, vara disjunkta tidpunkter för övergång från tillstånd sjuk till antingen tillståndet frisk eller avliden. Index D antas vara det totala antalet avvecklingstidpunkter för de två undersökta urvalen. Vid tidpunkten t_i observeras d_{ij} avvecklingar för grupp 1 bland R_{ij} observerade personer som står under risk. Låt R_i och d_i stå för det totala antalet avvecklingar respektive personer under risk för de två grupper som ska jämföras.

Den viktade testhypotesen H_0 är baserad på Nelson-Aalen skattning av den integrerade intensiteten $\beta(t)$. Om nollhypotesen är sann kommer skattningen av den förväntade intensiteten vara lika med den aggregerade intensiteten d_i/R_i . Testet av H_0 kommer således att baseras på statistikan:

$$Z_j(\tau) = \sum W(t_i) \cdot \left(d_{ij} - R_{ij} \cdot \frac{d_i}{R_i} \right), \quad j = 1, 2. \quad (7.1)$$

där $W(t_i)$ är en viktfunktion som antar värdet noll då antalet under risk är noll. Statistikan (7.1) är således den viktade differensen mellan det observerade antalet övergångar och det förväntade antalet övergångar under H_0 för den observerade gruppen. Det förväntade antalet övergångar antas vara proportionen under risk för den undersökta gruppen R_{ij}/R_i , multiplicerat med antalet avvecklingar vid tiden t_j .

Variansen för Z ges av:

$$\sigma_{jj} = \sum W(t_i)^2 \frac{R_{i1}}{R_i} \cdot \frac{R_{i2}}{R_i} \left(\frac{R_i - d_i}{R_i - 1} \right) \cdot d_i, \quad j = 1, 2. \quad (7.2)$$

Termen $(R_i - d_i)/(R_i - 1)$ i formel (7.2), antar värdet ett om det sker en ensam avveckling i en viss tidpunkt t_i d v s termen är en korrektion för avvecklingar som är noterade i samma tidpunkt. De övriga två termerna i uttrycket härrör från variansen av en multinomial stokastisk variabel med parametrarna d_i , $p_i = R_{ij}/R_i$, $j = 1, 2$.

Avslutningsvis kan teststatistikan för två urval med en kombination av formel (7.1) och (7.2) skrivas som:

$$T_j = \frac{Z_j}{\sigma_{jj}}, \quad j = 1, 2, \quad (7.3)$$

för stora urval, och när H_0 är sann har (7.3) en standard normalfördelning.

En viktig aspekt när ett sådant test skall utföras är valet av viktfunktion. I de flesta undersökningar används ett s.k. log-rank test som antar en viktfunktion $W(t) = 1$. Denna vikt ger optimal styrka till test om intensiteterna i de två undersökta urvalen är proportionerliga mot varandra. Vidare finns andra test som tar mer hänsyn till tidigare avvecklingar, en av dessa vikter kallas för Gehan och ges av $W(t_i) = R_i$ som är en generalisering av Mann-Whitney-Wilcoxon testet. En mer flexibel viktfunktion baserad på Product-Limit skattningen (6.2.1) av avvecklingsfunktionen framtofs av Fleming och Harrington (1981) där log-rank testet ingår som specialfall. Denna funktion som baseras på det aggregerade urvalet kan skrivas som:

$$W_{p,q}(t_i) = \hat{\lambda}(t_{i-1})^p \left(1 - \hat{\lambda}(t_{i-1})\right)^q, \quad p > 0, \quad q > 0 \quad (7.4)$$

här används skattningarna av avvecklingsfunktionen, förskjutna ett tidssteg bakåt i tiden, detta för att försäkra sig om att skattningarna av viktfunktionen är kända i den tidpunkt som jämförelsen ska göras. Då parametrarna $p = 0$ och $q = 0$ erhålls log-rank testet ovan samt när $q = 0$ och $p > 0$ läggs mer vikt på tidigare övergångar.

Studien kommer att genomföra både ett log-rank test samt ett Gehan test, detta för att se om det finns någon skillnad mellan dessa. En viss skillnad kan förekomma eftersom de flesta avvecklingarna sker för små värden på durationen t . Testet är uppdelat så att sjukförsäkringar jämförs med premiebefrielseförsäkringar för de olika indelningarna. Vidare prövas förhållandet mellan män och kvinnor samt förhållandet mellan de två undersökta karens-typerna.

7.1 Log-rank test

I detta test gäller som ovan nämnts att vikten $W(t_i)$ är lika med ett. Detta val av vikt är optimalt om urvalen är proportionerliga mot varandra dvs om intensiteterna under nollhypotesen H_0 , är proportionerliga mot varandra. De urval som skattas antas vara av den storleksordningen att statistikan (7.3) anses vara approximativt normalfördelad, statistikan kan då skrivas som:

$$Z_{\text{Log-rank}} = \frac{\sum_{i=1}^D \left(d_{i1} - R_{i1} \cdot \left(\frac{d_i}{R_i} \right) \right)}{\sqrt{\sum_{i=1}^D \frac{R_{i1}}{R_i} \cdot \frac{R_{i2}}{R_i} \cdot \left(\frac{R_i - d_i}{R_i - 1} \right) \cdot d_i}} \quad (7.1.1)$$

De värden som erhållits för de olika indelningarna visas i *Tabell 7.1*, *7.2* och *7.3* nedan för respektive indelning försäkring, kön och karens.

Log-rank testet ger att det för samtliga indelningar råder signifikanta skillnader på 5 % nivån för de två försäkringstyperna, *Tabell 7.1*, liknande resultat gäller för de två karenstyperna gällande Pb-försäkringar, *Tabell 7.2*. Övriga indelningar indikerar i de flesta fall att hypotesen inte kan förkastas på 5 % signifikansnivå, dock finns vissa indelningar som indikerar det motsatta, exempelvis åldersklass 40 för sjukförsäkringar som klart visar på skillnad mellan man och kvinnor, *Tabell 7.2*. I stort bekräftas de resultat som kom av konfidensgränserna i Kapitel 6.2.1 och 6.2.2.

Tabell 7.1, Log-rank test, skillnad mellan långtidsjuk respektive premiefrielseförsäkring.

		Värden\Ålders grupp	35-44	45-54	55-64
Kvinna	R- karens	Z- värde	9,05	9,06	4,53
		p- värde	<0,001*	<0,001*	<0,001*
	3- månadskarens	Z- värde	5,81	7,32	2,18
		p- värde	<0,001*	<0,001*	0,029*
Man	R- karens	Z- värde	9,05	9,06	4,53
		p- värde	<0,001*	<0,001*	<0,001*
	3- månadskarens	Z- värde	9,05	9,06	4,53
		p- värde	<0,001*	<0,001*	<0,001*

* indikerar signifikant skillnad på 5 %-nivån

Tabell 7.2, Log-rank test, skillnad mellan kvinnor och män.

		Värden\Ålders grupp	35-44	45-54	55-64
Pb	R- karens	Z- värde	0,19	1,18	1,19
		p- värde	0,85	0,239	0,235
	3- månadskarens	Z- värde	2,58	1,19	1,03
		p- värde	0,01*	0,236	0,305
Sjuk	R- karens	Z- värde	5,33	1,68	1,54
		p- värde	<0,001*	0,092	0,122
	3- månadskarens	Z- värde	5,26	1,67	1,33
		p- värde	<0,001*	0,094	0,183

* indikerar signifikant skillnad på 5 %-nivån

Tabell 7.3, Log-rank test, Skillnad mellan R respektive 3 månadskarens

		Värden\Ålders grupp	35-44	45-54	55-64
Pb	Kvinna	Z- värde	5,61	5,64	3,74
		p- värde	<0,001*	<0,001*	<0,001*
	Man	Z- värde	8,16	5,57	4,95
		p- värde	<0,001*	<0,001*	<0,001*
Sjuk	Kvinna	Z- värde	0,45	0,005	0,24
		p- värde	0,450	0,996	0,808
	Man	Z- värde	0,44	0,006	0,08
		p- värde	0,663	0,995	0,933

* indikerar signifikant skillnad på 5 %-nivån

7.2 Gehan test

Ett Gehan test antar vikten $W(t_i) = R_i$, d.v.s. stor vikt läggs på avvecklingar som sker då antalet under risk är stort. Eftersom riskfunktionen (5.3.8) i denna studie är avtagande kommer mest hänsyn tas till tidiga övergångar. Eftersom de undersökta delmängderna i detta test är desamma som i Log-rank testet antas statistikan (7.2) vara approximativt normalfördelad och kan skrivas som:

$$Z_{\text{Gehan}} = \frac{\sum_{i=1}^D R_i \left(d_{i1} - R_{i1} \cdot \left(\frac{d_i}{R_i} \right) \right)}{\sqrt{\sum_{i=1}^D R_i \frac{R_{i1}}{R_i} \cdot \frac{R_{i2}}{R_i} \cdot \left(\frac{R_i - d_i}{R_i - 1} \right) \cdot d_i}}, \quad (7.2.1)$$

de resultat som (7.2.1) genererar för respektive indelning visas i *Tabell 7.4*, *7.5* och *7.6* nedan för respektive försäkring, kön och karens.

De resultat som framkommer av detta Gehan test, se nedan, är i stort sett lika med de resultat som Log-Rank testet genererade, vilket bekräftar de skillnader och likheter, för de olika del bestånden, som framkommit tidigare i detta kapitel.

Tabell 7.4, *Gehan test, Skillnad mellan långtidssjuk respektive premiefrielseförsäkring.*

		Värden\Ålders grupp	35-44	45-54	55-64
Kvinna	R- karens	Z- värde	10,05	9,06	4,52
		p- värde	<0,001*	<0,001*	<0,001*
	3- månadskarens	Z- värde	6,71	8,71	2,57
		p- värde	<0,001*	<0,001*	0,029*
Man	R- karens	Z- värde	11,11	13,05	9,92
		p- värde	<0,001*	<0,001*	<0,001*
	3- månadskarens	Z- värde	9,92	13,93	5,60
		p- värde	<0,001*	<0,001*	<0,001*

* indikerar signifikant skillnad på 5 %-nivån

Tabell 7.5, *Gehan test, Skillnad mellan kvinnor och män*

		Värden\Ålders grupp	35-44	45-54	55-64
Pb	R- karens	Z- värde	0,18	0,87	1,51
		p- värde	0,86	0,39	0,13
	3- månadskarens	Z- värde	2,55	1,08	1,14
		p- värde	0,11	0,28	0,25
Sjuk	R- karens	Z- värde	4,44	1,28	1,63
		p- värde	<0,001*	0,20	0,10
	3- månadskarens	Z- värde	1,54	1,24	4,53
		p- värde	<0,001*	0,21	0,12

* indikerar signifikant skillnad på 5 %-nivån

Tabell 7.6, Gehan test, Skillnad mellan R respektive 3 månadskarens

		Värden\Ålders grupp	35-44	45-54	55-64
Pb	Kvinna	Z- värde	6,38	6,41	4,96
		p- värde	<0,001*	<0,001*	<0,001*
	Man	Z- värde	8,02	6,41	6,20
		p- värde	<0,001*	<0,001*	<0,001*
Sjuk	Kvinna	Z- värde	0,17	0,01	0,78
		p- värde	0,44	0,99	0,87
	Man	Z- värde	0,46	0,04	0,22
		p- värde	0,65	0,97	0,83

* indikerar signifikant skillnad på 5 %-nivån

8 Parametriska modeller för Avvecklingsfunktionen

I tidigare studier har det fokuserats på icke-parametriska fördelningar, så även i denna undersökning, för att förklara avvecklingar från tillståndet sjukdom/skadad. Dessa metoder kräver inga förbestämda antaganden om avvecklingsfunktionen. I följande kapitel kommer parametriska modeller undersökas för att skatta avvecklingsfunktionen. En parametrisk modell som har god anpassning till data tenderar att ge mer information om den underliggande mekanismen eftersom intensitetsfunktionen $\mu(t)$, får ett analytiskt uttryck. Tanken är också att skapa en enkel modell, i kontrast till den komplicerade modell som används idag, se Bilaga D, för att beräkna avvecklingsfunktionen, och med den ge en snabb, grov skattning på sjukreserver och premier baserade på senaste data. För att kunna utnyttja den parametriska modellen krävs givetvis att anpassningen blir bra.

De modeller som kommer att behandlas i denna studie antas ha en intensitetsfunktion som är monotont avtagande och begränsad vid observationsintervallens ändpunkter. Durationen t_i antas ligga i intervallet $I_i = [0.25, \tau_i)$, där τ_i är den största observerade tidpunkten för den undersökta delmängden i . Detta antagande är befogat eftersom flertalet parametriska modeller som förklarar överlevnadsdata och som för vissa parametrar antar en avtagande intensitet, divergerar mot oändligheten då $t \rightarrow 0$.

