Matematisk statistik
Stockholms universitet

Prissättning av sjukdomsförsäkringen Critical Illness

Jessica Rundqvist

Examensarbete 2011:4

Postadress:

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm
Sverige

Internet:

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

Matematisk statistik
Stockholms universitet
Examensarbete **2011:4**,
http://www.math.su.se/matstat

Prissättning av sjukdomsförsäkringen Critical Illness

Jessica Rundqvist*

Oktober 2011

Sammanfattning

Detta arbete är ett försök till att hitta en prissättningsmodell för en sjukdomsförsäkring med begränsad skadedata. Tanken är att se om det med hjälp av kredibilitetsskattning går att kombinera nationell data med portföljdata för att få en bra modell. Tyvärr visar det sig att detta inte är lämpligt då portföljen har ett tydligt moturval. Sannolikheten att drabbas av skada i den försäkrade populationen är betydligt högre än motsvarande sannolikhet i den nationella populationen. Således kvarstår problemet. Produkten är för lågt prissatt och har lönsamhetsproblem . Därför föreslås en alternativ prissättningsmodell. Den alternativa modellen är en funktion av det nationella datat anpassad till portföljnivå.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: jessica.rundqvist@home.se. Handledare: Erland Ekheden.

Abstract

This work is an attempt to find a pricing model for a critical illness insurance portfolio with limited injury data. The idea is to use credibility theory to combine estimates from the insured portfolio with estimates from the entire population to achieve a good and stable model. Unfortunately it proved not to be a suitable model since the portfolio suffers from adverse selection. The probability of being diagnosed with a critical illness in the insured population is significantly higher than the corresponding probability in the overall population. However, the problem remains; the critical illness insurance portfolio is too low priced and has profitability problems. Therefore an alternative pricing model is proposed. The alternative model is a function of the national data fitting to the portfolio level.

Förord

Detta arbete utgör ett exarmensarbete om 30 högskolepoäng på matematiska institutionen vid Stockholms Universitet.

Ett stort tack till min handledare Erland Ekheden som har läst och kommenterat mitt arbete.

Innehåll

1 Introduktion 5
2 Problemformulering och lösningsförslag 5
2.1 Bakgrund 5
2.2 Problemformulering 6
2.3 Metodik 6
2.4 Metod om skadedata finns 7
2.5 Metod om det inte finns någon skadedata 8
2.6 Begränsningar 9
3 Estimering av sannolikheten för critical illness eller dödsfall 10
3.1 Skattning av sannolikheten för skada 10
4 Estimera sannolikheten för CI och dödsfall i Storbritannien 13
4.1 Sannolikheten för skada i Storbritannien baserat på skadedata. 13
4.2 Sannolikheten för skada i Storbritannien baserat nationell data 16
4.3 Analys av skillnaden i skadedata och nationsdata 18
5 Kredibilitetsteori 20
5.1 Introduktion 20
5.2 Definition och notation 20
5.3 Bühlmanns kredibilitetsmodell 21
6 Tillämpa kredibilitetsteori på UK skattningarna av $\hat{\varphi}^{acc}$ 23
6.1 Kredibilitetsskattning av $\hat{\varphi}_I^{acc}$ och $\hat{\varphi}_N^{acc}$ 23
7 Prissätta Sverige 25
7.1 Sannolikheten för skada i Sverige 26
7.1.1 Förbättring av den nationella skattningen 27
8 Diskussion 30
8.1 Metoddiskussion 30
8.2 Jämförelse med befintlig rate 30
8.3 Förslag till alternativ metod 31
8.3.1 Män 32
8.3.2 Kvinnor 33
9 Slutsats 34
10 Referenslista 36
11 Appendix 37

1 Introduktion

Den befintliga Critical Illness portföljen är för lågt prissatt och har lönsamhetsproblem. Det finns en önskan om en ny premiemodell som kan hantera både när skadedata finns om än i begränsad omfattning och när skadedata inte finns. Vi ska i detta arbete ta fram och utvärdera en modell för att prissätta sjukdomsförsäkringen Critical Illness i Storbritannien och i Sverige. Försäkringen Critical Illness är en försäkring som utfaller när försäkringstagaren diagnostiseras med en av de sjukdomar som definierats i försäkringsvillkoret. Eftersom omfattningen av försäkringsdatat är mycket begränsat i Storbritannien och inte existerar alls i Sverige är tanken att utreda om offentlig nationell statstik för respektive land kan vara till hjälp. Det finns en misstanke om att sannolikheten att drabbas av sjukdom i den försäkrade populationen är högre än motsvarande sannolikhet i den nationella populationen. I Storbritannien där skadedata finns men omfattningen är för litet för att få bra estimat tänker vi prova att använda skattningar från både skadedata och hela nationen. Metoden vi tänker tillämpa för att kombinera ihop de estimaten från skadedatat och nationen är kredibilitetsteori. I Sverige där det bara finns nationell data att tillgå är tanken att försöka justera nationsdatat med hänsyn till den eventuella skillnad som finns mellan försäkrad och övrig population i Storbritannien.

2 Problemformulering och lösningsförslag

2.1 Bakgrund

Under 1980-talet breddades livförsäkringsmarknaden från att erbjuda traditionella livförsäkringsprodukter som utfaller när försäkringstagaren avlider till att även erbjuda andra försäkringar. Det utvecklades olika typer av försäkringsskydd där försäkringstagaren kunde få pengar utbetalat under tiden han fortfarande var i livet. En av de produkter som utvecklades går under samlingsnamnet Dread disease och är ett försäkringsskydd som utfaller vid diagnos av ett antal specificerade sjukdomar. Produkten började lanseras i Sydafrika under tidigt 80-tal där den blev mycket framgångsrik. Utvecklingen kom dock från USA där en variant av försäkringen hade existerat en längre tid. Den tidigare varianten av försäkringen från USA täckte enbart diagnos av cancer och var avsvedd för att täcka medicinska kostnader som uppkom i samband med sjukdomen.

I det här arbetet ska vi titta närmre på en Dread Disease produkt som Försäkringsbolaget börjande lansera i Storbritannien i början av 90-talet och som de gav namnet Critical Illness. Villkoren för Critical Illness innebär att försäkringstagaren får en engångssumma utbetalt i händelse av att han diagnostiserats med någon av sjukdomarna specificerade i avtalet. Critical Illness säljs både som ett tillägg till en traditionell livförsäkring och som en ensam försäkring. I båda fallen utbetalas försäkringssumman till den försäkrade om han diagnostiserats med någon av de specificerade sjukdomarna och överlever en karenstid. Om försäkringstagaren däremot diagnostiserats med en critical illness i samband med dödsfall och har tecknat en försäkring utan livförsäkring (sk Additional benefit) så utbetalas ingen ersättning. Om försäkringen istället tecknats tillsamman med en livförsäkring (sk Accelerated benefit) så betalas försäkringen ut även om diagnosen av critical illness görs i samband med dödsfall, dvs. en för tidig död ingår i skyddet.

Den här typen av produkter har mottagits mycket väl på försäkringsmarknaden även i länder där statlig sjukvårdsförsäkring täcker kostnaderna för vård. Den mycket snabba utvecklingen i medicinsk vetenskap är en bidragande orsak. Majoriteten av de individer som insjuknar i allvarliga sjukdomar som cancer, hjärtinfarkt och stroke överlever men flertalet behöver göra stora livsstilförändringar. Den här försäkringen kan bidra till att göra dessa omställningar ekonomiskt möjliga.

2.2 Problemformulering

Uppgiften i detta arbete är att ta fram och utvärdera en prissättningsmodell för produkten Critical Illness både i Storbritannien och i Sverige samt att försöka hantera några av de utmaningar som Försäkringsbolaget har brottats med. I Storbritannien har försäkringen sålts en tid medan försäljningen i Sverige är mycket blygsam. Även om produkten har sålts några år är det inte säkert att omfattningen av skadedatat är tillräcklig stort för ett tillförlitligt estimat. I detta fall har vi en begränsad mängd skadedata för Storbritannien men ingen skadedata alls i Sverige. Därför behövs två modeller för prissättningen. En modell som tar med skadedata i beräkningsmodellen vilket borde ge ett mervärde till estimatet av priset och en modell som kan hantera fall där det inte finns tillgång till skadedata.

Följande har identifierats som centrala frågor:

Utmaning 1: Data
Försäkringsbolaget har sålt försäkringen Critical Illness i Storbritannien sedan tidigt 90-tal. Antalet skador är fortfarande för litet för att skadedatat ska kunna ligga till grund för ett tillförlitligt estimat. I Sverige säljs Critical Illness men i en mycket begränsad omfattning. Det finns därför en önskan om en modell som kan hantera både när skadedata finns om än i begränsad omfattning och när skadedata inte finns.

Utmaning 2: Urval
Försäkringsbolagets erfarenhet är att de individer som köper eller har köpt denna försäkring inte har samma risk för skada som populationen i övrigt vilket måste tas i beaktande. Den försäkrade populationen skulle då inte vara ett slumpmässigt urval ur övriga befolkningen. För att få teckna denna försäkring måste man vara fullt frisk och inte ha någon belastande medicinsk historia. Dessutom har de personer som köper denna typ av försäkring råd och möjlighet att ta hand om sin hälsa. Dessa två omständigheter skulle kunna innebära att de försäkrade individerna löper en lägre risk av att drabbas av sjukdom än befolkningen i övrigt. Erfarenheten är dock den motsatta. Den försäkrade populationen har en högre sannolikhet för att insjukna i svåra sjukdomar än befolkningen i övrigt.

2.3 Metodik

Riskpriset för en Critical Illness försäkring är lika med summan av alla sannolikheter att diagnostiseras med någon av sjukdomarna i avtalet multiplicerat med försäkringssumman.

Försäkringssumman är fastställd från början varför det endast är sannolikheten för skada som vi behöver estimera för att kunna beräkna priset. Detta skulle vara relativt enkelt om data fanns tillgängligt.

Bristen på data gör att det inte är fullt så enkelt och att det krävs en del antaganden. I Storbritannien där försäkringen har sålts i flera år finns skadedata men omfattningen är begränsad med för få observationer för att enbart kunna använda dem för att skatta riskpriset. I Sverige där försäkringen har sålts i ytterst begränsad omfattning finns ingen skadedata att tillgå. Ett naturligt alternativ är att använda nationell data och göra antagandet att hela befolkningen ingår i försäkringspopulationen. Problemet med detta är att det finns tecken på att risken att diagnostiseras med en Critical Illness är högre i den försäkrade populationen jämfört med den nationella populationen. I vårt fall har vi en skattning av den sökta sannolikheten baserad på försäkrad population i Storbritannien, dock är skattningen osäker då det finns få observationer. Å andra sidan har vi ytterligare en skattning som är baserad på en stor datamäng och som ger ett stabilt estimat men där estimatet troligen inte helt motsvarar den sökta sannolikheten.

För att det ska vara möjligt att prissätta ett enskilt försäkringsavtal enbart baserat på analys av den enskilda försäkringens skadehistorik, måste försäkringen ha en mycket hög skadefrekvens. I de allra flesta fall är det omöjligt och därför delar försäkringsbolaget in alla risker i olika riskportföljer där varje portfölj innehåller likartade risker. Målsättningen är att varje portfölj ska vara lönsam dvs. försäkringsbolaget ska få in mer premier än vad som betalas ut i skadekostnader och övriga kostnader. All premiemodellering görs på portföljnivå eftersom den innehåller många försäkringar med liknande risk och mängden skadedata mer omfattande. Större riskportföljer har också en bredare premiebas och är mindre känslig för enskilda skador. När man pratar om prissättning av en försäkringsportfölj ska först och främst hela portföljen vara rätt prissatt. Sedan vill ju försäkringsbolagen självklart locka till sig bra kunder, dvs. de som löper mindre risk att få skador och eliminera de kunder som har större risk att få en skada. Detta försöker man göra genom att styra priset. För att behålla bra risker kan det löna sig att sänka priset för de försäkringstagare med lägre risk än genomsnittet och höja priser för de med högre risk. Det är därför vanligt att vid prissättning både arbeta med risken för portföljen och risken för den enskilda kunden. En metod som ofta används är kredibilitetsteori där man försöker hitta den optimala kombinationen mellan risken för hela portföljen och risken för det individuella avtalet.