I Avsnitt 8.1 kommer ett antal modeller att studeras med avseende på de kriterier som nämnts ovan. Vidare definieras i Avsnitt 8.2 likelihoodskattningen baserad på data som är utsatt för censurering och trunkering. Avslutningsvis anges i Avsnitt 8.3 den parametriska modell som passar bäst till den observerade datamängden.

8.1 Modeller

Det är svårt att teoretiskt motivera vilken fördelning som ska användas, vilket ger anledning till att testa olika överlevnadsmodeller, i detta fall avvecklingsmodeller, för att på så sätt hitta en fördelning som passar bra. Fördelningen bör vara flexibel och lättarbetad och givetvis ha god anpassning till data. De överlevnadsmodeller vars struktur har studerats är hämtade ur Cox och Oaks [1984] och från Klein och Moeschberger [1997]. Följande modeller har undersökts: Weibull, Log logistisk, Exponential Power, en omgjord Gompertz och även fördelningarna Invers Gaussisk, Normal, Log-normal, Pareto. De fyra sistnämnda fördelningarna har visat sig ha en något sämre anpassning än de övriga modellerna och kommer således att uteslutas ur den fortsatta studien av parametriska modeller. Nedan definieras de modeller som kan vara av intresse för studien.

8.1.1 Weibull

Modell som framtogas av Weibull 1951 som även namngav denna fördelning, antogs beskriva livslängden för material. Avvecklings- och intensitetsfunktionen definieras enligt:

$$\lambda(t) = \exp[-\beta t^\alpha], \quad \beta, \alpha > 0, t \geq 0 \quad (8.1.1.1)$$

och

$$\mu(t) = \beta \alpha t^{\alpha-1}, \quad \beta, \alpha > 0, t \geq 0. \quad (8.1.1.2)$$

För specialfallet $\alpha = 1$ övergår Weibull fördelningen till den mer kända exponentialfördelningen. Fördelningens intensitetsfunktion har en viss flexibilitet i sin utformning, på så sätt att intensiteten växer då $\alpha > 1$, och avtar då $\alpha < 1$ och är konstant för $\alpha = 1$, utseendet beror således fullständigt på vilket värde α antar. Som synes ovan har denna fördelning enkla uttryck för både avvecklings- och intensitetsfunktionen, samtidigt som den har en viss flexibilitet. Detta har gjort den till en populär modell att använda. Som antyds i början av detta kapitel är det en Weibull modell med $\alpha < 0$ som är av intresse för denna studie.

8.1.2 Log logistisk

En variabel som följer en log logistisk fördelning är en variabel $Y = \log(T)$ där T följer en logistisk fördelning. Denna är närbesläktad med normalfördelningen men har betydligt enklare uttryck för både avvecklingsfunktionen och intensiteten att avgå från tillståndet sjuk, se nedan:

$$\lambda(t) = \frac{1}{1 + \beta t^\alpha}, \quad \beta, \alpha > 0, t \geq 0 \quad (8.1.1.3)$$

och

$$\mu(t) = \frac{\alpha \beta t^{\alpha-1}}{1 + \beta t^\alpha}, \quad \beta, \alpha > 0, t \geq 0. \quad (8.1.1.4)$$

Denna modell har för vissa värden på α en intensitetsfunktion som är monotomt avtagande. Eftersom uttrycken (8.1.1.3) och (8.1.1.4) är relativt enkla vore denna modell intressant att studera vidare.

8.1.3 Exponential Power

Denna fördelning har en badkarsformad intensitetsfunktion vilket innebär avtagande till en början för att sedan börja växa. Exponential power fördelningen har avvecklings- respektive intensitetsfunktion enligt:

$$\lambda(t) = \exp(1 - \exp(\beta t)^\alpha), \quad \alpha, \beta > 0, t \geq 0 \quad (8.1.1.5)$$

samt

$$\mu(t) = \alpha \beta^\alpha t^{\alpha-1} \exp(-(\beta t)^\alpha), \quad \alpha, \beta > 0, t \geq 0, \quad (8.1.1.6)$$

Om avvecklingsfunktionen skattas i ett så pass snävt intervall att funktionen är konstant avtagande, visar sig denna modell ha en relativt god anpassning till den observerade datamängden.

8.1.4 Gompertz

Modellerna Weibull och Log logistisk har för vissa värden på parametern α monotont avtagande intensitetsfunktioner men divergerar då t går mot noll. Exponential power har ett gränsvärde för intensiteten när t går mot noll men den behöver inte vara avtagande under hela det observerade intervallet. Detta föranleder till att undersöka en utveckling av Gompertz modell vars intensitetsfunktion ges av:

$$\mu(t) = \beta \exp(\alpha t), \quad \alpha, \beta > 0, t \geq 0,$$

denna är som synes strängt växande. Genom att sätta ett minustecken framför parametern α erhålls en strängt avtagande funktion som går mot β då t går mot noll, d v s divergerar ej.

Den omgjorda Gompertzfördelningen har då, enligt relationen (3.1.7), följande utseende på avvecklings och intensitetsfunktionen:

$$\lambda(t) = \exp\left[-\frac{\beta}{\alpha}(1 - \exp(-\alpha t))\right], \quad \alpha, \beta > 0, t \geq 0 \quad (8.1.4.1)$$

respektive

$$\mu(t) = \beta \exp(-\alpha t) \quad \alpha, \beta > 0, t \geq 0, \quad (8.1.4.2)$$

Ett problem med denna omgjorda Gompertzfördelningen är att väntevärdet inte existerar samt att sannolikheten för att övergå från sjukdom konvergerar mot $\exp(-\beta/\alpha)$ då t blir stort, vilket strider mot ett av de antaganden som avvecklingsfunktionen bygger på, se Kapitel 3, antagandet om att avvecklingsfunktionen ska konvergera mot noll då t går mot oändligheten. Detta är dock inget som nämnvärt stör teorin utan enbart ändrar lite på grundförutsättningarna. Avvecklingsfunktionen kan under detta antagande definieras som:

$$\lambda(t) = P(T > t) = \exp\left(-\int_0^t \mu(t) dt\right) \xrightarrow{t \rightarrow \infty} \exp\left(-\int_0^{\infty} \mu(t) dt\right) = P(T = +\infty) = e \geq 0.$$

d v s det finns en positiv sannolikhet att individer aldrig tillfrisknar inom det studerade intervallet. Vidare kan konvergensen av intensiteten $\mu(t)$ förklaras genom relation (3.1.6) på följande sätt:

$$\mu(t) = \frac{f(t)}{\lambda(t)} \rightarrow \left[f(t) \xrightarrow{t \rightarrow \infty} 0, \lambda(t) \xrightarrow{t \rightarrow \infty} e > 0 \right] \xrightarrow{t \rightarrow \infty} 0, \quad (8.4.1.3)$$

intensiteten (8.4.1.3) konvergerar således monotont mot noll. Eftersom överlevnadsteorin även stämmer för denna omgjorda Gompertz modell kan också denna fördelnings parametrar skattas, likt andra parametriska avvecklingsmodeller, med hjälp av en maximum likelihood skattning. I och med att intensiteten för att tillfriskna har acceptabla egenskaper undersöks även denna modell.

8.2 Likelihood

Maximum likelihood funktionen kommer att formuleras under antagandet att tidpunkterna i avvecklingsfunktionen är kontinuerliga. Om en person som är under observation avgår, tillförs likelihood funktionen termen $f(t, \Phi)$, där Φ står för parametervektorn, d v s värdet på tätheten i avvecklingstidpunkten t . Denna term måste kompenseras eftersom det observerade urvalet är vänstertrunkerat, se Avsnitt 5.3. Täthetsfunktionen måste således betingas på att individen är sjuk i minst tre månader, vilket gör att täthetsfunktionen för vänstertrunkerade data erhåller följande uttryck:

$$f_{VT}(t, \Phi) = \frac{f(t, \Phi)}{\lambda(0.25, \Phi)} \quad (8.2.1)$$

Den rörliga karensen kommer även den att vara betingad på att den försäkrade individen varit sjuk i minst tre månader, vilket gör att den parametriska skattningen avseende R-karensen kommer i grunden att vara oriktig. Återigen härrör detta från att den erhållna informationen är ofullständig, se även Avsnitt 5.3 och 6.2.

Det som en höger- eller stokastisk censurerad observation tillför modellen är värdet på avvecklingsfunktionen $\lambda(c, \Phi)$ i censureringstidpunkten c , d v s sannolikheten att fortfarande vara sjuk efter durationstiden c . Även denna sannolikhet måste betingas på samma sätt som täthetsfunktionen ovan. Likelihoodfunktionen L , för n oberoende individer, indexerade med i , ges av:

$$L = \prod_{IC} \frac{f(t_i, \Phi)}{\lambda(0.25, \Phi)} \cdot \prod_C \frac{\lambda(t_i, \Phi)}{\lambda(0.25, \Phi)} \quad (8.2.2)$$

där IC indikerar icke-censurerade data och C censurerade. Detta uttryck kan omformuleras med hjälp av relationerna (3.1.6) enligt:

$$L = \prod_{IC} \mu(t_i, \Phi) \cdot \prod_{IC+C} \frac{\lambda(t_i, \Phi)}{\lambda(0.25, \Phi)} \quad (8.2.3)$$

Här betecknar $\mu(t_i, \Phi)$ intensiteten avseende parametervektorn Φ . Avslutningsvis logaritmeras det högra uttrycket i (8.2.3) för att då erhålla log likelihooden, l :

$$l = \sum_{IC} \ln(\mu(t_i, \Phi)) + \sum_{IC+C} \ln\left(\frac{\lambda(t_i, \Phi)}{\lambda(0.25, \Phi)}\right) \quad (8.2.4)$$

Maximum log likelihood skattningen (8.2.4) är således den logaritmerade intensitetsfunktionen summerad över de individer vars exakta durationslängd kunnat observerats plus den logaritmerade avvecklingsfunktionen summerad över alla individer i det undersökta intervallet, betingat av att någon gång under tidsintervallet I vara sjuk med en duration $t \geq 3$ månader.

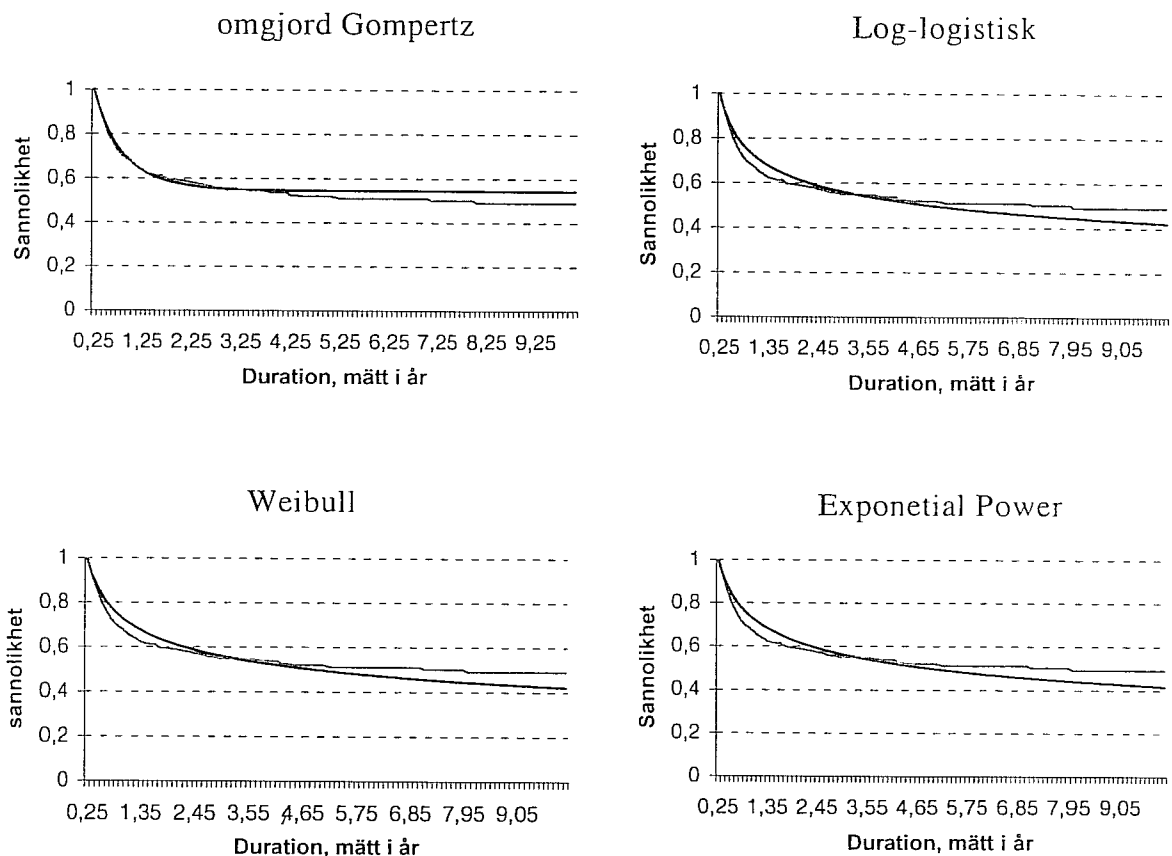
8.3 Val av modell

Dugligheten i de parametriska modellerna jämförs både grafiskt och genom skattade värden på log likelihooden. För att en modell ska anses passa den censurerade

datamängden bör inte den parametriska fördelning generera värden som nämnvärt skiljer sig från den icke-parametriska skattningen. Värdena representeras som tidigare av kvinnor i åldersklassen 50 år med en sjukförsäkring med tre månaderskarens, denna grupp representerar väl hur övriga observerade gruppindelningar förhåller sig till de parametriska modellerna.

I *Diagram 8.1* nedan visas resultaten av likelihood skattningarna för respektive modell tillsammans med motsvarande Nelson-Aalen skattning och *Tabell 8.1* visar maximum likelihood skattningarna.

Diagram 8.1 Skattade parametriska avvecklingsmodeller (mörk linje) i relation med den icke parametriskt skattade avvecklings modellen (ljus linje).



Tabell 8.1, Maximum log likelihood skattningar av de parametriska modellerna.

	Weibull	Log logistisk	Exponential Power	omgjord Gompertz
Log(L)	-443.1	-442.89	-445.34	-168.13

Enligt *Diagram 8.1* är det en modell, baserad på (8.4.3), som mest efterliknar utseendet för den ickeparametriska modellen, dock överskattar den observerade gruppen ovan sannolikheten att tillfriskna för stora värden på t . Övriga modeller har en tendens att underskatta sannolikheten att avvecklas för små värden på t och överskatta för höga värden på t . Vidare indikerar maximum likelihood skattningarna *Tabell 8.1*, att det är den omgjorda Gompertz modellen som bäst speglar den censurerade datamängden. Övriga grupper indelningar visar på liknande resultat som för den undersökta gruppen,

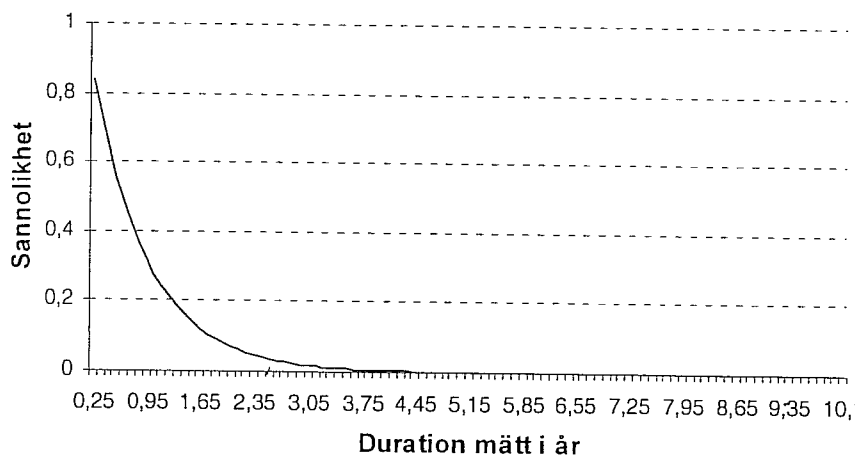
d v s att modell (8.4.3) är den modell som bäst anpassar data. Resultaten gör att studien i fortsättningen kommer koncentrera sig på denna parametriska modell.

Denna studie tyder således på att en avvecklingsfunktion skiljer sig något från en s.k. vanlig överlevnadsfunktion i den meningen att en individ har en positiv sannolikhet att i oändlig tid kvarstå i det undersökta tillståndet, i detta fall sjuk. Detta innebär att sjukligheten är baserad på en stokastisk variabel T , som antar en fördelningsfunktion F , vars sannolikhetsmassa är mindre än ett, d v s $F(\infty) < 1$.

I Bilaga E är den omgjorda Gompertz modellen för övriga grupper uppritad med motsvarande icke-parametrisk skattning, vilket visar på en riktigt god anpassning för samtliga grupper förutom Pb-försäkringar med rörlig karens, vilket avspeglar den oriktiga skattningen avseende den rörliga karensen. Ett problem med denna omgjorda Gompertz modell är, som ovan nämnts, att den saknar väntevärde och således tillhörande osäkerhetsskattning.

I *Diagram 8.2* ritas den analytiska intensiteten att tillfriskna för den valda parametriska modellen (8.4.2). Den uppritade kurvan visar tydligt hur snabbt intensiteten konvergerar mot noll då durationstiden t växer, vilket innebär att efter en viss durationstid t^* sker inga avgångar från tillståndet sjuk. Detta återspeglar konvergensen i avvecklingsfunktionen. Modellen missar således avvecklingar som inträffar efter en lång sjukperiod, då framförallt avvecklingar som sker p.g.a. dödsfall. Den parametriska fördelningens anpassning för små och medelstora värden på sjuktiden t , antas dock vara god.

Diagram 8.2 Intensiteten för att tillfriskna för den omgjorda Gompertz modellen.



För långa durationstider bör det gälla att avveckling sker på andra preferenser än vad som gäller för korta durationstider. Detta skulle kunna motivera en undersökning som delade upp beståndet i korta och långa durationstider och sedan utför separata undersökningar. Intensitetsfunktionen kan då ge ett grovt mått för vid vilken durationstid uppdelningen bör ske. En modell som ska anpassa långa sjuktider bör förmodligen i större grad bygga på intensiteten för att avlida än på intensiteten för att tillfriskna. Denna utveckling kommer dock att utelämnas i denna undersökning.

De parametervärden som fås för den omgjorda Gompertz modellen för respektive delbestånd framgår av *Tabell 8.2*.

Tabell 5.2 De skattade parametervärdena för den omgjorda Gompertz modellen med tillhörande log likelihood skattning.

Förs.	Par / Grupp	α	β	Log(L)	Förs.	Par / Grupp	α	β	Log(L)
Pb	M40,R	0,013	-0,018	-120,0	Sjuk	M40,R	0,608	0,946	-787,6
	K40,R	0,042	0,019	-99,76		K40,R	0,724	0,864	-336,4
	M40,3	0,398	0,326	-799,1		M40,3	0,659	1,019	-799,1
	K40,3	0,350	0,223	-578,5		K40,3	0,907	0,810	-411,8
	M50,R	-0,009	0,031	-572,8		M50,R	0,749	0,646	-1577
	K50,R	0,018	0,035	-264,1		K50,R	0,946	0,666	-421,7
	M50,3	0,422	0,188	-1835		M50,3	0,772	0,658	-1296
	K50,3	0,425	0,204	-1378		K50,3	1,381	0,840	-168,1
	M60,R	-0,008	0,035	-119,6		M60,R	1,268	0,583	-474,6
	K60,R	-0,009	0,029	-193,7		K60,R	0,945	0,354	-107,6
	M60,3	0,611	0,210	-545,2		M60,3	1,395	0,634	-454,2
	K60,3	0,589	0,187	-689,8		K60,3	0,938	0,373	-107

Som nämndes i avsnittet om den omgjorda Gompertz fördelningen, existerar inget väntevärde för denna modell, vilket leder till att en statistisk jämförelse inte går att göra mellan denna och den icke-parametriska skattningen. Detta motiverar en beräkning av sjukreserven med avseende på respektive modell för att på så sätt se hur de två modellerna avviker från varandra.

9 Reserver

Sjukreserven definieras som nuvärdet för en pågående sjukförsäkring till en person som är $(x + t)$ år gammal och som vid $x + k$ års ålder blev berättigad till en månadsvis sjukpension fram till och med åldern z , $z > x + t$. Med en karenstid av k år definieras sjukreserven, för en utbetalning på en krona enligt Carl Gösta Dillner [1974] som:

$$a(x, t, z - x - t) = \int_t^{z-x} \frac{\lambda(x, u)}{\lambda(x, t)} \cdot \exp(-\delta(u-t)) du, \quad t \geq k \quad (9.1)$$

där δ är ränteintensiteten, d v s. $\delta = \log(1 + r)$, r är räntesatsen.

9.1 Utjämning

På det sätt som Nelson-Aalen skattningen är beräknad, fås enbart punktvisa skattningar för sannolikheten att avvecklas från sjukdom. För att beräkna sjukreserven måste således en utjämnande skattning av avvecklingsfunktionen definieras.

Enligt Grunderna, se Bilaga D, kan avvecklingsfunktionen uttryckas med en summa av exponentialfunktioner. Antag därför att den skattade avvecklingsfunktionen kan anpassas med hjälp av följande:

$$\lambda^*(t) = \sum_{i=1}^n a_i \cdot \exp(-b_i t), \quad (9.1.1)$$

där parametrarna a_i och b_i , $i = 1, \dots, n$, är okända. Tanken är att anpassa denna parametriska funktion till Nelson-Aalen skattningen där parametrarna skattas med den modifierade minimum chi-2metoden, d v s genom att minimera följande uttryck med avseende på a_i , b_i enligt:

$$Q^2 = \sum_{i: t_i \leq t} \frac{1}{\text{Var}(\hat{\lambda}(t_i))} (\hat{\lambda}(t_i) - \lambda^*(t_i))^2 \quad (9.1.2)$$

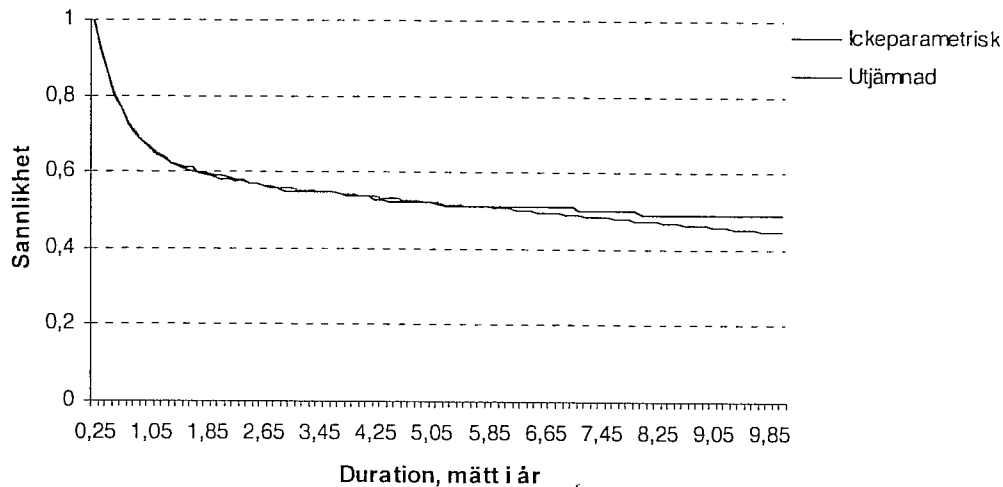
där:

$$\text{Var}(\hat{\lambda}(t_i)) = \left(\hat{\lambda}(t_i) \right)^2 \sum_{j: t_j \leq t_i} \frac{1}{R(t_j)^2}. \quad (9.1.3)$$

Denna metod används i FTN's dödlighetsundersökning [1999] för att utjämna dödsintensiteterna. Skattningar skapade med metod (9.1.2) är asymptotiskt effektiva, d v s alla andra skattningar har asymptotiskt större varians, Hoem [1976,b]. Ett stort problem med detta förfarandesätt är att (9.1.2) minimerar med avseende på de punkter där avveckling sker, vilket gör att skattningen är bra för små värden på durationen t , där ett stort antal avgångar inträffar, däremot kan anpassningen vara något sämre för stora värden på t . För att illustrera problemet visas i *Diagram 9.1*, representerad av samma grupp som

tidigare där sista avvecklingen sker efter en sjuktid på 7,9 år. Övriga grupper redovisas i Bilaga E där den modifierade chi-2 skattningen är uppritad med motsvarande icke-parametrisk- och parametrisk skattning.

Diagram 9.1, den modifierade chi-2 skattningen i relation med den ickeparametriska skattningen.



Beräkningen av sjukreserven avseende den icke-parametriska modellen kommer därför att bygga på ett antagande om att avvecklingsfunktionen är linjärt avtagande mellan två efterföljande avvecklingstidpunkter. Antagandet gäller också mellan åldern $x + t$ år och första avvecklingen och mellan sista avvecklingen och slutåldern z . Som ovan framkommit, kan efter en viss durationstid t , skattningar av $\lambda^*(t)$ saknas, vilket leder till att dessa värden måste uppskattas. Uppskattningen hanteras på samma sätt som i FTN's sjuklighetsundersökning [1993] d v s om det i undersökningen finns en skattning av exempelvis $\lambda^*(10)$ så har detta värde ställts i relation till $\lambda_{GR}(10)$ enligt de tekniska grunderna. Relationen $[\lambda^*(10) / \lambda_{GR}(10)]$ används sedan som uppräkningsstal för att exempelvis erhålla skattningen $\lambda^*(25)$.