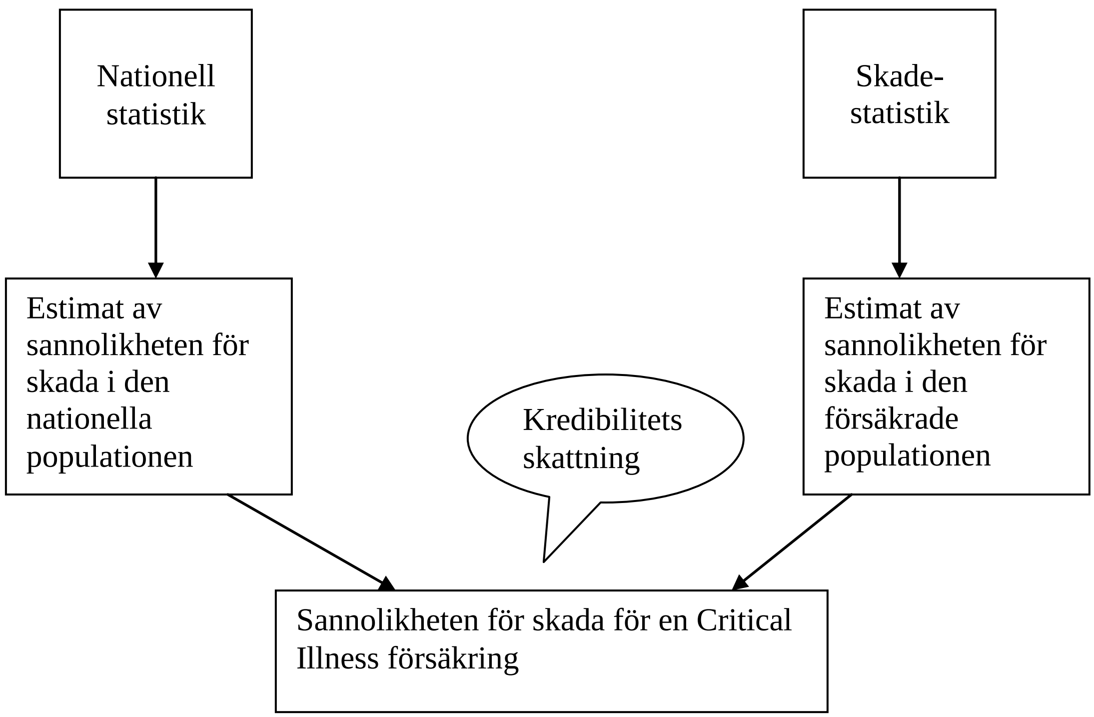
Vi kan inte tillämpa kredibilitetsteori på det sättet eftersom hela portföljen är för liten. Tar vi tanken ett steg till, så att nationell data speglar hela försäkringssportföljen och det faktiska skadedatat för den försäkrade portföljen speglar den individuella risken, så skulle kredibilitetsteori vara ett möjligt sätt att modellera priset på. Det är detta som vi kommer att göra i detta arbete.

2.4 Metod om skadedata finns

Som vi tidigare har konstaterat är omfattningen av skadedata litet. Vi kommer därför att använda både skadedata från redan såld försäkring och nationell data för att estimera sannolikheten för skada. Dessa två estimat kommer att viktas ihop med hjälp av kredibilitetsteori. Estimatet baserat på nationell data är mer pålitligt ur statistisk synvinkel men inte ekvivalent med estimatet från skadedatat pga. av urvalet.

Som beskrivits ovan finns två varianter av Critical Illness försäkringen, dels som tillägg till en livförsäkring och dels som separat försäkring. Sannolikheten för skada för en Critical Illness försäkring såld som ett tillägg till en livförsäkring och inkluderar plötslig död beräknas som antalet individer som diagnostiserat med cancer eller avlidit pga. cancer dividerat med antalet friska individer i populationen. För en Critical Illness försäkring som köpts separat och endast utbetalas om försäkringstagaren lever beräknas sannolikheten för skada som antalet individer som diagnostiserats med cancer dividerat med antalet friska individer i populationen.

Metoden beskrivs schematiskt i figur 1 nedan.

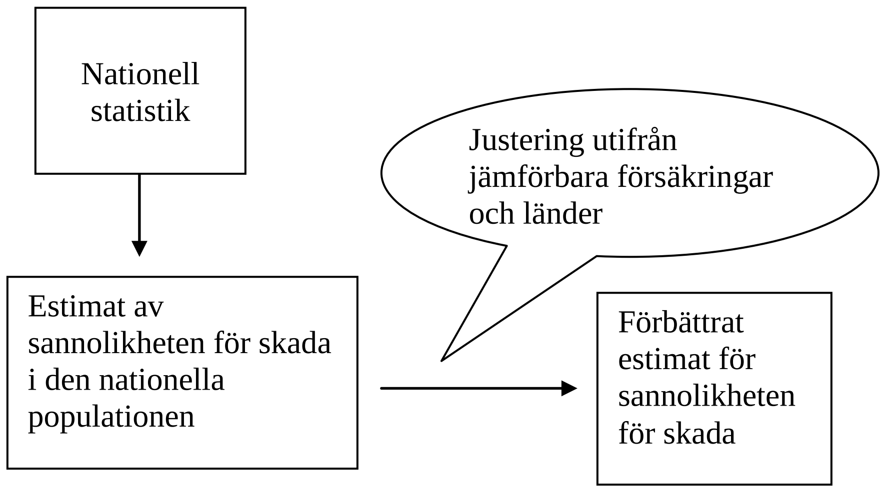


Figur 2.4. Metod om skadedata finns.

2.5 Metod om det inte finns någon skadedata

Om det inte finns någon skadedata är nationell data den enda data som är tillgänglig att använda. Vi har tidigare diskuterat möjligheten att nationell data inte helt är överensstämmande med skadedata. Den försäkrade populationen borde vara friskare än befolkningen i övrigt eftersom de vid tecknandet av försäkringen är dokumenterat friska. Misstankarna är dock de motsatta, dvs att de försäkrade har en högre sannolikhet att insjukna i någon av de specificerade sjukdomarna. Om det visar sig att det är skillnader mellan den försäkrade populationen och befolkningen i övrigt så borde prissättningsmodellen kompensera för detta faktum. Ett sätt att justera är att titta på jämförbara länder när det gäller standard och hälsa och som har skadedata. Skillnaden mellan estimat baserat på skadedata och estimat baserat på nationell överförs till estimatet som ska förbättras.

Metoden beskrivs schematiskt i figur 2 nedan.



Figur 2.5 Metod om skadedata finns.

2.6 Begränsningar

I detta arbete gör vi antagandet att försäkringen enbart täcker sjukdomen cancer. Antagandet är en konsekvens av tillgången på data. Cancer står för närmre femtio procent av alla rapporterade skador och är den sjukdom som är bäst registrerad i nationella databaser. Modellen går dock lätt att utvidga till att täcka flera sjukdomar. Definitionen av sjukdomarna i villkoren kan variera mellan olika försäkringsbolag beroende på hur försäkringsbolagen har valt att utforma produkten. I vårt fall gäller följande definition av cancer:

En malignttumör karaktäriseras av en okontrollerad tillväxt och spridning av maligna celler och invasion i vävnad. Termen cancer inkluderar leukemi och Hodgkin’s sjukdom men följade är exkluderat.

- Alla tumörer som är histologiskt beskrivna som pre-maligna, icke invasiva eller cancer in situ.
- Kaposl´s sarcoma i samband med Human Immunodeficiency Virus.
- Annan hudcancer än malignt melanom.

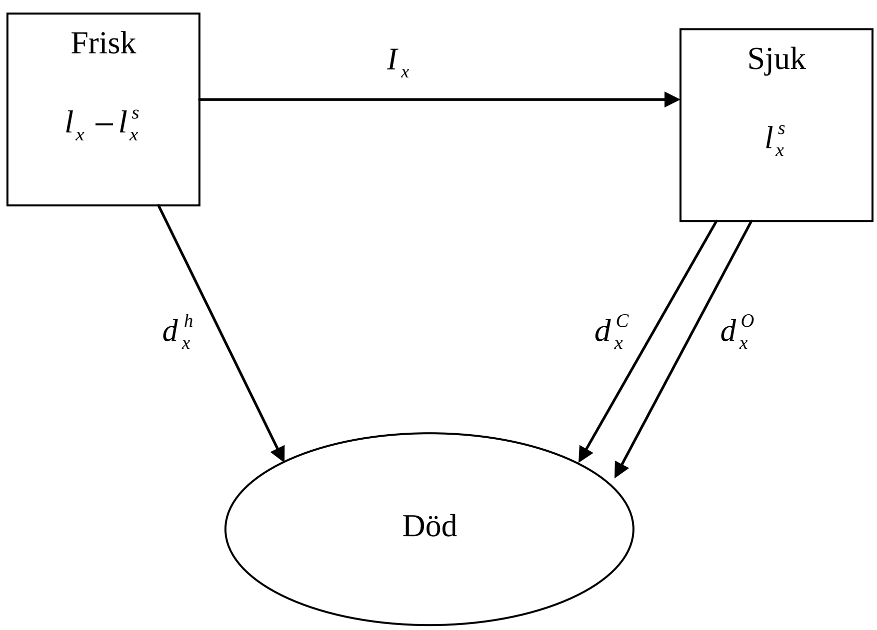
De allra flesta Critical Illness försäkringar i vårt datamaterial är försäkringar som inkluderar plötslig död dvs Critical Illness försäkringen är såld tillsammans med en livförsäkring. Den begränsade mängden data gör att vi i detta arbete endast kan estimera priset för en Critical Illness såld som ett tillägg till en livförsäkring. Vi kommer dock att redovisa hur skattningar kan göras för Critical Illness såld som en ensam försäkring när beräkningarna skiljer sig åt.

3 Estimering av sannolikheten för critical illness eller dödsfall

3.1 Skattning av sannolikheten för skada

Låt CI beteckna Critical Illness. För en försäkring som inkluderar plötsligt dödsfall är sannolikheten för skada lika med sannolikheten för CI eller för tidig död under ett år. För en försäkring som endast täcker risken för CI är sannolikheten för skada lika med sannolikheten för CI under ett år. Vid ingången av året är individerna i populationen antingen friska, dvs ej diagnostiserade med CI, eller sjuka och diagnostiserade med CI.

Antag följande beteckningar:
$l_x =$ Antal individer i populationen vid ingången av året
$l_x^s =$ Antal individer i populationen som har CI vid ingången av året
$I_x =$ Antalet individer i populationen som diagnostiseras med CI under året
$d_x^h =$ Antalet individer i den friska populationen som avlider av andra orsaker än CI under året
$d_x^C =$ Antalet individer som avlider av CI i populationen l_x^s under året
$d_x^O =$ Antalet individer som avlider av andra orsaker än CI i populationen l_x^s under året
$d_x =$ Antalet individer som avlider i populationen under året.
där x=året



Figur 3.1

För en Critical Illness såld som en tilläggsförsäkring till en livförsäkring, en sk Accelerated benefit är antalet skador lika med antalet individer som diagnostiserat med CI under ett år oavsett om de avlider av CI eller ej plus antalet individer som är friska vid ingången av året men som avlider av andra orsaker än CI under året. Beteckna denna sannolikhet med φ^{acc}.

För en Critical Illness försäkring som sålts som en ensam försäkring en sk Additional benefit är antalet skador lika med antalet individer som diagnostiseras med CI året. Beteckna denna sannolikhet med φ^{add}.

$$\varphi^{acc} = \frac{I_x + d_x^h}{l_x - l_x^s} \qquad\qquad \varphi^{add} = \frac{I_x}{l_x - l_x^s}$$

Tyvärr saknas informationen om antalet döda som avlider av andra orsaker än CI i den friska populationen, d_x^h

Vi tar hjälp av Sats 3.1 för att lösa detta problem att skatta φ^{acc}.

Sats 3.1

a) ***Accelerated benefit.*** *Sannolikheten för skada i den nationella populationen under ett år är:*

$$\varphi^{acc} = i_x + (q_x - k_x q_x)\frac{l_x - l_x^s - I_x}{l_x - l_x^s}$$

b) ***Additional benefit.*** *Sannolikheten för skada i den nationella populationen under ett år är:*

$$\varphi^{add} = i_x$$

där

$i_x = \dfrac{I_x}{l_x - l_x^s}$ *sannolikheten att diagnosteras med CI under år 1 för individer som var friska vid ingången av år 1.*

$k_x = \dfrac{d_x^C}{d_x}$ *Sannolikheten att dö av CI givet att död inträffar.*

$q_x = \dfrac{d_x}{l_x}$ *Sannolikheten för att dö i den nationella populationen.*

Bevis:

Antalet döda i den nationella populationen är $d_x = d_x^h + d_x^C + d_x^O$. *(3.1)*

Vi visar nu satsen för accelerated benefit (a).

Accelerated benefit utfaller om den försäkrade diagnostiseras med CI eller för tidig död så totalt antal skador under ett år är lika med $d_x^h + I_x$ *och sannolikheten för skada under ett år är lika med* $\varphi^{acc} = \frac{I_x + d_x^h}{l_x - l_x^s}$

Problemet är att d_x^h *är okänt. För att undervika detta problem och göra estimatet möjligt att skatta är det nödvändigt att anta följande. Sannolikheten för död pga. andra orsaker än CI i den friska populationen är lika med sannolikheten för död pga. andra orsaker än CI i den populationen diagnostiserade med CI.* I_x *inkluderar de fall där en individ diagnostiseras med CI och avlider inom kort därefter.*

Antagandet är:

$$\frac{d_x^h}{l_x - l_x^s - I_x} = \frac{d_x^O}{l_x^s + I_x} \tag{3.2}$$

Använd antagande (3.2) i ekvation (3.1):

$$d_x - d_x^C = d_x^h + d_x^O = d_x^h + d_x^h \frac{l_x^s + I_x}{l_x - l_x^s - I_x} = d_x^h \frac{l_x}{l_x - l_x^s - I_x} \Rightarrow \frac{d_x^h}{l_x - l_x^s} = \frac{d_x - d_x^C}{l_x} * \frac{l_x - l_x^s - I_x}{l_x - l_x^s}$$

Sätt

$$k_x = \frac{d_x^C}{d_x} \qquad q_x = \frac{d_x}{l_x} \qquad i_x = \frac{I_x}{l_x - l_x^s}$$

Genom att använda definitionen ovan erhålls önskad formel:

$$\varphi^{acc} = \frac{I_x + d_x^h}{l_x - l_x^s} = \frac{I_x}{l_x - l_x^s} + \frac{d_x - d_x^C}{l_x} * \frac{l_x - l_x^s - I_x}{l_x - l_x^s} = i_x + (q_x - k_x q_x) * \frac{l_x - l_x^s - I_x}{l_x - l_x^s}$$

4 Estimera sannolikheten för CI och dödsfall i Storbritannien

Uppgiften är att estimera premien för produkten Critical Illness i Sverige och Storbritannien. På grund av tillgången på data antar vi att den enda CI som täcks av försäkringen är cancer. Vi har grupperat riskerna efter kön och ålder. Begränsningen i datamängden har gjort att vi har varit tvungna att göra grupperingar med stora ålderspann på 14 år.

4.1 Sannolikheten för skada i Storbritannien baserat på skadedata.

Vi startar med estimatet för sannolikheten för skada i den försäkrade populationen för försäkringar sålda som tillägg till livförsäkring (accelerated benefit).

$$\hat{\varphi}_I^{acc} = \frac{antal\ skador}{antal\ försäkrade}$$

Nedan följer en sammanställning av datamaterialet. Tabell 4.1.1 visar vår exponering, antalet försäkrade individer grupperade efter kön och ålder. I tabell 4.1.2 följer antalet anmälda skador dvs. antalet individer i vår exponering som diagnostiserat med cancer.

Tabell 4.1.1 Antal försäkrade individer

Kön	Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000	summa
F	20-34	221	200	275	273	882	800	738	734	504	4 627
F	35-49	237	224	297	326	1 234	1 136	1 037	896	834	6 221
F	50-64	15	16	33	47	212	205	252	250	294	1 324
											0
M	20-34	802	656	865	758	2 121	2 072	2 039	1 693	1 028	12 034
M	35-49	898	759	944	982	3 657	3 321	2 750	2 015	1 282	16 608
M	50-64	55	58	89	149	641	583	516	441	428	2 960
	summa	2 228	1 913	2 503	2 535	8 747	8 117	7 332	6 029	4 370	43 774

Tabell 4.1.2 Antal skador

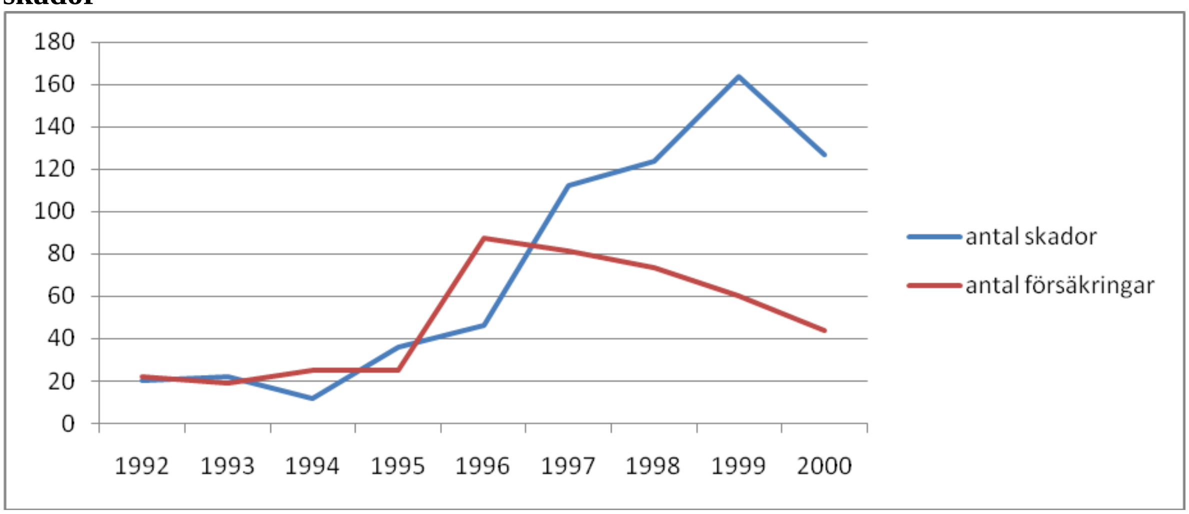
Kön	Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000	summa
F	20-34	2	2	2	2	0	8	10	2	8	36
F	35-49	4	0	0	6	12	20	40	42	32	156
F	50-64	0	0	0	0	2	10	16	18	18	64
											0
M	20-34	10	14	6	6	4	16	4	14	7	81
M	35-49	4	2	2	16	16	32	28	53	43	196
M	50-64	0	4	2	6	12	26	26	35	19	130
	summa	20	22	12	36	46	112	124	164	127	663

Vårt data omfattar 9 år och totalt 43 774 försäkrade individer. Av dem har 663 individer diagnostiserats med cancer eller avlidit en för tidig död. Vi kan se att antalet försäkrade individer ökade kraftigt mellan år 1995 och 1996. Antalet försäkrade individer mer än tredubblades för att sedan gå tillbaka något. Ökningen i antalet skador gick långsammare och nådde sin högsta nivå 1999, tre år senare. Vi kan se detta i Graf 4.1.1 där antalet försäkrade individer är delat med 100 för att grafen ska få en tydligare skala. Detta syns även i skattningen av $\hat{\varphi}_I^{acc}$, se tabell 4.1.3, där sannolikheten för skada har ökat de sista åren både för män och kvinnor.

Tabell 4.1.3 Skattning av $\hat{\varphi}_I^{acc}$

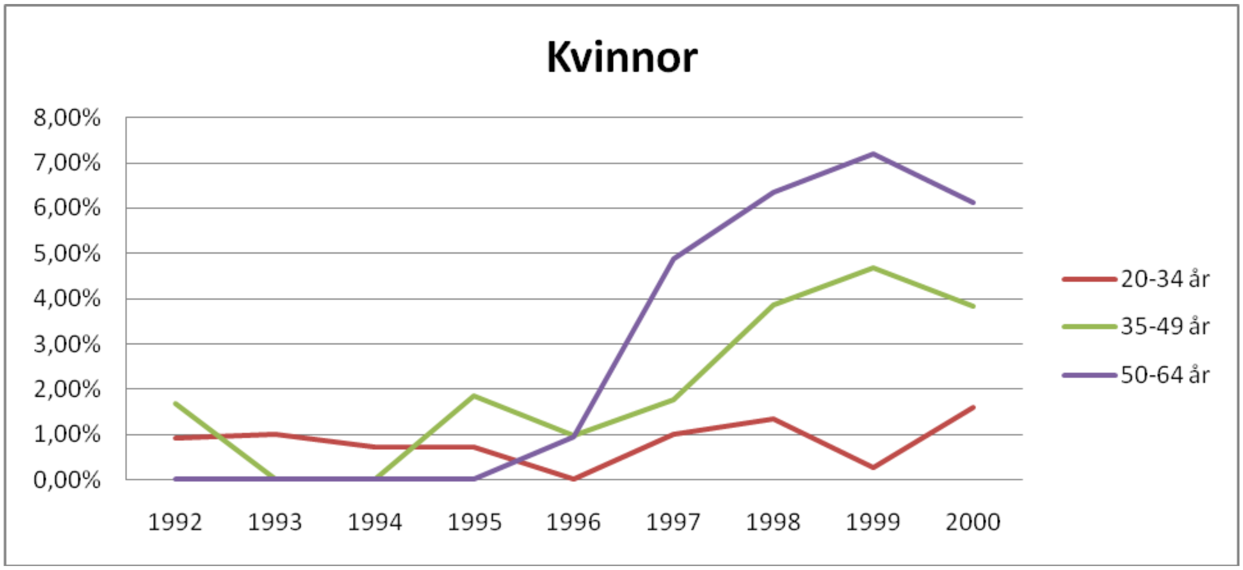
Kön	Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000
F	**20-34**	0,90%	1,00%	0,73%	0,73%	0,00%	1,00%	1,36%	0,27%	1,59%
F	**35-49**	1,69%	0,00%	0,00%	1,84%	0,97%	1,76%	3,86%	4,69%	3,84%
F	**50-64**	0,00%	0,00%	0,00%	0,00%	0,94%	4,88%	6,35%	7,20%	6,12%
M	**20-34**	1,25%	2,13%	0,69%	0,79%	0,19%	0,77%	0,20%	0,83%	0,68%
M	**35-49**	0,45%	0,26%	0,21%	1,63%	0,44%	0,96%	1,02%	2,63%	3,35%
M	**50-64**	0,00%	6,90%	2,25%	4,03%	1,87%	4,46%	5,04%	7,94%	4,44%

Graf 4.1.1 Utvecklingen av antal försäkringar och skador

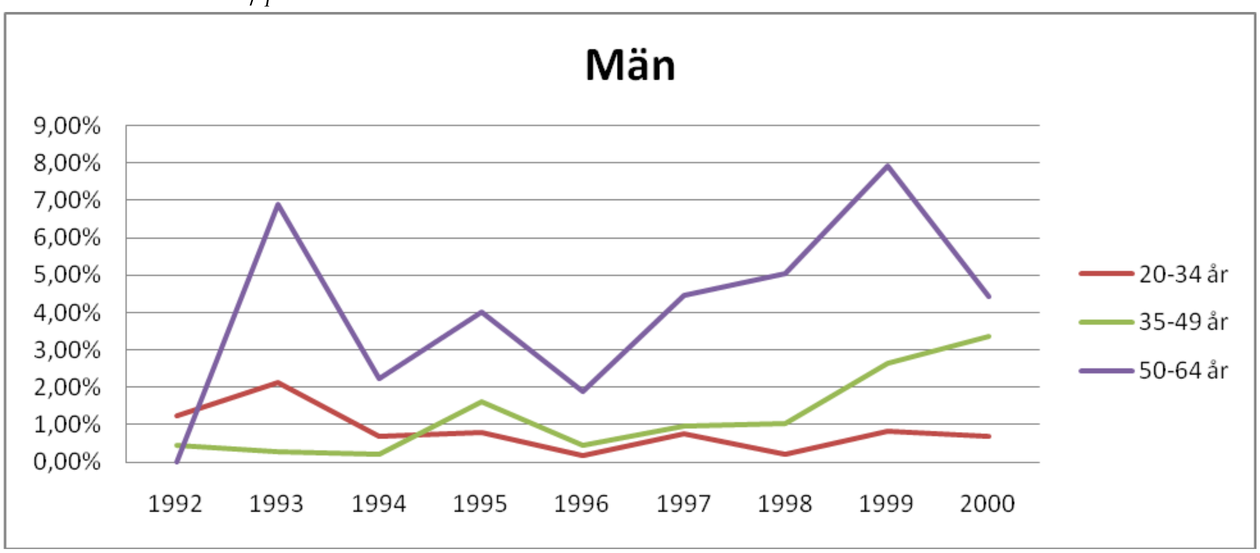


I graferna 4.1.2a och 4.1.2b nedan ser vi estimerad sannolikhet för skada, $\hat{\varphi}_I^{acc}$, för kvinnor respektive män. Från år 1996 då portföljen växer kraftigt är det tydligt att sannolikheten för skada accelererar. Detta visar att de nya individerna i portföljen har en högre sannolikhet för skada än de redan försäkrade. Eftersom priset för de nya individerna är det samma som för de redan försäkrade individerna så innebär detta att lönsamheten i portföljen minskar. Vi ser även att ålder är en viktig parameter.