Skattningen av sjukreserven (9.1), baserad på Nelson-Aalens ansats, beräknas nu i denna studie enligt:

$$a(x, t, z - x - t) = \frac{1}{\lambda^*(t)} \left[\sum_{t_i \leq t} [(\lambda^*(t_i) + \lambda^*(t_{i+1})) \cdot (t_{i+1} - t_i) \cdot \exp(-\delta(t_{i+1} - t))] + \right. \\ \left. [(\lambda^*(t) + \lambda^*(t_1))(t_1 - t) + (\lambda^*(t_{z-x-t}) + \lambda^*(t_\tau))(t_{z-x-t} - t_\tau)] \right] \quad (9.1.4)$$

där τ är den största observerade tidpunkten för en avveckling. Vidare antas ränteintensiteten δ , i denna studie vara 0 vilket motsvarar en räntesats på 0 %.

Skattning (9.1.4) jämförs sedermera med sjukreserven beräknad avseende den parametriska modellen (8.1.4.1). En osäkerhetsskattning är inlagd i tabellen, denna kommer av att beräkna sjukreserven avseende det 95 procentiga konfidensband som togs fram i Kapitel 6.

I Tabell 9.1 redovisas resultaten avseende män och kvinnor i åldersklasserna 40, 50 och 60 år med 3 månadskarens. Övriga förhållanden mellan parametriska och icke-parametriska modeller avseende sjukreserver redovisas i Bilaga F.

De framräknade värdena på sjukreserven indikerar att den parametriska modellen generellt genererar högre värden på reserven. Detta skulle innebära att sannolikheten att avgå från tillståndet sjuk, baserad på den parametriska modellen, är lägre i förhållande till en skattning baserad på Nelson-Aalen. Vidare gäller att de framräknade reserverna i denna studie är relativt höga, vilket till stor del beror på den låga räntesatsen och att sjukligheten på senare tid har ökat, se Kapitel 6. För övrigt bekräftas de resultat som tidigare har framkommit i undersökningen, i och med att olika indelningar genererar olika värden på reserven, dock, som tidigare nämnts, är inte alla skillnader signifikanta.

Tabell 9.1 sjukreserver för en årlig krona upp för åldersklasserna 40, 50, och 60 för män och kvinnor med sjukförsäkring där karens tiden är 3 månader.

Åldersklass 40,

Duration t	Män		Kvinnor	
	P	IP	P	IP
0,25	7,45	5,89±1,6	12,61	10,32±2,59
2	15,67	13,75±4,2	20,28	17,39±3,37
5	18,97	15,64±5,2	19,82	17,03±3,40

Åldersklass 50,

Duration t	Män		Kvinnor	
	P	IP	P	IP
0,25	7,85	7,29±1,3	9,81	8,10±1,2
2	11,04	11,2±2,25	12,54	11,70±1,72
5	9,85	9,73±2,32	10,00	10,09±1,56

Åldersklass 60,

Duration t	Män		Kvinnor	
	P	IP	P	IP
0,25	3,63	3,28±0,27	3,73	3,55±0,57
2	2,94	2,87±0,25	2,88	2,89±0,51
5	0	0	0	0

10 Slutsatser

Studien visar på signifikanta skillnader på 5 procents nivå mellan långtidssjuk- och premiefrielseförsäkringar avseende t-frekvenser och avvecklingsfunktioner. Vidare visar studien på att olikheter finns mellan bestånden män och kvinnor i intensiteten att insjukna och kvarstå som sjuk/skadad, däremot finns ingen signifikant skillnad i avvecklingsfunktionen. Detta skulle innebära att reserverna för män och kvinnor bör vara snarlika däremot ska premier för kvinnor vara högre än för män. Analysen gällande de två karenstyperna 3 månads- och R-karens är dock i sin helhet bristfällig då ej fullständig information varit tillgänglig, angående längden på den rörliga karensen. För att kunna studera R- karensen behövs information på när en person blev beviljad förtidspension eller när en utbetalning på en sjuk eller Pb- försäkring inleds. Ett sätt är då att låta durationstiden starta den dag en förtidspension påbörjas, avvecklingsfunktionen tar således ej hänsyn till längden på sjukdomen eller skadan utan enbart på hur lång tid en utbetalning pågår. Detta bör ge en mer sann bild på hur avvecklingar för försäkringar med R- karenser ser ut.

De skillnader som framkommit mellan tidigare sjuklighetsundersökningar och denna studie indikerar att upprepade undersökningar är befogade för att fånga upp de förändringar som sker i sjukligheten. Detta gäller både för intensiteten att insjukna/skadas och kvarstå som sjuk/skadad en tid längre än karenstiden k år, t- frekvens, som för sannolikheten att avgå från tillståndet sjuk/skadad, avvecklingsfunktionen. Eftersom skillnaden är sådan att sjukligheten har ökat på senare tid bör dessa två iakttagelser leda till att beräkningsunderlaget avseende premier och reserver ses över.

Den parametriska fördelning som bäst anpassade data var en omgjord Gompertzfördelning. Denna fördelning ägde dock inget existerande väntevärde vilket skulle tyda på att sjukligheten är baserad på en stokastisk variabel T , som antar en fördelningsfunktion F , vars sannolikhetsmassa är mindre än ett, d v s $F(\infty) < 1$. Detta skulle innebära att avvecklingsfunktion skiljer sig något från en s.k. vanlig överlevnadsfunktion i den meningen att en individ har en positiv sannolikhet att i oändlig tid kvarstå i det undersökta tillståndet, i detta fall sjuk/skadad.

Den omgjorda Gompertzfördelningen verkar ha en relativ god anpassning för små och medelstora värden på durationen t , dock missar fördelningen det som händer i svansarna vilket till stor del beror på att de flesta avvecklingarna sker för små värden på t . En forsatt undersökning, som tidigare har nämnts, kan således försöka anpassa en fördelning för sjuktider kortare än en viss tid k , och en annan modell för sjuktider längre än k , för att på så sätt fånga upp egenskaperna för så väl korta som långa durationstider.

APPENDIX A

Beräknade T-frekvenser med duration t >= 6 månader

Ålder	Pb R		Sj R	
	Kvinnor	Män	Kvinnor	Män
27	0,6	0,0	6,8	3,0
32	0,0	1,2	5,7	4,2
37	1,4	0,2	11,0	6,6
42	0,5	0,7	12,0	7,0
47	1,2	1,8	15,1	8,9
52	2,0	2,1	13,5	13,1
57	4,1	4,8	26,8	17,8
62	3,3	3,9	8,8	25,1

Ålder	Pb 13		Sj 13	
	Kvinnor	Män	Kvinnor	Män
27	1,4	1,5	6,9	2,8
32	1,5	2,0	6,6	3,8
37	3,1	2,0	11,6	6,3
42	3,6	2,2	13,3	7,0
47	5,3	4,1	14,9	8,9
52	7,3	5,9	14,4	12,7
57	10,7	9,5	25,7	18,3
62	8,4	11,1	21,9	24,3

Beräknade T-frekvenser med duration t >= 1 år

Ålder	Pb R		Sjuk R	
	Kvinnor	Män	Kvinnor	Män
27	0,8	0,0	3,6	1,6
32	0,0	0,7	3,4	2,6
37	2,0	0,2	7,2	3,5
42	0,3	0,8	9,2	4,5
47	0,7	1,9	10,0	6,5
52	1,7	1,9	9,8	10,0
57	4,1	4,3	26,6	14,5
62	2,1	3,8	8,2	19,9

Ålder	Pb 13		Sj 13	
	Kvinnor	Män	Kvinnor	Män
27	1,3	1,2	3,3	1,6
32	1,4	1,3	3,4	2,5
37	2,6	1,3	6,6	3,2
42	2,5	1,7	9,6	4,7
47	3,6	2,7	10,4	6,6
52	5,7	5,2	11,5	9,4
57	10,5	8,3	24,4	14,3
62	7,8	10,9	22,7	18,2

APPENDIX B

Nelson-Aalen skattningen av avvecklingsfunktionen

Pb- försäkringar med duration $t \geq 3$ m.

Sjukförsäkringar med duration $t \geq 3$ m.

Kvinnor ålder 30		
Duration	13kar	Rkar
0,25	1000	1000
0,5	946	1000
0,75	882	1000
1	826	1000
2	808	1000
3	677	1000
4	567	961
5	485	922
6	419	922
7	405	824

Man ålder 30		
Duration	13kar	Rkar
0,25	1000	1000
0,5	928	1000
0,75	883	1000
1	845	1000
2	655	1000
3	538	948
4	516	948
5	504	921
6	476	921
7	421	921

Kvinnor ålder 30		
Duration	13kar	Rkar
0,25	1000	1000
0,5	667	659
0,75	466	460
1	362	347
2	230	232
3	183	166
4	175	158
5	165	145
6	165	145
7	151	145

Män ålder 30		
Duration	13kar	Rkar
0,25	1000	1000
0,5	640	1000
0,75	512	746
1	435	505
2	295	489
3	212	465
4	179	440
5	165	305
6	150	217
7	130	189

Kvinnor ålder 40		
Duration	13kar	Rkar
0,25	1000	1000
0,5	945	1000
0,75	902	1000
1	842	985
2	825	985
3	721	953
4	657	945
5	614	936
6	584	928
7	574	898

Män ålder 40		
Duration	13kar	Rkar
0,25	1000	1000
0,5	860	1000
0,75	795	977
1	758	969
2	670	961
3	613	954
4	569	946
5	530	938
6	500	938
7	487	938

Kvinnor ålder 40		
Duration	13kar	Rkar
0,25	1000	1000
0,5	820	842
0,75	696	718
1	629	653
2	506	532
3	454	485
4	438	471
5	438	465
6	426	451
7	426	443

Män ålder 40		
Duration	13kar	Rkar
0,25	1000	1000
0,5	770	1000
0,75	634	858
1	513	611
2	361	593
3	298	559
4	268	528
5	259	376
6	251	308
7	240	275

Kvinnor ålder 50		
Duration	13kar	Rkar
0,25	1000	1000
0,5	967	997
0,75	935	986
1	874	969
2	861	959
3	767	923
4	725	904
5	693	885
6	667	864
7	651	837

Män ålder 50		
Duration	13kar	Rkar
0,25	1000	1000
0,5	939	983
0,75	904	977
1	870	966
2	792	957
3	742	901
4	714	871
5	694	857
6	678	843
7	665	838

Kvinnor ålder 50		
Duration	13kar	Rkar
0,25	1000	1000
0,5	823	820
0,75	724	729
1	679	682
2	589	589
3	549	552
4	538	541
5	520	520
6	510	511
7	510	505

Män ålder 50		
Duration	13kar	Rkar
0,25	1000	1000
0,5	815	1000
0,75	739	874
1	655	726
2	545	709
3	502	683
4	486	659
5	472	546
6	462	500
7	452	484

Kvinnor ålder 60		
Duration	13kar	Rkar
0,25	1000	1000
0,5	977	992
0,75	938	981
1	901	977
2	897	969
3	813	952
4	776	952
5	754	934
6	735	920
7	735	920

Män ålder 60		
Duration	13kar	Rkar
0,25	1000	1000
0,5	938	990
0,75	905	974
1	873	964
2	794	957
3	767	917
4	745	893
5	727	874
6	711	863
7	711	851

Kvinnor ålder 60		
Duration	13kar	Rkar
0,25	1000	1000
0,5	895	889
0,75	840	848
1	810	826
2	746	757
3	722	734
4	668	688
5	668	688
6	668	688
7	668	688

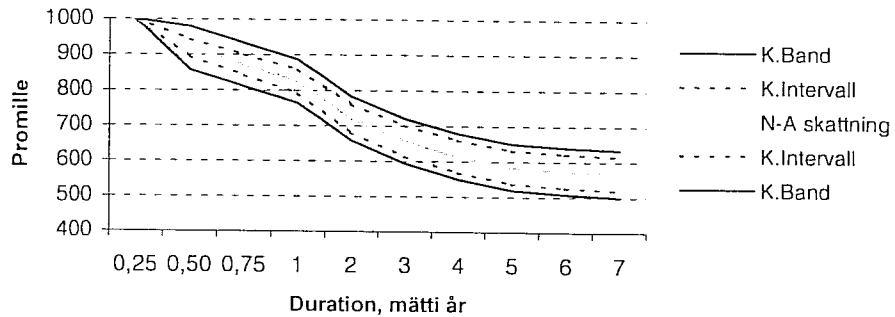
Män ålder 60		
Duration	13kar	Rkar
0,25	1000	1000
0,5	858	1000
0,75	784	922
1	734	770
2	674	763
3	659	753
4	647	741
5	636	676
6	636	660
7	636	651

APPENDIX C (a)