Graf 4.1.2a: Kvinnor $\hat{\varphi}_I^{acc}$



Graf 4.1.2b: Män $\hat{\varphi}_I^{acc}$



Den här typen av försäkring, där sannolikheten för skada är relativt låg kräver en omfattande försäkringsportfölj för att det ska vara möjligt att göra säkra skattningar baserat på skadedata. I vårt material, med relativt få försäkrade individer, blir konsekvensen att antalet skador blir mycket få. När vi sedan gör indelningen i ålder och kön saknar vi helt observationer i vissa riskgrupper och de skattningar som kan beräknas blir mycket osäkra. Vi måste dock göra någon indelning efter premieargumenten kön och ålder eftersom vi vet att dessa har stor betydelse för risken att få cancer eller att avlida.

4.2 Sannolikheten för skada i Storbritannien baserat nationell data

Vi fortsätter med att skatta sannolikheten för skada i den Storbritanniska populationen. Sannolikheten för skada för acceleration benefit är lika med sannolikheten att diagnostiseras med cancer eller att avlida i förtid. För additional benefit är sannolikheten för skada lika med risken för att diagnostiseras med cancer. För att estimera risken använder vi sats 3.1.

$$\varphi_N^{acc} = i_x + (q_x - k_x q_x) * \frac{l_x - l_x^s - I_x}{l_x - l_x^s}$$

$$\varphi_N^{add} = i_x$$

där

$$\hat{i}_x = \frac{I_x}{l_x - l_x^s} \qquad \hat{k}_x = \frac{d_x^c}{d_x} \qquad \hat{q}_x = \frac{d_x}{l_x}$$

Vi ser att risken att diagnostiseras av sjukdom är åldersberoende och det är stor skillnad på risken inom ett 15 års spann. I den försäkrade populationen är åldersfördelningen i de tre åldersgrupperna mycket stabil över åren. Vi väljer därför att i försäkringsdatat beräkna åldersfördelning inom respektive åldersklass över alla år (se tabell 4.2.1) och viktar det nationella datat därefter. Detta för att få en mer rättvisande skattning av nationsestimaten när vi senare ska jämföra dem med estimaten från försäkringsdatat.

Tabell 4.2.1 Åldersfördelningen inom resp. åldersklass i den försäkrade populationen.

	F	M
20-24	**5%**	**4%**
25-29	**35%**	**32%**
30-34	**60%**	**64%**
35-39	**46%**	**47%**
40-44	**32%**	**33%**
45-49	**22%**	**20%**
50-54	**80%**	**80%**
55-59	**18%**	**18%**
60-64	**2%**	**2%**

Ett annat problem är att l_x^s inte är tillgängligt. En lösning är att l_x^s är relativt litet i förhållande till l_x och därför ersätter vi $l_x - l_x^s$ med l_x. En annan möjlig approximation är att anta att en person i genomsnitt är sjuk i cancer i 5 år så att $l_x - l_x^s$ är approximerat lika med $l_x - 5 * I_x$.
Vi provar båda approximationerna (se tabell 4.2.2 och 4.2.3) och ser om det ger någon skillnad.

Tabell 4.2.2 Skattning av $\hat{\varphi}_N^{acc}$ med approximationen $l_x - l_x^s = l_x$

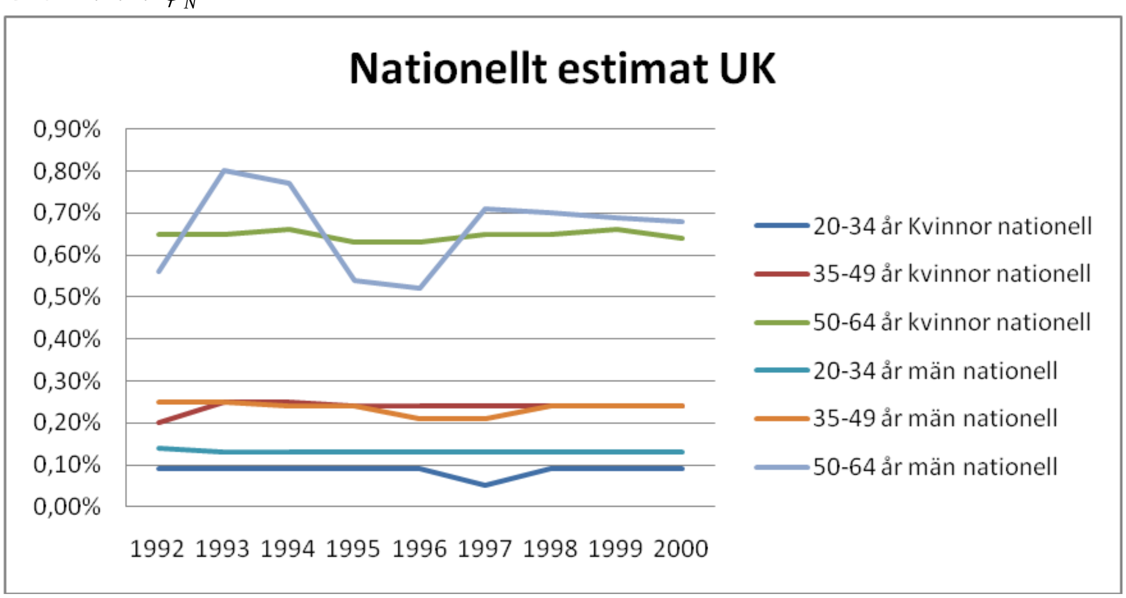
Kön	Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000
F	20-34	0,09%	0,09%	0,09%	0,09%	0,09%	0,05%	0,09%	0,09%	0,09%
F	35-49	0,20%	0,25%	0,25%	0,24%	0,24%	0,24%	0,24%	0,24%	0,24%
F	50-64	0,65%	0,65%	0,66%	0,63%	0,63%	0,65%	0,65%	0,66%	0,64%
M	20-34	0,14%	0,13%	0,13%	0,13%	0,13%	0,13%	0,13%	0,13%	0,13%
M	35-49	0,25%	0,25%	0,24%	0,24%	0,21%	0,21%	0,24%	0,24%	0,24%
M	50-64	0,56%	0,80%	0,77%	0,54%	0,52%	0,71%	0,70%	0,69%	0,68%

Tabell 4.2.3 Skattning av $\hat{\varphi}_N^{acc}$ med approximationen $l_x - l_x^s = l_x - 5 * I_x$

Kön	Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000
F	20-34	0,09%	0,09%	0,09%	0,09%	0,09%	0,05%	0,09%	0,09%	0,09%
F	35-49	0,20%	0,25%	0,25%	0,24%	0,24%	0,24%	0,24%	0,24%	0,24%
F	50-64	0,65%	0,65%	0,66%	0,63%	0,63%	0,65%	0,65%	0,66%	0,64%
M	20-34	0,14%	0,13%	0,13%	0,13%	0,13%	0,13%	0,13%	0,13%	0,13%
M	35-49	0,25%	0,25%	0,24%	0,24%	0,21%	0,21%	0,24%	0,24%	0,24%
M	50-64	0,56%	0,80%	0,77%	0,54%	0,52%	0,71%	0,70%	0,69%	0,68%

Vi ser att skillnaden mellan de två olika approximationerna är obetydliga, det skiljer sig först på tredje decimalen. Antalet individer i åldern 20-64 i Storbritannien är mycket stor, över 30 miljoner, vilket innebär att skattningarna blir mycket stabila vilket syns i graf 4.2.1 nedan. Risken att drabbas av cancer eller för tidig död ökar kraftigt med åldern.

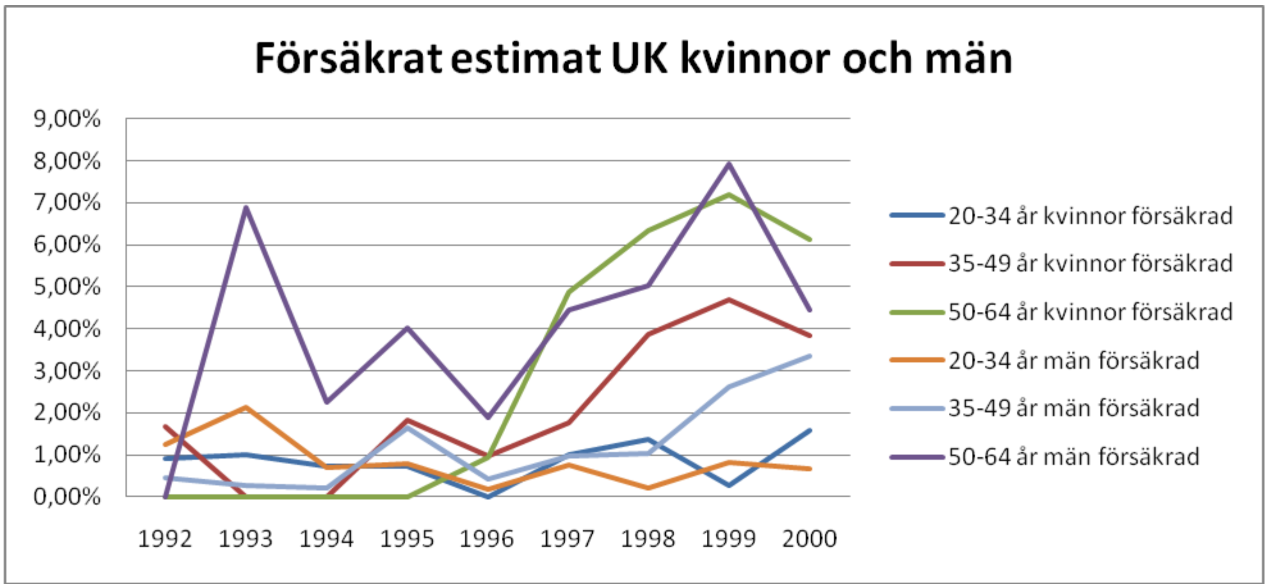
Graf 4.2.1: $\hat{\varphi}_N^{acc}$



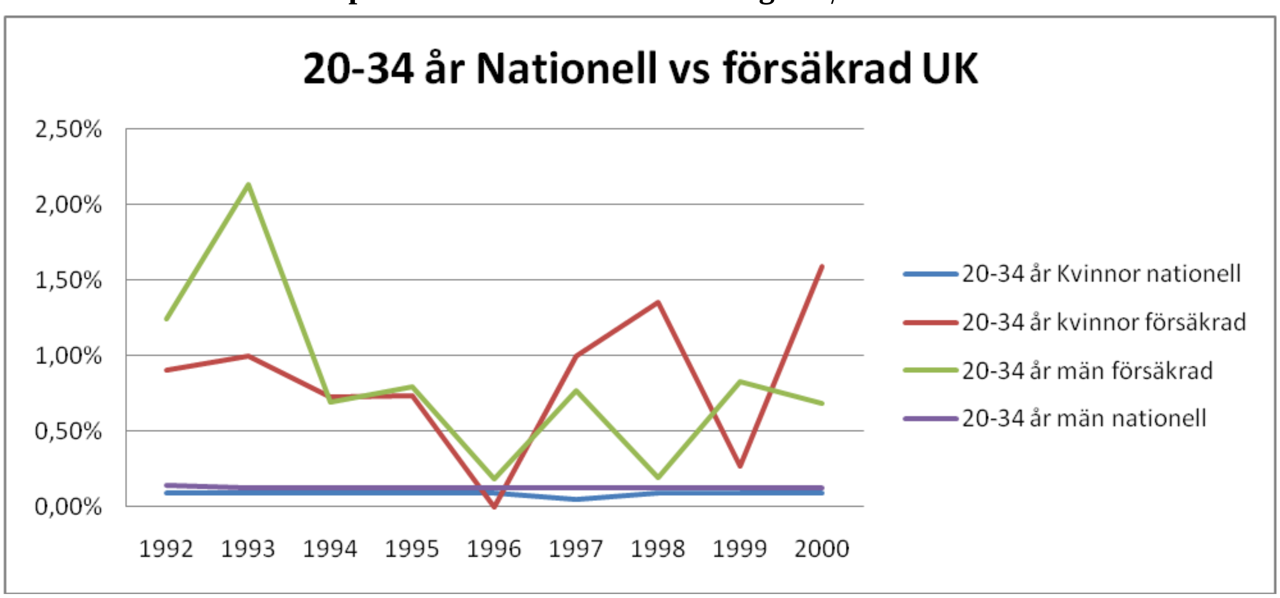
4.3 Analys av skillnaden i skadedata och nationsdata

Det är stor skillnad mellan sannolikheterna för den nationella och den försäkrade populationen. Det är tydligt att sannolikheten för skada är betydligt högre i den försäkrade populationen, även om skattningarna är mycket osäkra. Tittar man på åren 1996 och framåt i försäkrad population, då portföljen är väsentligt större, ser vi i grafen 4.3.1 och 4.3.2 samma tendens som i graf 4.2.1 för nationell population att kurvorna för kvinnor och män följs åt i ålderklass. Detta tyder på att observationerna i den försäkrade populationen inte helt är slumpmässiga.

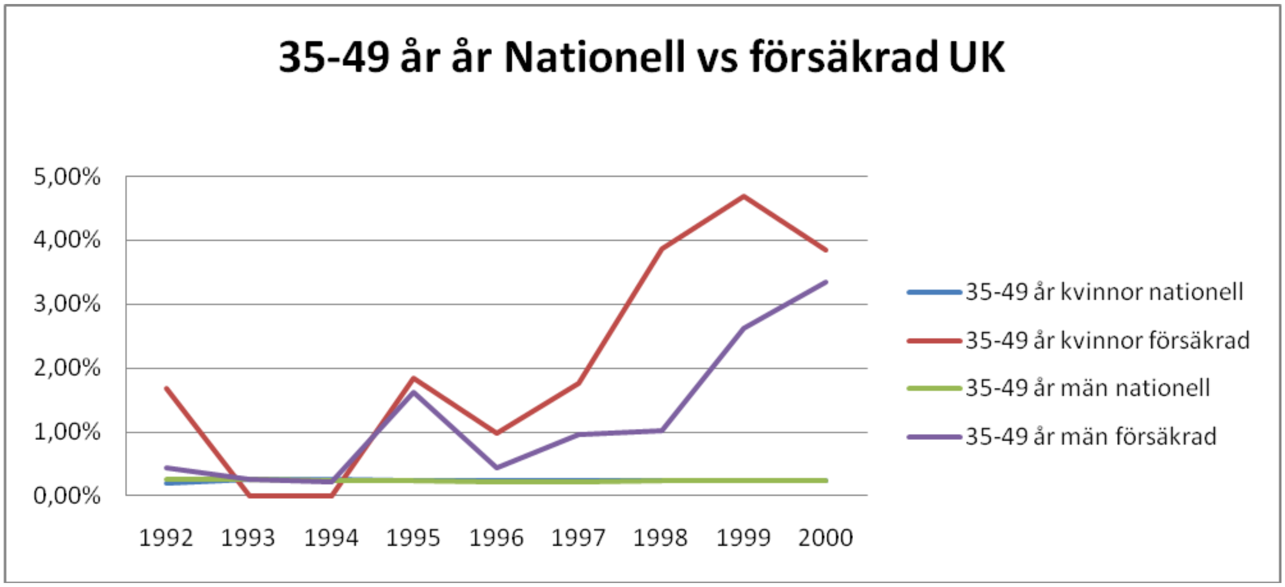
Graf 4.3.1: $\hat{\varphi}_I^{acc}$



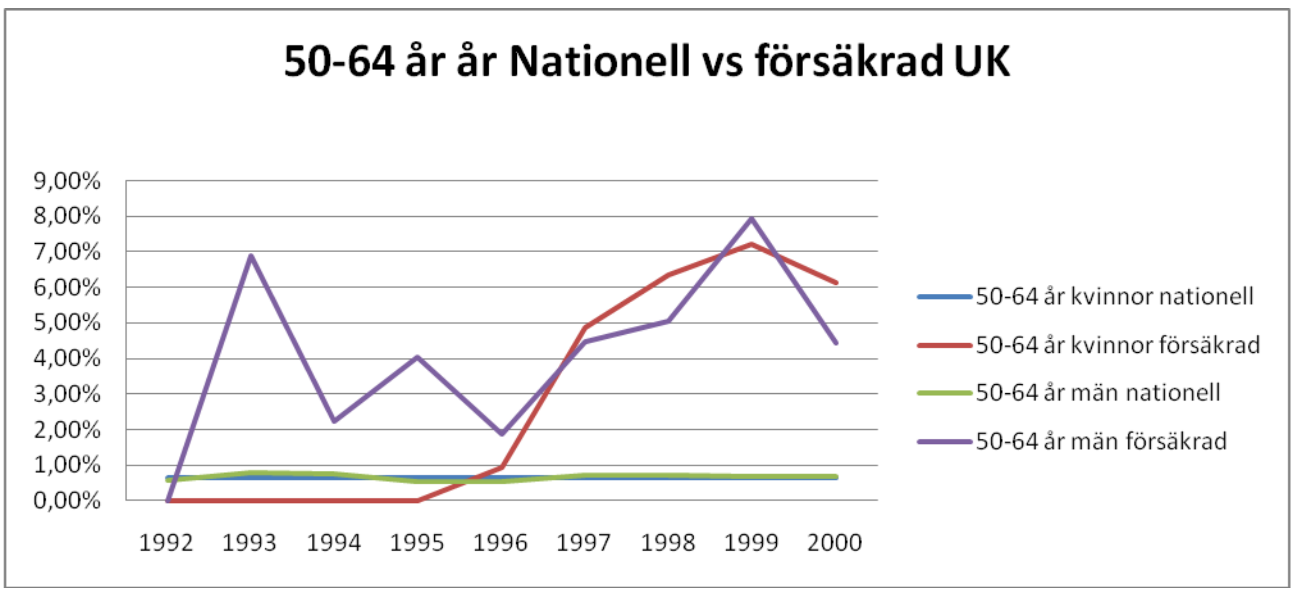
Graf 4.3.2a: Nationell resp. försäkrad 20-34 år skattning av $\hat{\varphi}^{acc}$



Graf 4.3.2b: Nationell resp. försäkrad 35-49 år skattning av $\hat{\varphi}^{acc}$



Graf 4.3.2c: Nationell resp försäkrad 50-64 år skattning av $\hat{\varphi}^{acc}$



5 Kredibilitetsteori

5.1 Introduktion

När man arbetar med att prissätta försäkringar tittar man på de skador som redan har inträffat för att på så sätt kunna uppskatta eventuella skador i framtiden och därmed kunna bestämma priset. Alla risker grupperas utifrån de egenskaper som ger dem liknande risk och varje riskgrupp får sitt pris beroende av dessa riskparametrar. För att ta fram priset för en individuell försäkring justeras den gemensamma klassraten ofta med skadestatistik från den individuella försäkringen för att få ett pris som bättre speglar den unika risken. Det gemensamma priset för försäkringarna i en riskgrupp är baserat på många observationer och är därmed relativt stabilt men kanske mindre träffsäkert än om priset skulle beräknas enbart på den unika riskens skadedata. De försäkringar som har ett lägre individuellt genomsnittspris än riskklassens genomsnittspris är goda risker och det finns en möjlighet att de kan få ett bättre pris någon annanstans. Försvinner de bra riskerna försämras portföljen, och även om priset justeras på sikt förlorar försäkringsbolaget premievolym vilket därmed minskar chanserna till god avkastning.

Hur ska då klassraten och den individuella raten kombineras? Kredibilitetsteorin hjälper till att svara på den frågan. Kredibilitetsteorin tar hänsyn till två saker; hur homogen riskklassen är och hur stor risk spridning det är inom en unik risks individuella skadedata. Om alla riskerna i klassen är identiska och har samma förväntade skadeutfall finns det ingen mening med att ta hänsyn till den individuella skadutfallet. Finns det däremot skillnader mellan riskerna i samma riskklass så finns det en nytta med att justera riskklasspriset med den unika riskens skadedata. Är det stora variationer inom den unika riskens skadedata kan det faktiska skadeutfallet för den enskilda risken vara långt ifrån det individuella förväntade värdet vilket gör det individuella skadeutfallet mindre användbart. Kredibilitetsteorin försöker ta reda på dessa två omständigheter och hitta den optimala kombinationen av utfallet för en individuell försäkring som är relevant men har större osäkerhet och utfallet för riskklassen som är stabil men mindre relevant.

Vi kommer att se att förväntat utfall för det individuella avtalet är ett viktat genomsnitt av förväntat utfall för riskklassen och ett genomsnitt för utfallet för det enskilda avtalet.

Utfallet för den unika risken är:

$$Estimat = Z\bar{X} + (1 - Z)\mu$$

där
$\bar{X}$= medelvärdet för den enskilda risken
μ= medelvärdet för hela riskklassen
Z = kredibilitetsfaktorn för risken

5.2 Definition och notation

Sätt den individuella risken j i en riskklass under tiden t till den stokastiska variabeln X_{jt} med det faktiska utfallet x_{jt}.

X_{jt} kan representera olika saker t.ex. antal skador under period t eller skadeprocent under år t.

Varje risk j har en riskparameter θ_j som särskiljer den individuella riskens egenskaper. θ_j är en stokastisk variabel som betecknas ϑ med täthetsfunktionen $f_\vartheta(\theta)$. Om två risker har samma parameter θ så har de samma riskegenskaper och samma väntevärde $\mu(\theta)$.

Nästan alla modeller inom kredibilitetsteorin gör antagandet att θ_{jt} är tidshomogen, dvs riskparametern för en risk j ändras inte över tiden . Vi utlämnar därför notationen t och skriver θ_j.

Ett försäkringsavtal j beskrivs alltså av följande vektor $(\theta, X_1, X_2, ..., X_t)$ av stokastiska variabler där variabeln θ inte är observerbar. Fördelningen av den stokastiska variabeln X_t beror på värdet av θ. Om X_{jt} är en kontinuerlig variabel sätts den betingade täthetsfunktionen till $f_{X|\vartheta}(x_{jt}|\theta)$

För den enskilda risken j är:
Väntevärdet för X_t givet $\vartheta = \theta$:

$$E_{X|\vartheta}[X_t|\vartheta = \theta] = \int x_t f_{X|\vartheta}(x_t|\theta)dx_t = \mu(\theta)$$

Det obetingade väntevärdet av X_t är väntevärdet över alla risker i riskklassen

$$E[X_t] = \int \left[\int x_t f_{X|\vartheta}(x_t|\theta)dx_t\right] f_\vartheta(\theta)d\theta = E_\vartheta\left[E_{X|\vartheta}[X_t|\vartheta]\right] = E_\vartheta[\mu(\theta)] = \mu$$

Den betingande variansen för X_t givet $\vartheta = \theta$ beskriver en risks slumpmässiga utfall från sitt förväntade värde:

$$Var_{X|\vartheta}[X_t|\vartheta = \theta] = E_{X|\vartheta}[(X_t - \mu(\theta))^2|\vartheta = \theta] = \int\int (X_t - \mu(\theta))^2 f_{X|\vartheta}(x_t|\theta)dx_t = \sigma^2(\theta)$$

Den obetingade variansen för X_t kallas ofta för den totala variansen:
$Var[X_t] = Var_\vartheta\left[E_{X|\vartheta}[X_t|\vartheta]\right] + E_\vartheta\left[Var_{X|\vartheta}[X_t|\vartheta]\right]$

5.3 Bühlmanns kredibilitetsmodell

Bühlmanns kredibilitetsmodell förutsätter att alla observationerna har samma vikt och liknande riskvolym.

Låt väntevärdet för hela riskklassen:
$\mu = E_{\vartheta}[\mu(\theta)] = E_{\vartheta}\left[E_{X|\vartheta}[X_t|\vartheta]\right]$

förväntat värde av variansen för en risks slumpmässiga utfall från sitt förväntade värde:
$EPV = E_{\vartheta}[\sigma^2(\theta)] = E_{\vartheta}\left[Var_{X|\vartheta}[X_t|\vartheta]\right]$

variansen för variationen av den enskilda riskens medelvärde kring riskklassmedelvärdet μ:
$VHM = Var_{\vartheta}[\mu(\theta)] = E_{\vartheta}[(\mu(\theta) - \mu)^2]$

Bühlmanns kredibilitetsmodell antar att:

1. För någon vald risk j=1,…,k så är paren (θ_j, X_{jt}) oberoende och likafördelade.
2. För varje risk j=1,…k och konstant θ_j för varje tid t så är de betingade variablerna $X_{1j} | \theta_j, X_{2j} | \theta_j, \ldots, X_{mj} | \theta$ oberoende och lika fördelade.
3. $E(X^2) < \infty$

Det betyder att riskerna är oberoende av varandra och inom en risk med givet θ_j är de betingade stokastiska variablerna lika över tiden.