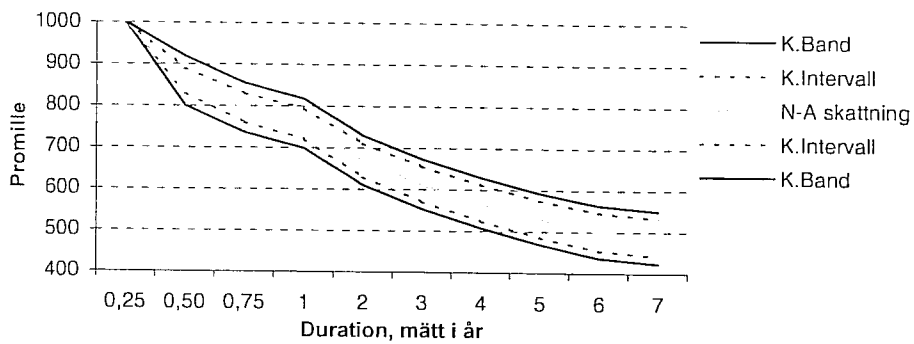
Nelson-Aalen skattningen med tillhörande konfidensband/intervall

Premiebefrielseförsäkringar för ålder 35-44 år

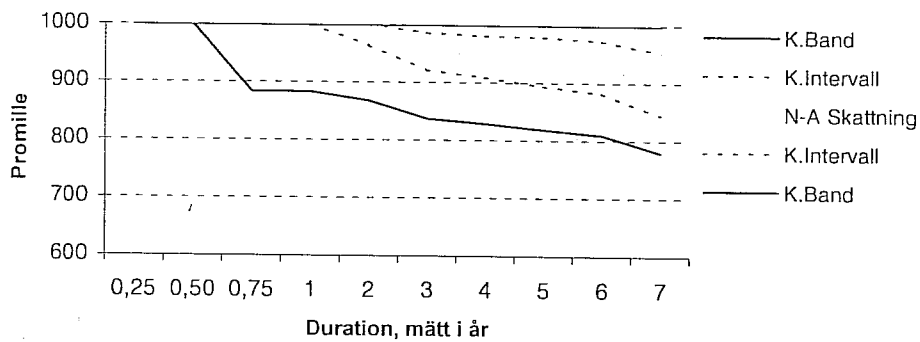
Kvinnor med 3 månaders karens



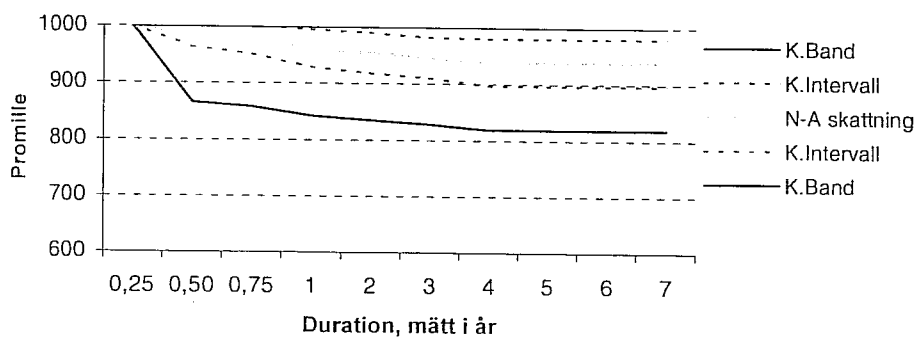
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

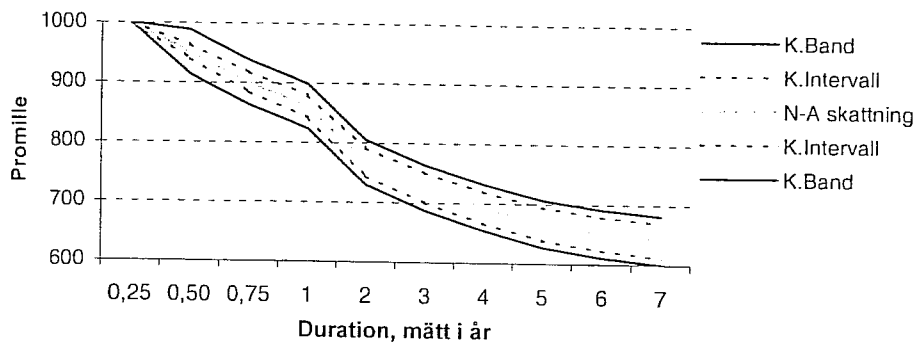


APPENDIX C (b)

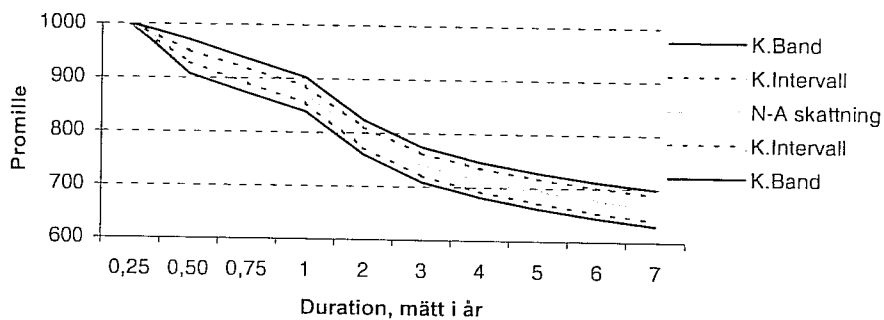
Nelson-Aalen skattningen med tillhörande konfidensband/intervall

Premiefrielseförsäkringar för ålder 45-54 år

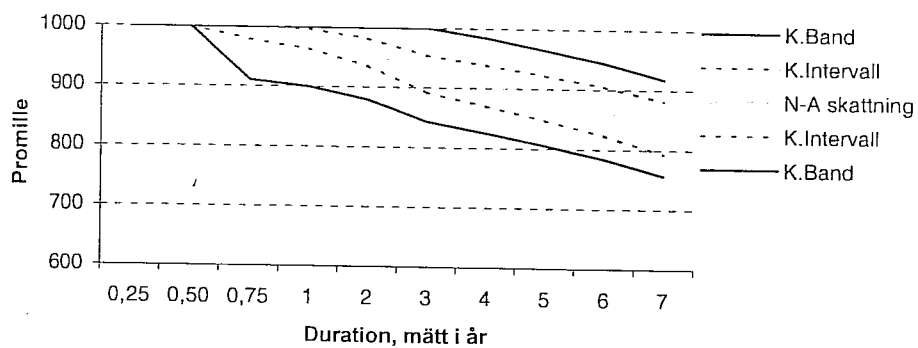
Kvinnor med 3 månaders karens



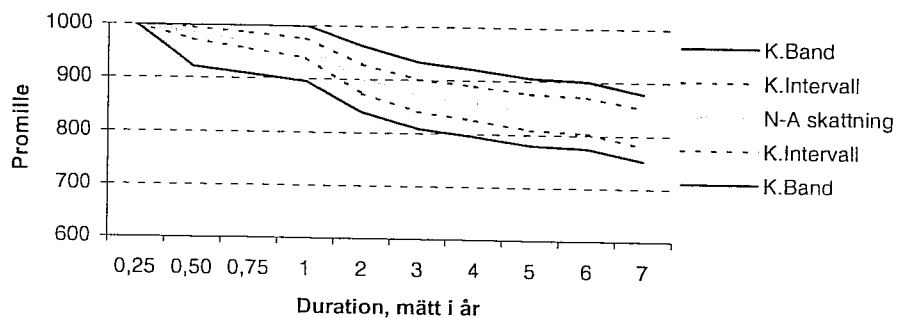
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

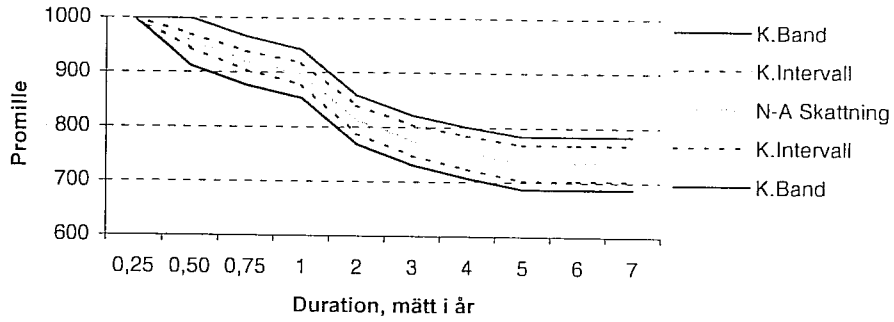


APPENDIX C (c)

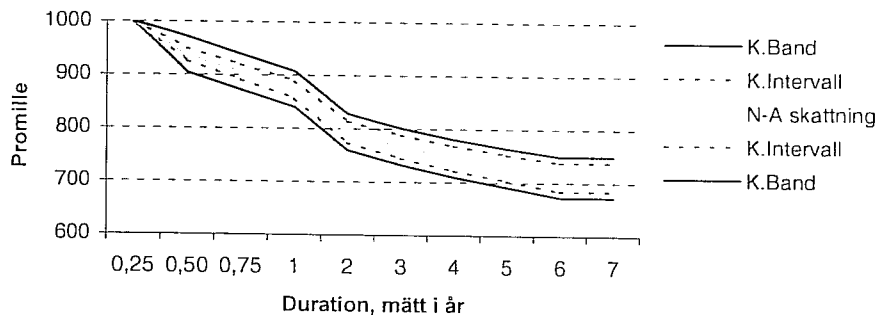
Nelson-Aalen skattningen med tillhörande konfidensband/intervall

Premiebefrielseförsäkringar för ålder 55-64 år

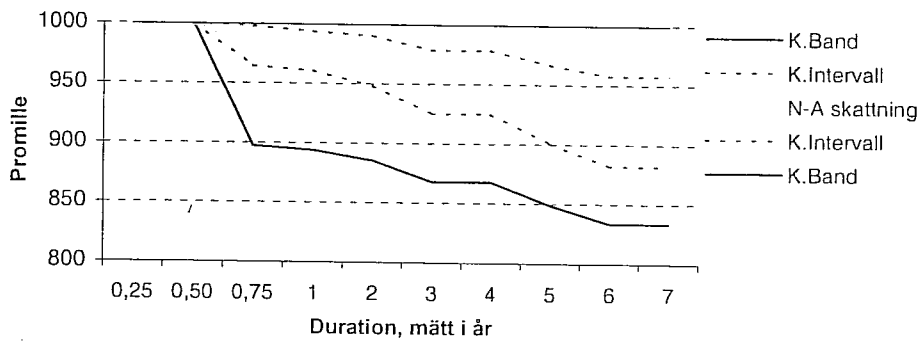
Kvinnor med 3 månaders karens



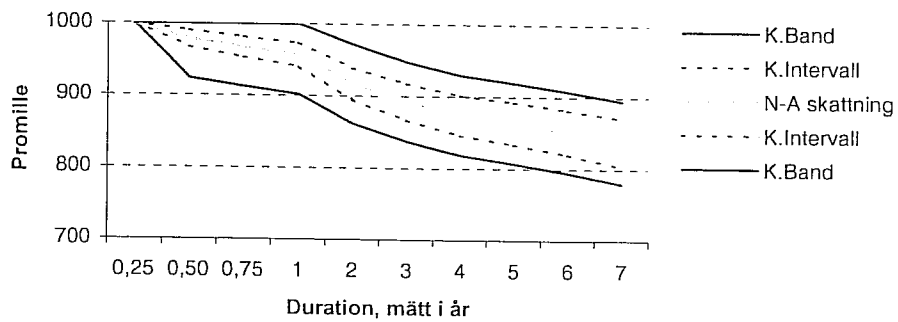
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

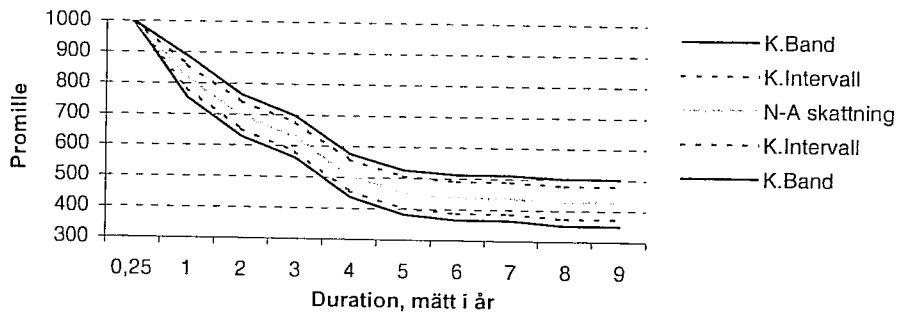


APPENDIX C (d)