$E(X_{jr} | \theta_j) = E(X_{js} | \theta_j)$
$Var(X_{jr} | \theta_j) = Var(X_{js} | \theta_j)$

Bühlmanns kredibilitetsviktade skattning för $\mu(\theta) = E_{X|\vartheta}[X_t|\theta]$ för t=1,2,…,K är

$$\hat{\mu}(\theta) = Z\bar{X} + (1 - Z)\mu$$

där $= \frac{N}{N+K}$ och $K = \frac{EPV}{VHM}$

Se bevis i appendix.

$\mu(\theta)$ är okänd för den valda risken j därför används medelvärdet $\bar{X} = \left(\frac{1}{N}\right)\sum_{t=1}^{N} X_t$ för att skatta väntevärdet. Medelvärdet $\bar{X}$ är en unbiased skattning.

$$E_{X|\vartheta}[\bar{X}|\theta] = E_{X|\vartheta}\left[\left(\frac{1}{N}\right)\sum_{t=1}^{N} X_t\,|\theta\right] = \left(\frac{1}{N}\right)\sum_{t=1}^{N} E_{X|\vartheta}[X_t|\theta] = \left(\frac{1}{N}\right)\sum_{t=1}^{N} \mu(\theta) = \mu(\theta)$$

Den betingade variansen av $\bar{X}$:

$$Var_{X|\vartheta}[\bar{X}|\theta] = Var_{X|\vartheta}\left[\left(\frac{1}{N}\right)\sum_{t=1}^{N} X_t\,|\theta\right] = \left(\frac{1}{N}\right)^2 \sum_{t=1}^{N} Var_{X|\vartheta}[X_t|\theta] = \left(\frac{1}{N}\right)^2 \sum_{t=1}^{N} \sigma^2(\theta) = \frac{\sigma^2(\theta)}{N}$$

Den obetingade variansen av $\bar{X}$:

$$Var[\bar{X}] = Var_{\vartheta}\left[E_{X|\vartheta}[\bar{X}|\vartheta]\right] + E_{\vartheta}\left[Var_{X|\vartheta}[\bar{X}|\vartheta]\right] = Var_{\vartheta}[\mu(\theta)] + \frac{E_{\vartheta}[\sigma^2(\theta)]}{N} = VHM + \frac{EPV}{N}$$

Och Z kan uttryckas som $Z = \frac{VHM}{Var[\bar{X}]} = \frac{Var_{\vartheta}[\mu(\theta)]}{Var[\bar{X}]}$

Täljaren uttrycker hur spridda medelvärdena är för de olika riskerna i riskgruppen och nämnaren är ett mått på den totala variansen.

6 Tillämpa kredibilitetsteori på UK skattningarna av $\hat{\varphi}^{acc}$

Vi har estimaten $\hat{\varphi}_I^{acc}$ och $\hat{\varphi}_N^{acc}$ för Storbritannien. Vi vill nu vikta ihop dessa två estimat av sannolikheten att diagnostiseras med cancer under ett år på bästa sätt för få en så bra skattning som möjligt.

6.1 Kredibilitetsskattning av $\hat{\varphi}_I^{acc}$ och $\hat{\varphi}_N^{acc}$

Antag att X_{jt} är lika med sannolikheten för CI. Vi har två observationer i riskklassen. Den första observationen är från den försäkrade populationen $\hat{\varphi}_I^{acc}$, tabell 4.1.3 och den andra observationen är den nationella populationen $\hat{\varphi}_N^{acc}$, tabell 4.2.2. Båda observationerna är från tidsperioden 1992 till 2000 vilket är 9 år och ger t=9 och j=2.

Beräkning av EPV

Ett estimat av förväntat värde av variansen för de enskilda riskernas slumpmässiga utfall från deras förväntade värde.

$$\widehat{EPV} = \frac{1}{R}\sum_{j=1}^{R}\sigma_j^2 = \frac{1}{R(N-1)}\sum_{j=1}^{R}\sum_{t=1}^{N}(X_{jt-}\bar{X}_j)^2$$

Beräkning av VHM
Den totala variansen kan utryckas som

$$Var(\bar{X}_j) = \mathrm{Var}\left(\mathrm{E}(\bar{X}_j|\theta)\right) + \mathrm{E}\left(\mathrm{Var}(\bar{X}_j|\theta)\right) = \mathrm{Var}\left(\mu(\theta_j)\right) + \mathrm{E}(\mathrm{Var}(\bar{X}_j|\theta))$$
$$= \mathrm{Var}\left(\mu(\theta_j)\right) + \mathrm{E}\left(\sigma^2\,(\theta_j)\frac{1}{N}\right) = \mathrm{VHM} + \frac{\mathrm{EPV}}{\mathrm{N}}$$

$$\widehat{VHM} = Var(\bar{X}_j) + \frac{\mathrm{EPV}}{\mathrm{N}} = \left(\frac{1}{\mathrm{R}-1}\right)\sum_{j=1}^{R}(\bar{X}_j - \bar{X})^2 - \left[\frac{1}{R(N-1)}\sum_{j=1}^{R}\sum_{t=1}^{N}(X_{jt-}\bar{X}_j)^2\right]/N$$

När dessa är estimerade kan man beräkna det viktade medelvärdet:

$$\hat{\boldsymbol{\mu}}(\boldsymbol{\theta}_j) = \hat{\boldsymbol{Z}}\bar{\boldsymbol{X}}_j + (\mathbf{1} - \hat{\boldsymbol{Z}})\bar{\boldsymbol{X}}$$

$där \qquad \hat{K} = \frac{\widehat{EPV}}{\widehat{VHM}} \qquad \hat{Z} = \frac{N}{N+K}$

I tabell 6.1 nedan redovisas beräkningen av kredibilitetsskattningen av $\hat{\varphi}_I^{acc}$ och $\hat{\varphi}_N^{acc}$

Tabell 6.1: Kredibilitetsskattning av sannolikheten för CI eller plötslig död

Kön	ålder	$\bar{X}_1$	$\bar{X}_2$	$\widehat{EPV}$	$\widehat{VHM}$	$\hat{K}$	Z	$\bar{X}$	$\mu(\hat{\theta}_1)$	$\mu(\hat{\theta}_2)$
F	20-34	0,84%	0,09%	0,0012%	0,0026%	0,47	0,95	0,46%	0,82%	0,10%
	35-49	2,07%	0,24%	0,0145%	0,0132%	1,10	0,89	1,15%	1,97%	0,34%
	50-64	2,83%	0,65%	0,0513%	0,0111%	4,64	0,66	1,74%	2,46%	1,02%

Kön	ålder	$\bar{X}_1$	$\bar{X}_2$	$\widehat{EPV}$	EPV	$\hat{K}$	Z	$\hat{X}$	$\mu(\hat{\theta}_1)$	$\mu(\hat{\theta}_2)$
M	20-34	0,84%	0,13%	0,0017%	0,0021%	0,83	0,92	0,48%	0,81%	0,16%
	35-49	1,22%	0,24%	0,0062%	0,0033%	1,91	0,82	0,73%	1,13%	0,32%
	50-64	4,10%	0,66%	0,0306%	0,0515%	0,59	0,94	2,38%	4,00%	0,77%

Ovan ser vi de justerade estimaten för sannolikheten att drabbas av CI i den försäkrade portföljen respektive hela nationen. Vi kan konstatera att medelvärdet för sannolikheten att drabbas av CI i den försäkrade populationen för kvinnor i ålder 20-34 år $\hat{\varphi}_I^{acc}$ justeras ned från 0,84% till 0,82%.

Samma sannolikhet för motsvarande urval i den nationella populationen $\hat{\varphi}_N^{acc}$ justeras upp från 0,13 % till 0,16 %. Skillnaden mellan estimatet för populationen och estimatet för nationen är mycket stora. För att förtydliga skillnaden följer ett exempel i tabell 6.2 nedan. Antag att vi har en försäkrad population på 100 000 individer med estimatet skattat från den försäkrade populationen skulle vi få 823 skador för kvinnor i ålder 20-34 år och med estimatet skattad från den nationella populationen skulle det endast bli 104 skador. Denna skillnad är så stor att vi kan konstatera att det inte är någon lösning att använda nationellt estimat för att skatta sannolikheten i den försäkrade populationen.

Tabell 6.2. Skillnad i antalet skador med de olika estimaten räknat på 100 000 försäkrade individer

kön	**Ålder**	**Försäkrade populationen**	**Nationen**
F	20-34	823	104
	35-49	1 971	338
	50-64	2 461	1 018
kön	**Ålder**		
M	20-34	807	161
	35-49	1 131	321
	50-64	3 995	770

Orsaken till de olika resultaten kan vara flera. För det första så är antalet skador i försäkringsdatat för litet för att några säkra resultat ska kunna beräknas. En enda skada påverkar storleken på estimatet mycket eftersom det blir en stor skillnad på en eller två skador. Dessutom är det troligt så att det sker ett moturval i försäkringsportföljen. Det är inte ett slumpmässigt urval ur populationen som köper denna typ av försäkring. Individer som på ett eller annat sätt tror sig veta att de har en större risk än andra att drabbas av CI är mer benägen att en teckna försäkring. Underwriting har försökt begränsa detta genom olika metoder, t.ex. genom att begränsa försäkringssumman samt skriva ned den ytterligare ifall försäkringstagaren har ytterligare försäkringar. Åtgärderna verkar inte ha räckt hela vägen fram.

7 Prissätta Sverige

I Sverige är försäljningen av Critical Illness försäkring som nämnts mycket begränsad. Det finns ingen skadedata tillgänglig och den enda data som är tillgänglig är nationell data. Som vi tidigare har framgått i detta arbete kan det skilja sig mycket mellan sannolikheten att diagnostiseras med cancer för den försäkrade populationen och den nationella populationen.

7.1 Sannolikheten för skada i Sverige

Det finns inget skadedata för Sverige och därför får vi gå direkt till skattningar på nationell data. För en acceleration benefit är sannolikheten för skada lika sannolikheten att diagnostiseras med cancer eller att avlida i förtid. För additional benefit är sannolikheten för skada lika med risken för att diagnostiseras med cancer. Precis som för Storbritannien använder vi sats 3.1 för att beräkna dessa sannolikheter.

$$\varphi_N^{acc} = i_x + (q_x - k_x q_x) * \frac{l_x - l_x^s - I_x}{l_x - l_x^s}$$

$$\varphi_N^{add} = i_x$$

Vi gör samma antaganden som i avsnitt 4.2 för att approximera $l_x - l_x^s$ och samma åldersfördelning inom ålderspannen som vi gjorde för Storbritannien så estimaten blir jämförbara. I tabell 7.1 och 7.2 redovisas skattningen med respektive approximation.

Tabell 7.1. Skattning av $\hat{\varphi}_N^{acc}$ med approximationen $l_x - l_x^s = l_x$

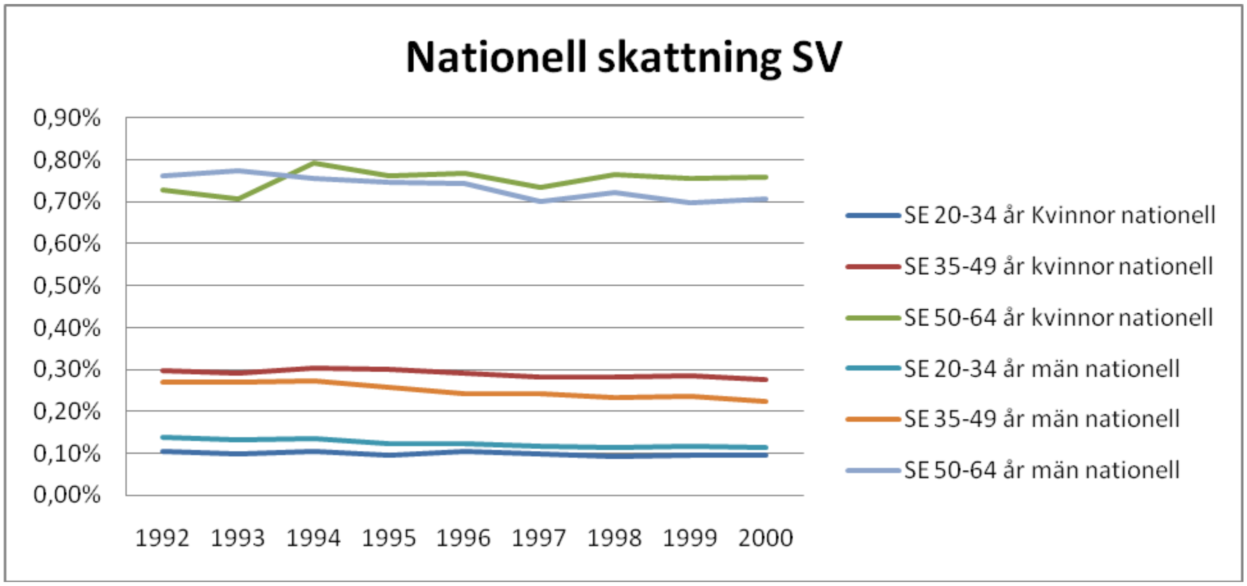
Kön	Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000
F	**20-34**	0,11%	0,10%	0,11%	0,10%	0,10%	0,10%	0,09%	0,09%	0,09%
F	**35-49**	0,30%	0,29%	0,30%	0,30%	0,29%	0,28%	0,28%	0,28%	0,28%
F	**50-64**	0,73%	0,71%	0,79%	0,76%	0,77%	0,73%	0,76%	0,76%	0,76%
M	**20-34**	0,14%	0,13%	0,13%	0,12%	0,12%	0,12%	0,12%	0,12%	0,11%
M	**35-49**	0,27%	0,27%	0,27%	0,26%	0,24%	0,24%	0,23%	0,24%	0,23%
M	**50-64**	0,76%	0,77%	0,75%	0,75%	0,74%	0,70%	0,72%	0,70%	0,71%

Tabell 7.2. Skattning av $\hat{\varphi}_N^{acc}$ med approximationen $l_x - l_x^s = l_x - 5 * I_x$

Kön	Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000
F	**20-34**	0,11%	0,10%	0,11%	0,10%	0,10%	0,10%	0,09%	0,09%	0,09%
F	**35-49**	0,30%	0,29%	0,30%	0,30%	0,29%	0,28%	0,28%	0,28%	0,28%
F	**50-64**	0,73%	0,71%	0,79%	0,76%	0,77%	0,73%	0,76%	0,76%	0,76%
M	**20-34**	0,14%	0,13%	0,13%	0,12%	0,12%	0,12%	0,12%	0,12%	0,11%
M	**35-49**	0,27%	0,27%	0,27%	0,26%	0,24%	0,24%	0,23%	0,24%	0,23%
M	**50-64**	0,76%	0,77%	0,75%	0,75%	0,74%	0,70%	0,72%	0,70%	0,71%

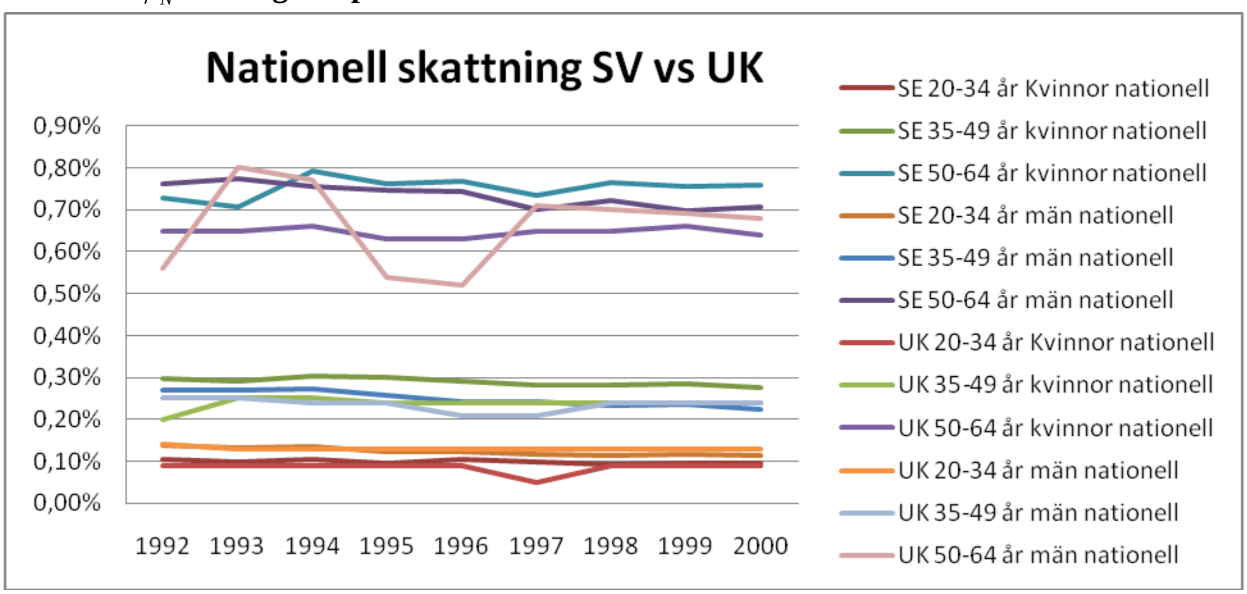
Precis som väntat är det ingen skillnad mellan approximationerna. Även om Sveriges population är betydligt mindre än Storbritanniens så är det fortfarande ett stort antal individer (1,8 miljoner) som är bas för beräkningen. Detta ger en mycket stabil skattning av $\hat{\varphi}_N^{acc}$, se graf 7.1

Graf 7.1: $\hat{\varphi}_N^{acc}$



När vi jämför de nationella estimaten för Sverige och Storbritannien ser vi att sannolikheterna är relativt lika för varje ålderskategori. Se graf 7.2 nedan.

Graf 7.2: $\hat{\varphi}_N^{acc}$ **Sverige respektive Storbritannien**



7.1.1 Förbättring av den nationella skattningen

Den ursprungliga tanken i detta arbete var att justera Sveriges nationella estimat med hänsyn till den eventuella skillnad som fanns i Storbritannien mellan försäkrat och nationellt estimat. Detta för att kompensera för misstanken att urvalet i den försäkrade populationen inte riktigt var jämförbart med nationen. Problemet är, som vi tidigare har konstaterat, att försäkringsestimatet i

Storbritannien är osäkert och att det skiljer sig så mycket från det nationella estimatet att det är svårt att dra en generell slutsats. I de resultat som vi tidigare har presenterat har vi sett att ålder är av mycket större vikt än kön. Blir resultatet något annat om vi tar bort parametern kön? Vi provar och ser om skattningarna blir något bättre när antalet observationer i de kvarvarnade grupperna ökar. Tabell 7.1.1 och 7.1.2 visar antal försäkrade individer respektive antal skador för både män och kvinnor.

Vi beräknar först sannolikheten för skada i den försäkrade populationen $\hat{\varphi}_I^{acc}$ i Storbritannien, se tabell 7.1.3.

Tabell 7.1.1 Antal försäkrade individer

Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000	summa
20-34	1 023	856	1 140	1 031	3 003	2 872	2 777	2 427	1 532	16 661
35-49	1 135	983	1 241	1 308	4 891	4 457	3 787	2 911	2 116	22 829
50-64	70	74	122	196	853	788	768	691	722	4 284
	2 228	1 913	2 503	2 535	8 747	8 117	7 332	6 029	4 370	43 774

Tabell 7.1.2 Antal skador

Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000	summa
20-34	12	16	8	8	4	24	14	16	15	117
35-49	8	2	2	22	28	52	68	95	75	352
50-64	0	4	2	6	14	36	42	53	37	194
	20	22	12	36	46	112	124	164	127	663

Tabell 7.1.3 Skattning av $\hat{\varphi}_I^{acc}$

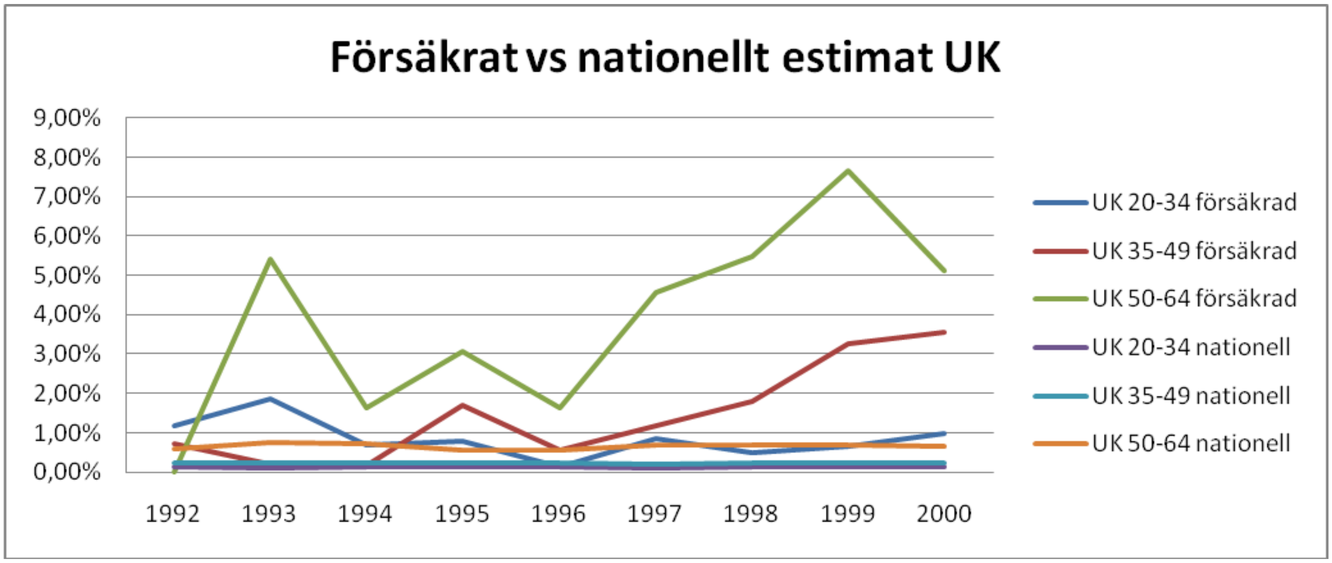
Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000
20-34	1,17%	1,87%	0,70%	0,78%	0,13%	0,84%	0,50%	0,66%	0,98%
35-49	0,70%	0,20%	0,16%	1,68%	0,57%	1,17%	1,80%	3,26%	3,54%
50-64	0,00%	5,41%	1,64%	3,06%	1,64%	4,57%	5,47%	7,67%	5,12%

Vi tittar på könsfördelningen i portföljen och gör antagandet att 30 % i portföljen är kvinnor och resten är män. Vi tillämpar fördelningen vid beräkningen av populationssannolikheterna och beräknar sannolikheten för skada i den nationella populationen precis som i avsnitt 4.2, se tabell 7.1.4. I graf 7.1.1 nedan ser vi sannolikheten för skada både för nationell och försäkrad population. Vi ser att sannolikheten i den nationella populationen är betydligt lägre än försäkrad population.

Tabell 7.1.4 Skattning av $\hat{\varphi}_N^{acc}$

Ålder	1992	1993	1994	1995	1996	1997	1998	1999	2000
20-34	0,12%	0,12%	0,12%	0,12%	0,12%	0,10%	0,12%	0,12%	0,12%
35-49	0,24%	0,25%	0,25%	0,24%	0,22%	0,22%	0,24%	0,24%	0,24%
50-64	0,59%	0,75%	0,74%	0,57%	0,55%	0,70%	0,68%	0,68%	0,67%

Graf 7.1.1: Nationell resp. försäkrad skattning av $\hat{\varphi}^{acc}$



Vi väger ihop dessa skattningar precis som tidigare i avsnitt 6.1. och vi väljer att endast använda de fem sista åren då försäkringsportföljen är större för att få ett stabilare estimat.

Tabell 7.1.5: Kredibilitetsskattning av sannolikheten för CI eller plötslig död

Ålder	$\bar{X}_1$	$\bar{X}_2$	$\widehat{EPV}$	$\widehat{VHM}$	$\hat{K}$	Z	$\bar{\mathbf{X}}$	$\mu(\hat{\theta}_1)$	$\mu(\hat{\theta}_2)$
20-34	0,62%	0,12%	0,0005%	0,0010%	0,53	0,90	0,37%	0,60%	0,14%
35-49	2,07%	0,23%	0,0084%	0,0127%	0,66	0,88	1,15%	1,96%	0,34%
50-64	4,89%	0,66%	0,0235%	0,0781%	0,30	0,94	2,78%	4,77%	0,78%

I tabell 7.1.5 kan utläsas att skattningarna närmar sig varandra men att de fortfarande är oanvändbara. Om vi gör samma exempel som tidigare med 100 000 försäkrade individer blir konsekvenserna mycket tydliga, vilket framgår i tabell 7.1.6 nedan. Att ha över 7000 skador jämfört med drygt 1000 skador betyder att priset skulle behöva vara sju gånger högre. Vi konstaterar att detta inte är en lämplig metod för att prissättning av Critical Illness portföljen, varken för Sverige eller Storbritannien.