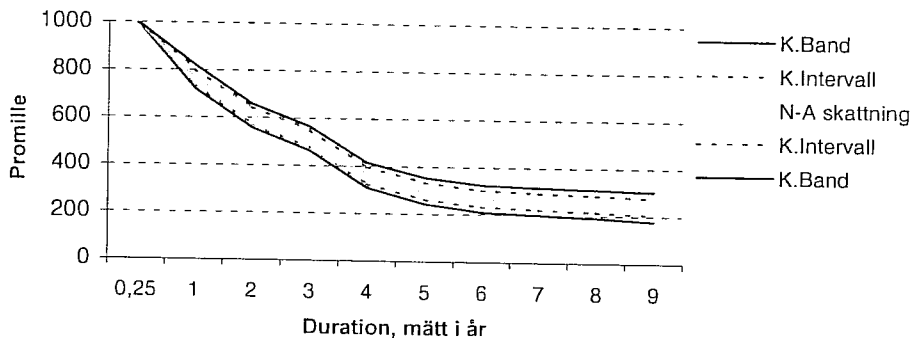
Nelson-Aalen skattningen med tillhörande konfidensband/intervall

Långtidssjukförsäkring för ålder 45-54 år

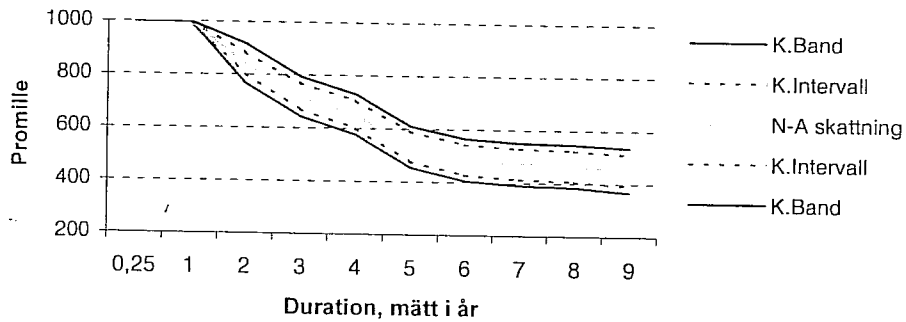
Kvinnor med 3 månaders karens



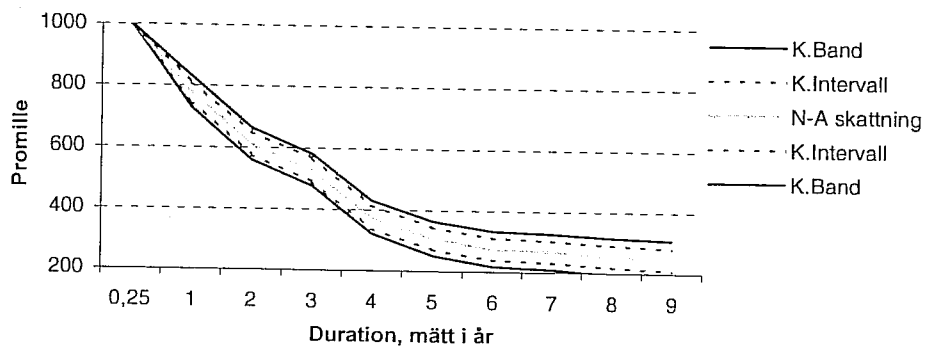
Män med 3 månaders karens



Kvinnor med R- karens



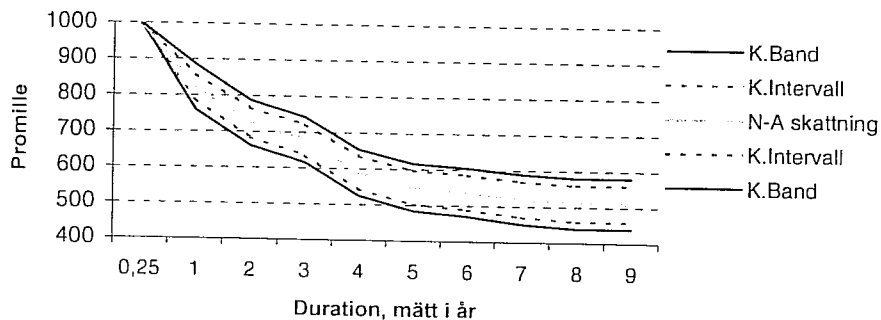
Män med R- karens



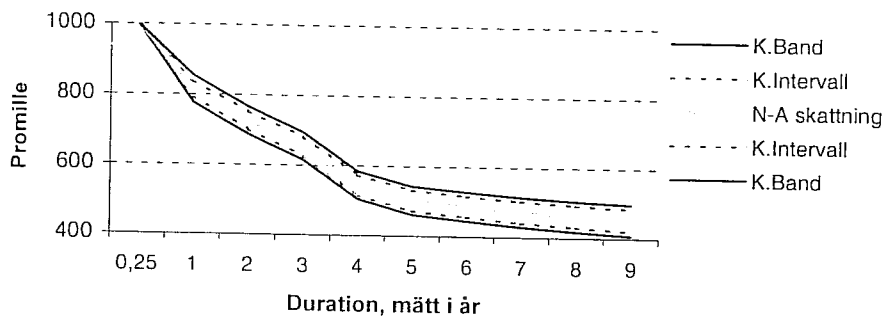
Nelson-Aalen skattningen med tillhörande konfidensband/intervall

Långtidssjukförsäkring för ålder 45-54 år

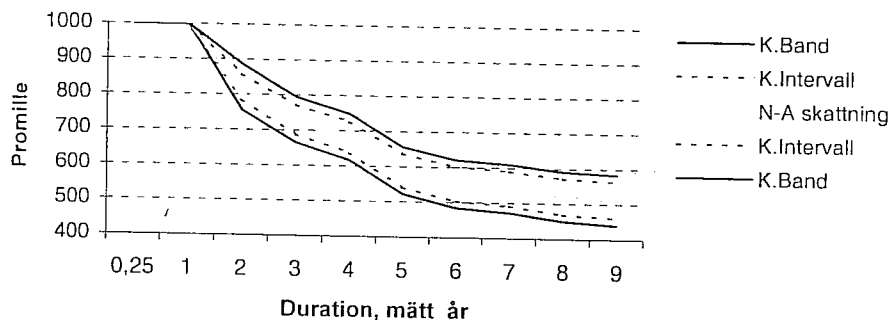
Kvinnor med 3 månaders karens



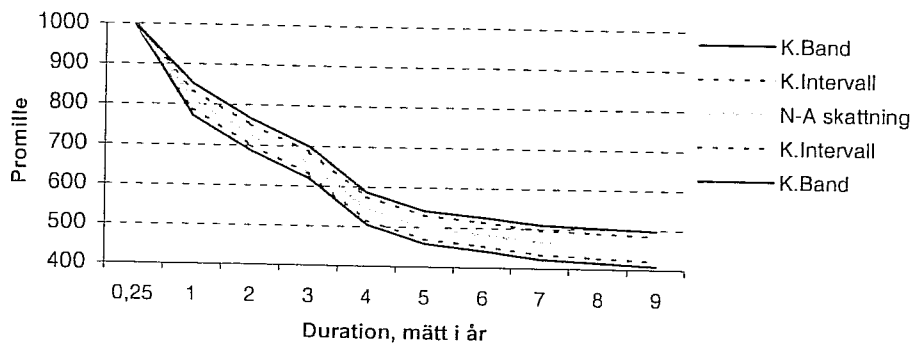
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

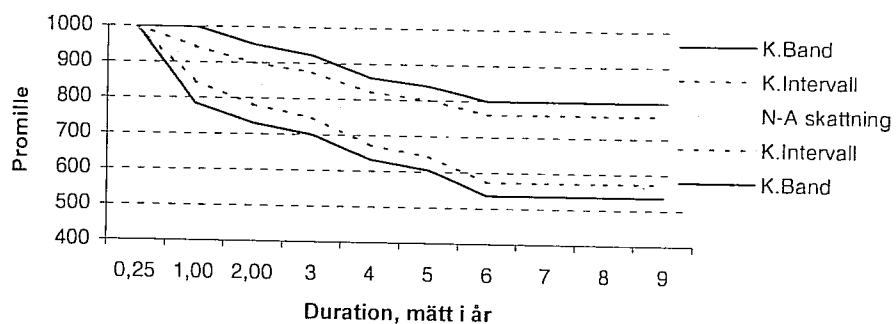


APPENDIX C (f)

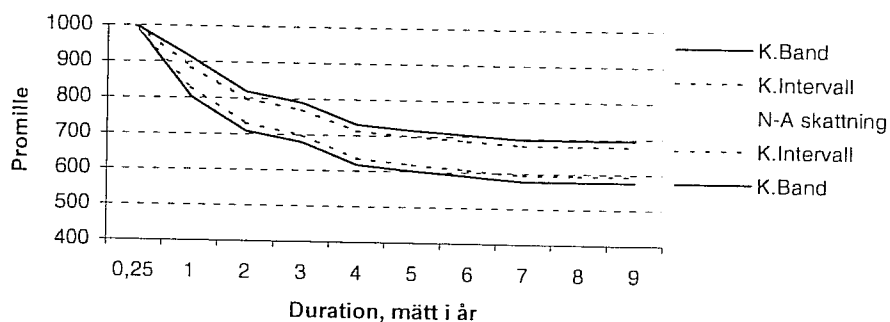
Nelson-Aalen skattningen med tillhörande konfidensband/intervall

Långtidssjukförsäkring för ålder 55-64 år

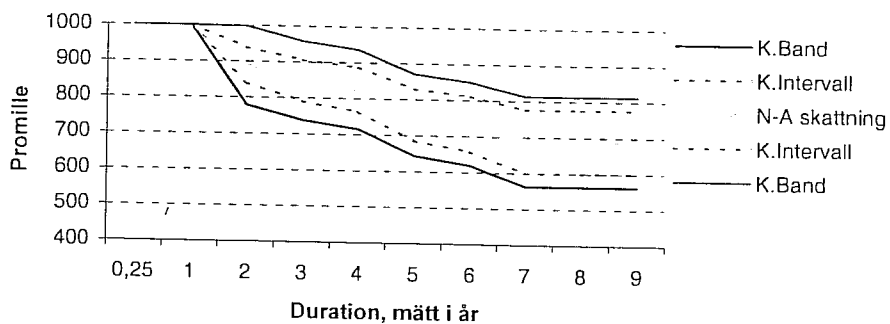
Kvinnor med 3 månaders karens



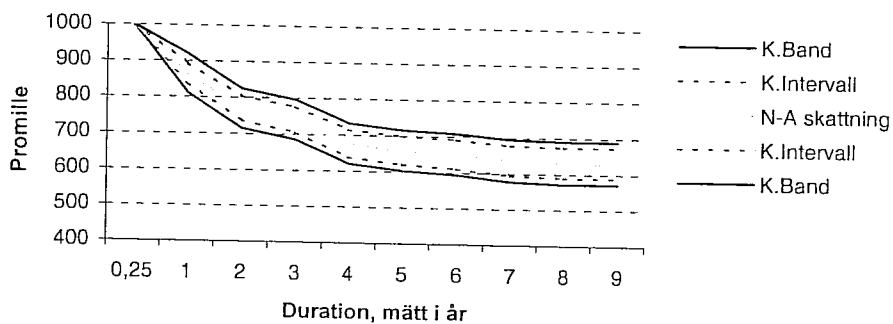
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

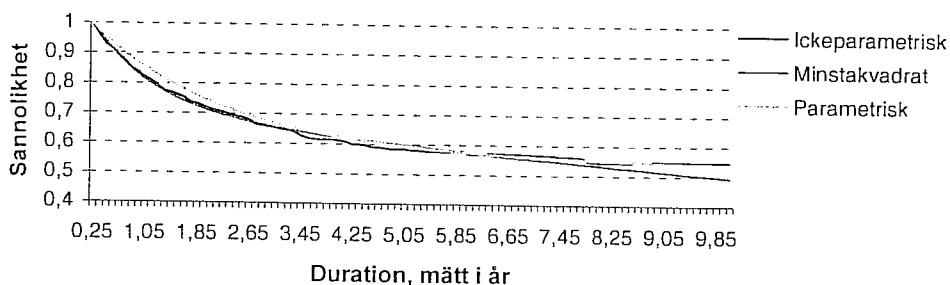


APPENDIX E (a)

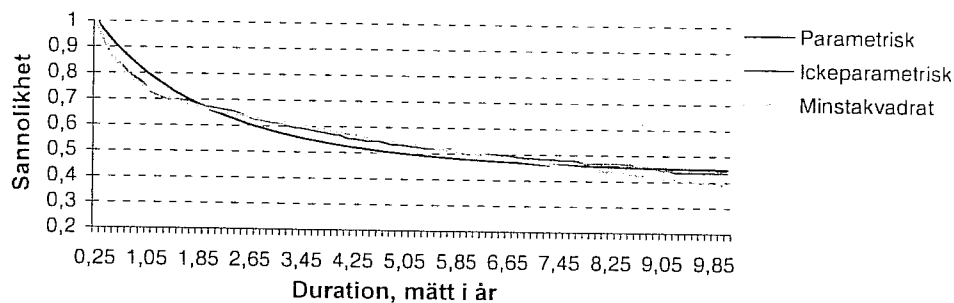
Förhållandet mellan avvecklingsfunktionen skattad med parametrisk, ickeparametrisk och minstakvadrat anpassad Nelson-Aalen skattning.

Premiefrielseförsäkringar för ålder 35-44 år

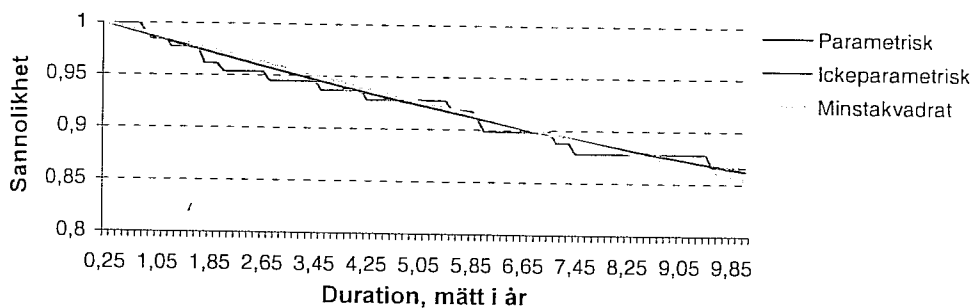
Kvinnor med 3 månaders karens



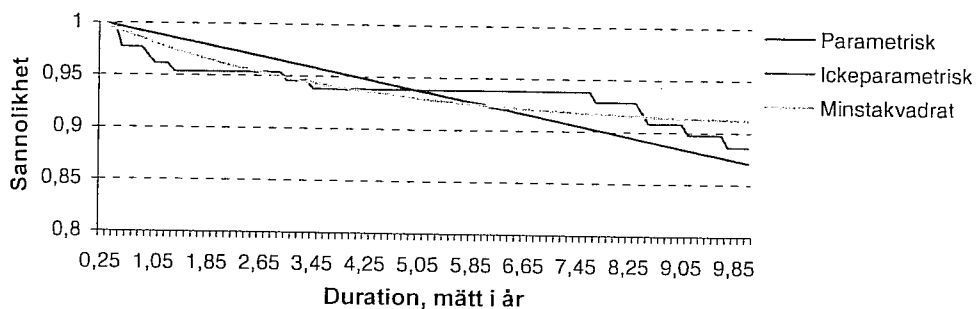
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

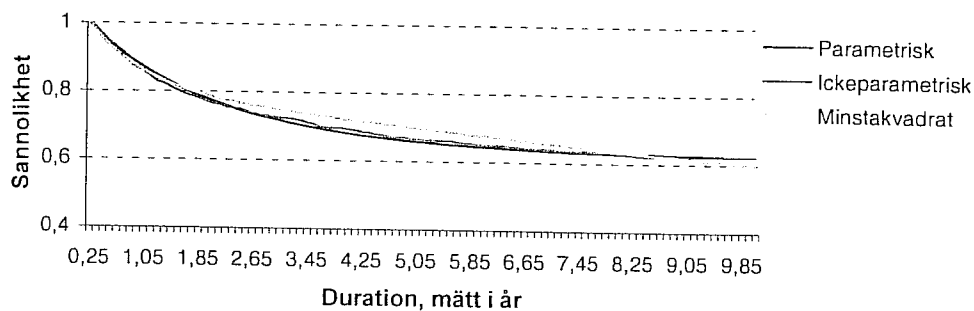


APPENDIX E (b)

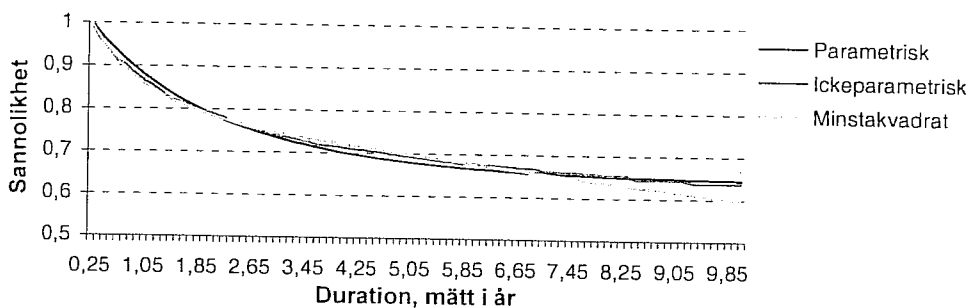
Förhållandet mellan avvecklingsfunktionen skattad med parametrisk, ickeparametrisk och minstakvadrat anpassad Nelson-Aalen skattning.

Premiebefrielseförsäkringar för ålder 45-54 år

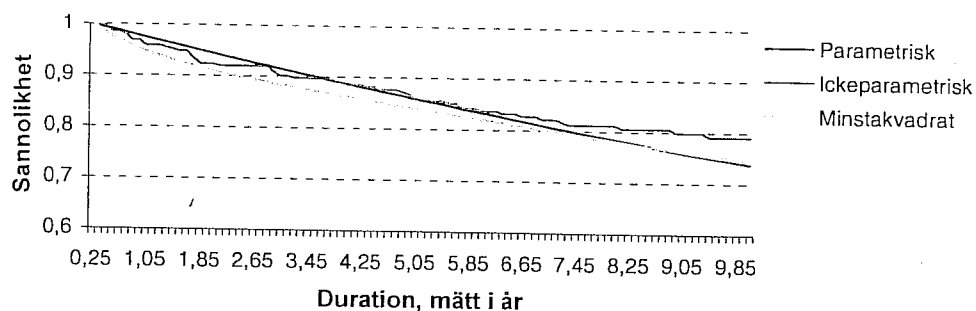
Kvinnor med 3 månaders karens



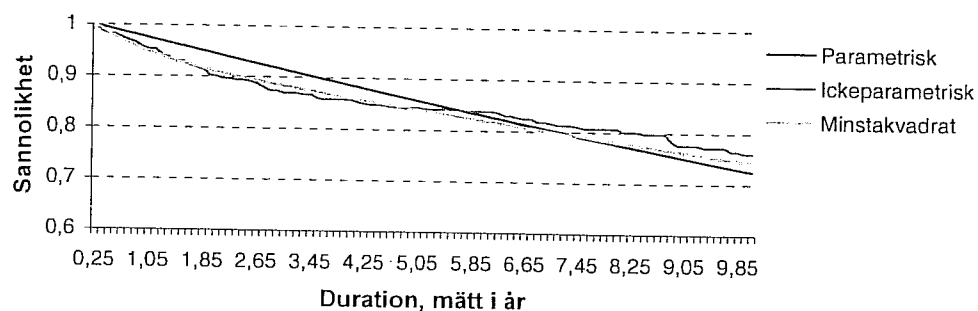
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

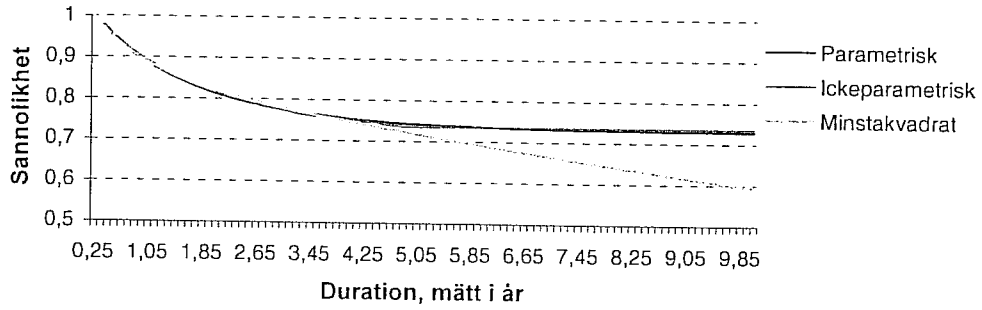


APPENDIX E (c)

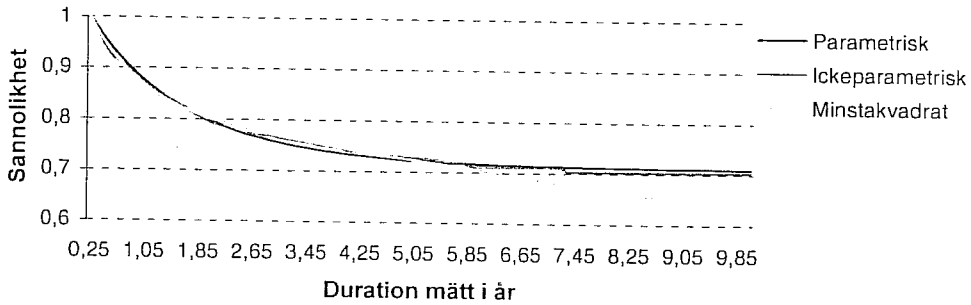
Förhållandet mellan avvecklingsfunktionen skattad med parametrisk, ickeparametrisk och minstakvadrat anpassad Nelson-Aalen skattning.

Premiebefrielseförsäkringar för ålder 55-64 år

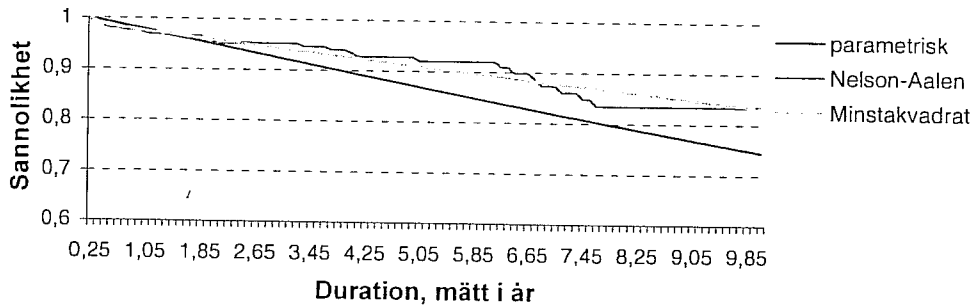
Kvinnor med 3 månaders karens



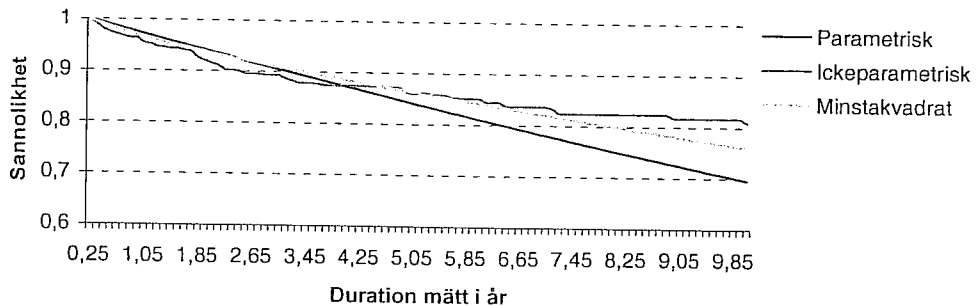
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

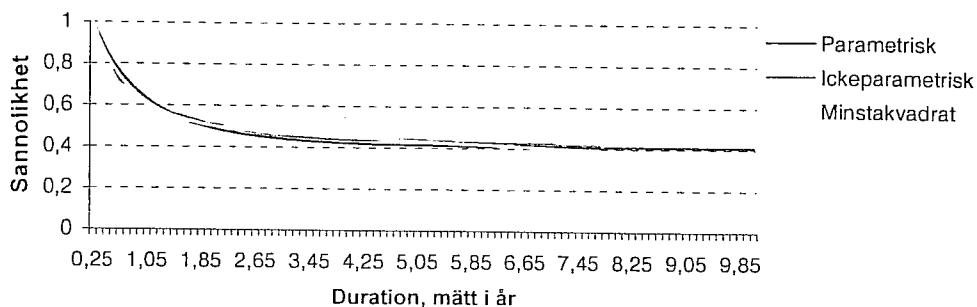


APPENDIX E (d)

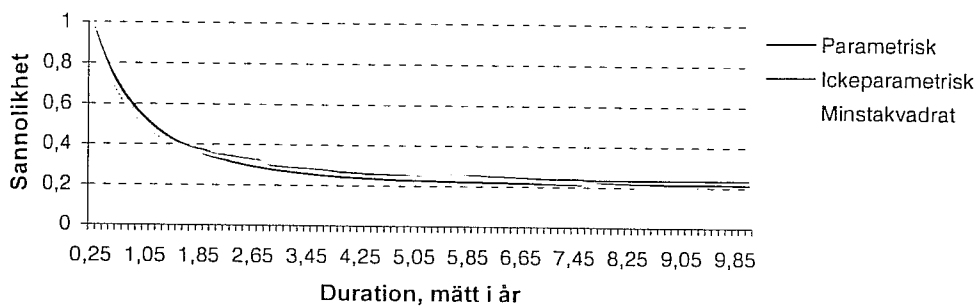
Förhållandet mellan avvecklingsfunktionen skattad med parametrisk, ickeparametrisk och minstakvadrat anpassad Nelson-Aalen skattning.

Långtidssjukförsäkringar åldersklass 35-44 åringar

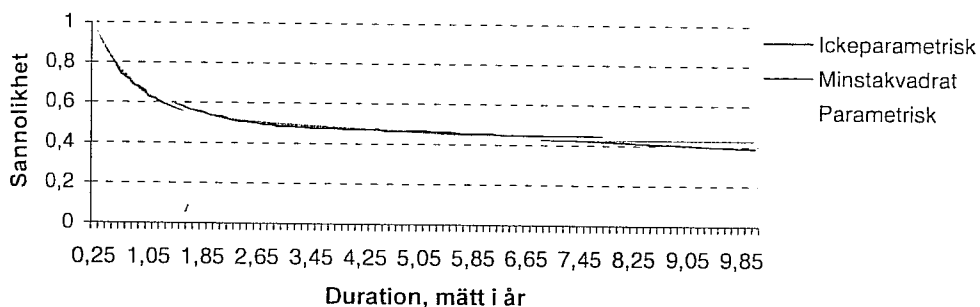
Kvinnor med 3 månaders karens



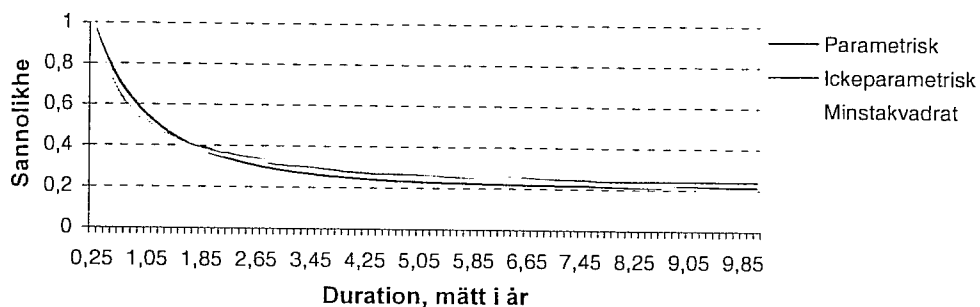
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

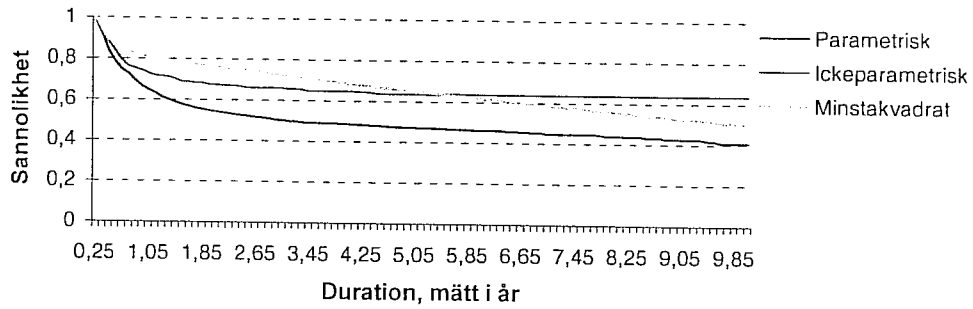


APPENDIX E (e)

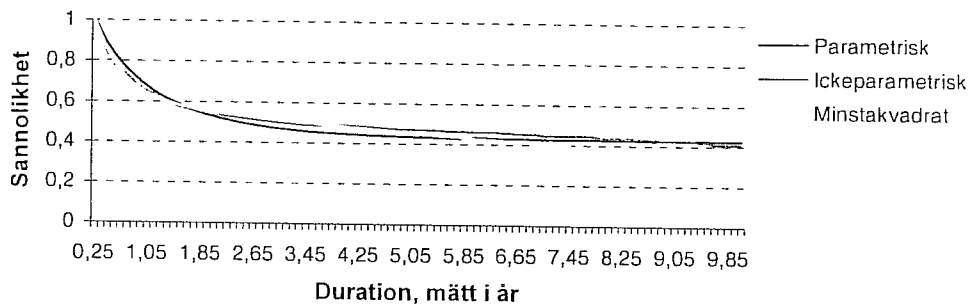
Förhållandet mellan avvecklingsfunktionen skattad med parametrisk, ickeparametrisk och minstakvadrat anpassad Nelson-Aalen skattning.

Långtidssjukförsäkringar för ålder 45-54 år

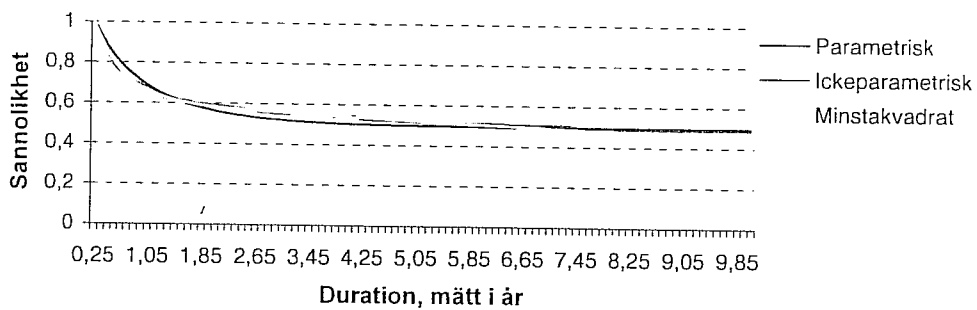
Kvinnor med 3 månaders karens



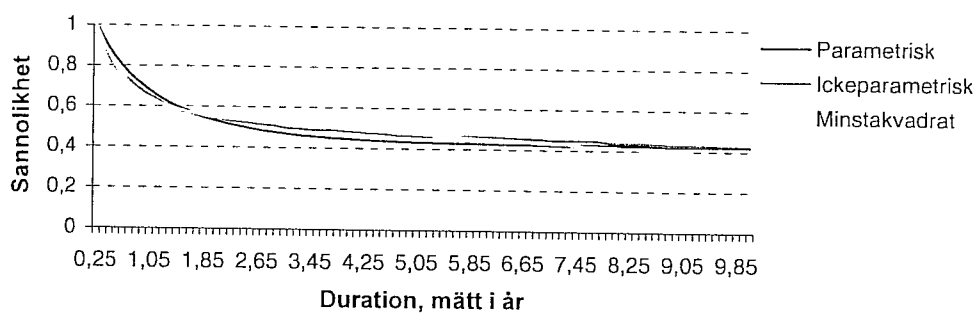
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens

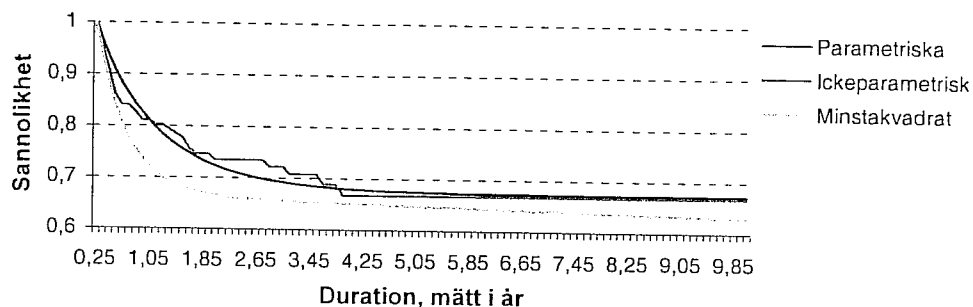


APPENDIX E (f)

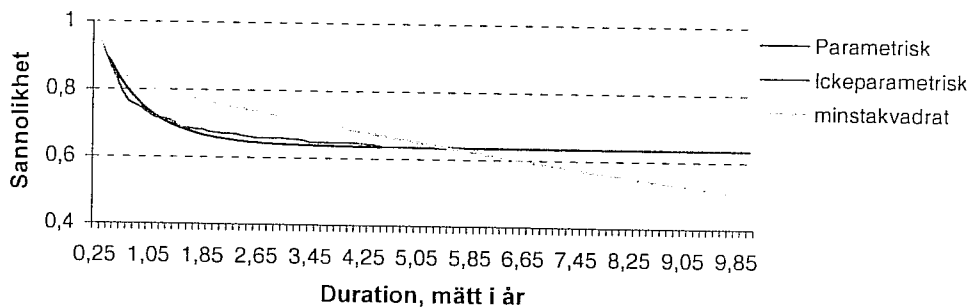
Förhållandet mellan avvecklingsfunktionen skattad med parametrisk, ickeparametrisk och minstakvadrat anpassad Nelson-Aalen skattning.

Långtidssjukförsäkringar åldersklass 55-64 åringar

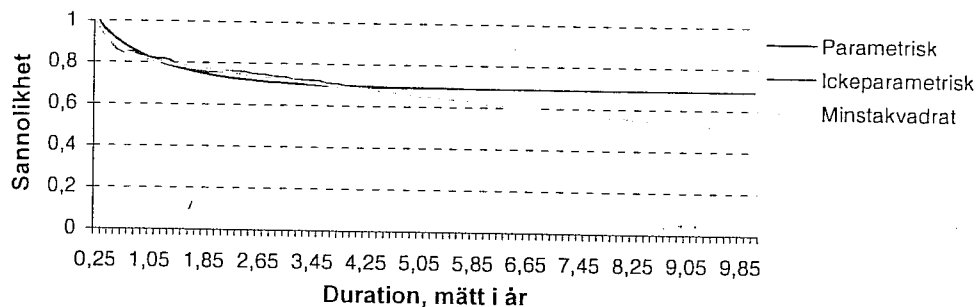
Kvinnor med 3 månaders karens



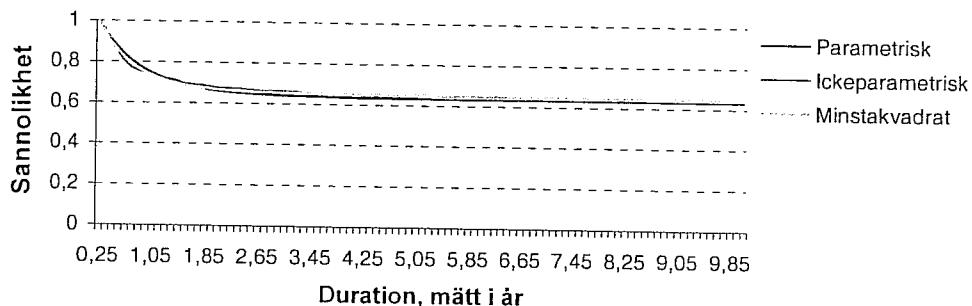
Män med 3 månaders karens



Kvinnor med R- karens



Män med R- karens



Referenser

- Aalen, O. (1978): *Nonparametric Inference for a Family of Counting Processes*; Annals of Statistics, volume 6 Issue 4, sid. 701-736.
- Ajne, B., Ohlin, J. (1994): *Livförsäkringsmatematik*, Institutet för försäkringsmatematik och Matematisk Statistisk, Stockholms Universitet, sid. 5-43.
- Andersen, P. K., Borgan, Ø.(1985): *Counting Process Models for Life History data*, Scandinavian J. Statistics 12, sid. 97-158.
- Andersson, H., Boulem, D. (1997): *An Introduction to Martingale Theory, Institute of actuarial mathematics and mathematical statistics*, Stockholm University, sid. 50,51.
- Björk, T. (1998): *Arbitrage Theory in Continuous Time*, Oxford University press, sid. 27-29.
- Dillner, C.-G. (1969): *New Bases for Non-cancellable Sickness Insurance in Sweden*, Scandinavian Actuarial journal, sid. 113-124.
- Dillner, C.-G. (1974): *New Bases for Long Term Sickness in Sweden 1973 Insurance*, Scandinavian Actuarial journal, sid. 167-173.
- Gill, R. D. (1994): *Understanding Cox's model: A Martingale Approach*, Journal of the American Statistical Association, volume 79 Issue 386, sid. 441-447.
- Klein, P. K. Moeschberger, M. L. (1997): *Survival Analysis*, Springer.
- Källström, L. (1990): *Long-Term Individual Sickness insurance according to the 1984 rules (G84)*, Sverige Reinsurance Company 75 year, Fyris tryck AB, sid. 31-60.
- Lewander S. (2001): *Livförsäkringsformer*, ifu, sid. 23-26.
- Mattson, P. (1977): *Sjukförsäkringen i Sverige- historisk, teknik och utveckling*, Gotab, sid. 18-65.
- Ramlau-Hansen, H. (1982): *Smoothing Counting Process Intensities by Means of Kernel Functions*, University of Copenhagen
- Sandström, A. (1990): *Long-Term Health Insurance*, Sverige Reinsurance Company 75 year, Fyris tryck AB, sid. 101-140.
- Steal, J. M. (2001): *Stochastic Calculus and Financial Applications*, Springer, sid. 137-143.
- Øksendal, B. (2000) *Stochastic differential Equations*, Springer, sid 7, 11-14, 31.
- Österberg, H. (1990): *Sjuklighetsavveckling – En empirisk undersökning bland individuella sjukförsäkrade*, kandidat uppsats, Uppsala Universitet.