Tabell 7.1.6. Skillnad i antalet skador med de olika estimaten räknat på 100.000 försäkrade individer

Ålder	**Försäkrade populationen**	**Nationen**
20-34	598	142
35-49	1 961	336
50-64	4 774	777
Summa	7 333	1 255

8 Diskussion

8.1 Metoddiskussion

Den befintliga Critical Illness portföljen är för lågt prissatt och har lönsamhetsproblem. Syftet med detta arbete var att ta fram en modell för att bestämma priset för en Critical Illness försäkring i Storbritannien och Sverige. De sökta sannolikheterna är mycket små vilket gör att antalet försäkringar i försäkringsportföljen måste vara mycket stort för det ska vara möjligt att beräkna säkra estimat baserat på skadedata. Den befintliga försäkringsportföljen är mycket liten därför var tanken att pröva om det går att kombinera estimaten från portföljen med estimat baserat på hela nationen. Med hjälp av kredibilitetsteori vägs de osäkra skattningarna från försäkringsdatat samman med de stabila nationella skattningarna för att på så sätt få en bra prissättningsmodell. Det visade sig dock att denna metod inte var lämplig.

Det största skälet till att metoden inte är användbar är att sannolikheten att drabbas av skada i den försäkrade populationen är mellan 5 till 9 gånger högre än motsvarande sannolikheten i den nationella populationen. Den nationella statistiken är inte ett bra substitut för den försäkrade populationen. Orsaken till detta kallas moturval och innebär att den försäkrade populationen inte är ett slumpmässigt urval ur den nationella befolkningen. Uppenbarligen har individerna som tecknar den här typen av försäkringen en ökad risk att drabbas av sjukdom och förtidig död jämfört med befolkningen i övrigt och är därför mer benägna att försäkra sig. De bakomliggande orsakerna till cancer är svåra att fånga upp i livsstilsfaktorer och därför är informationsövertaget hos kunderna svårt att komma till rätta med och fånga upp i en modell.

En annan betraktelse är att som vi tidigare har nämnt tar kredibilitetsskattningen hänsyn till dels hur skattningarna inom riskklassen skiljer sig mycket åt och dels hur stor spridning det är inom varje enskild risk. Eftersom sannolikheten för skada eller tidig död i nationen $\hat{\varphi}_N^{acc}$ baserar sig på en stor datamängd skiljer sig skattningarna mellan åren mycket lite. Variansen av de enskilda utfallen från deras förväntade värde blir därmed mycket liten. Även om de två riskerna i riskgruppen, den baserad på nationen och den baserad på försäkringsdata är olika så blir kredibilitetsfaktorn hög eftersom de enskilda riskernas varians är låg och skattningarna ändras inte så mycket i förhållande till deras genomsnitt.

8.2 Jämförelse med befintlig rate

Eftersom vi vet att den befintliga Critical Illness portföljen är för lågt prissatt kan det vara intressant att jämför befintlig rate med de resultat som har beräknats. I tabellen 8.2.1 nedan ser vi den befintliga raten för män i några olika åldersklasser. Vi ska dock ta i beaktande att den inkluderar fler Critical Illness diagnoser än cancer samt övriga kostnader. En jämförelse är därför inte helt enkelt. Låt oss dock göra ett försök: CI cancer utgör ca hälften av de skador som försäkringen täcker. Vi gör antagandet att vi får dubbelt antal skador och dubblerar sannolikheten för skada om vi inkluderar alla CI. Vi kan se att skattningen av $\hat{\varphi}_N^{acc}$ baserad på nationell data är relativt i linje med befintlig rate om man tar med ett kostnadspåslag på ca 20 %. Detta resultat är att vänta eftersom befintlig rate är baserad på nationell data.

Tabell 8.2.1: Jämförelse mellan befintlig rate och estimerad rate för män

Befintlig rate för män

ålder	sannolikhet
20	0,14%
25	0,15%
30	0,21%
35	0,32%
40	0,52%
45	0,85%
50	1,30%
55	1,99%
60	2,84%

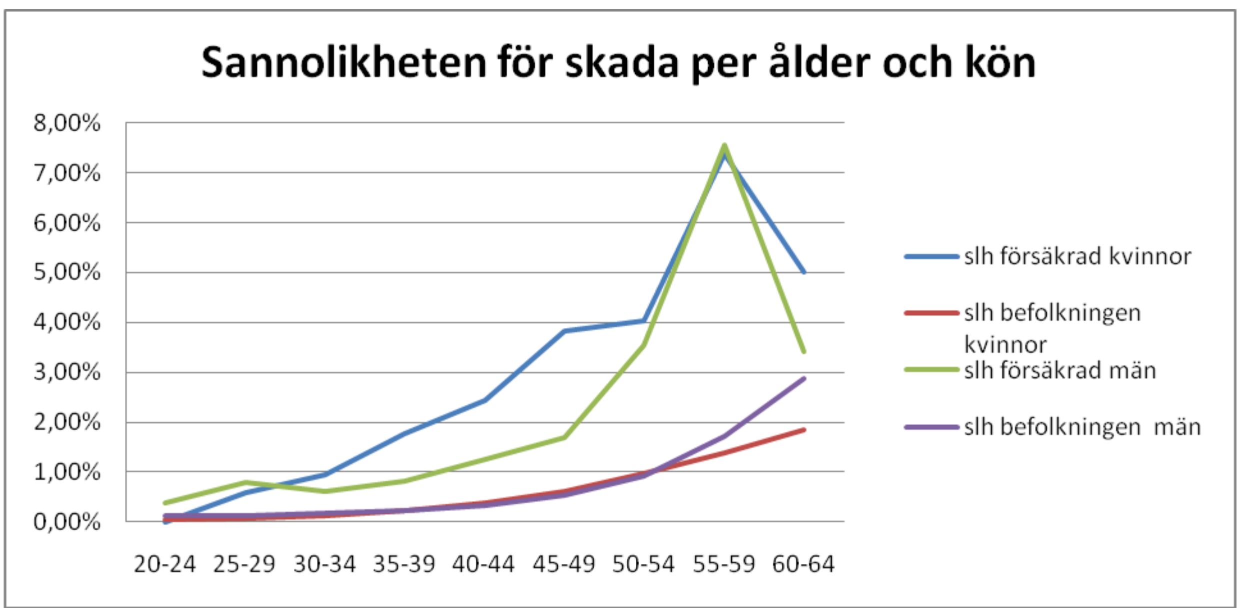
$\hat{\varphi}^{acc}$ x_1= skadedata och x_2= nationsdata multiplicerat med 2

ålder	$\bar{X}_1$	$\bar{X}_2$
20-34	1,67%	0,26%
35-49	2,43%	0,47%
50-64	8,20%	1,33%

8.3 Förslag till alternativ metod

Ansatsen i detta arbete var att kombinera estimaten baserat på den lilla men dock empiriska försäkringsportföljen med stabila estimat från hela befolkningen. Tyvärr visade det sig att den givna modellen inte var tillämpningsbar som vi har diskuterat i avsnittet ovan. Problemet kvarstår dock befintlig rate och raten baserad på $\hat{\varphi}_N^{acc}$ är för låg för att given portfölj ska vara lönsam, samtidigt är en rate baserad på $\hat{\varphi}_I^{acc}$ för osäker för att ligga till grund för en ny rate. I ett försök att gå vidare och göra ett förslag till ny modell börjar vi med att plotta sannolikheten för skada i den nationella population och i den försäkrade populationen per ålder indelat i 5 års intervall och ser om vi kan se något mönster, graf 8.3.1 nedan. Eftersom vi inte har någon data för Sverige väljer vi att enbart göra en alternativ modell för Storbritannien baserat på det data vi har tillgängligt.

Graf 8.3.1: Nationell resp. försäkrad skattning av $\hat{\varphi}^{acc}$ indelat per ålders och kön



För den nationella populationen är det ingen skillnad mellan könen att drabbas av cancer fram till 50 års ålder därefter ökar sannolikheten i snabbare takt för män. Det är dock en betydande skillnad mellan män och kvinnor i den försäkrade populationen där sannolikheten för kvinnor är generellt högre än för män i ålder 30 till 50 år. Det är därför lämpligt att göra separata modeller för könen eftersom vi främst är intresserade av att ta fram en modell som omfattar ålder 30 till 50 år då det är störst försäljning i detta intervall. Vi noterar att sannolikheten för kvinnor i den försäkrade populationen och nationella populationen är mer linjär medan motsvarande sannolikheter för män bättre beskrivs av en exponentiell funktion. Vi önskar att se om vi kan genom att finna en funktion för den försäkrade sannolikheten av värdena från den nationella populationen vid varje ålder och därmed få en mer jämn och tillförlitlig rate.

8.3.1 Män

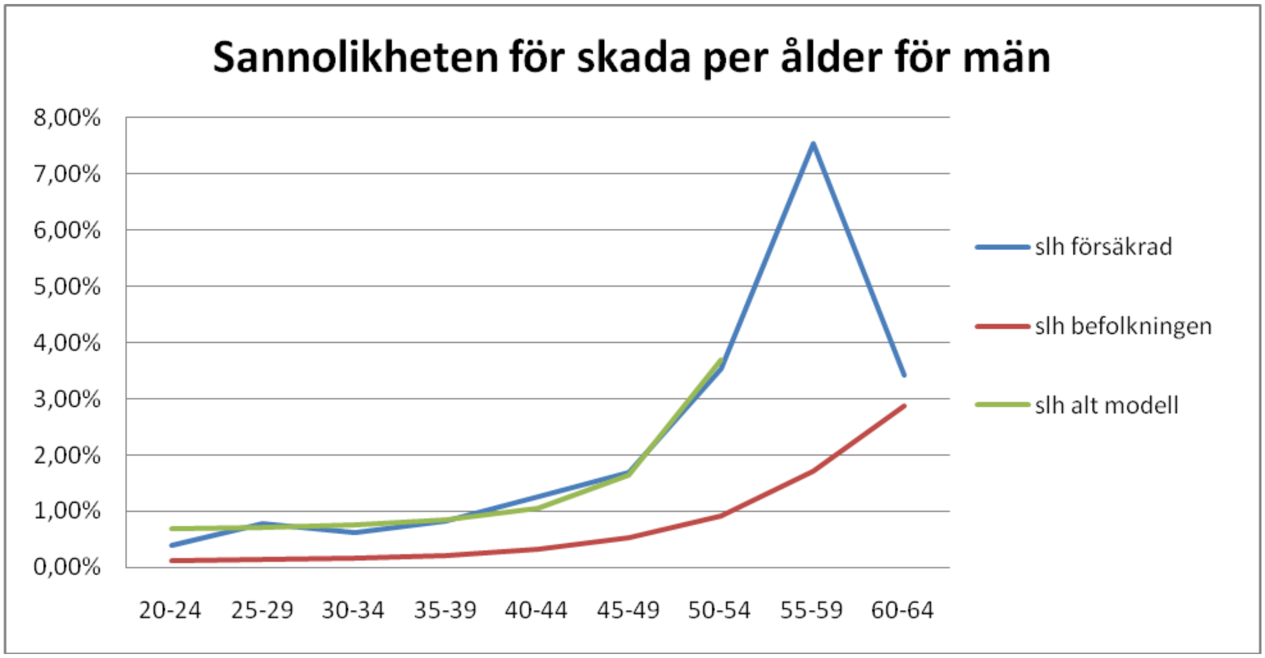
Vi noterade ovan att sambandet för män inte är linjärt. Vi transformerar därför anpassningsfunktionen så den blir linjära i parametrarna genom att logaritmera datapunkterna. I tabellen 8.3.1 nedan kan vi konstatera att antalet försäkringar i den lägsta ålderklassen 20-24 år samt i de två äldsta åldersklasserna 55-59 och 60-64 år är få vilket ökar osäkerheten i estimaten. Vi väljer därför att exkludera dessa data i regressionen då vi önskar att främst finna en god skattning i intervallet 30 till 50 år.

Tabell 8.3.1: Män

ålder	antal skador	exponering	slh försäkrad $\hat{\varphi}_I^{acc}$	slh befolkningen $\hat{\varphi}_N^{acc}$	slh alt modell $\hat{\varphi}_M^{acc}$
20-24	2	513	0,39 %	0,12 %	0,69 %
25-29	31	3963	0,78 %	0,14 %	0,72 %
30-34	48	7841	0,61 %	0,17 %	0,76 %
35-39	63	7613	0,83 %	0,22 %	0,85 %
40-44	71	5612	1,27 %	0,32 %	1,06 %
45-49	62	3686	1,68 %	0,53 %	1,64 %
50-54	80	2266	3,53 %	0,91 %	3,69 %
55-59	47	623	7,54 %	1,71 %	19,65 %
60-64	3	88	3,41 %	2,87 %	226,50 %

Vi gör en enkel linjär regression och finner att funktionen som bäst passar till vårt data för sannolikheten att drabbas av cancer eller för tidig död i den försäkrade populationen i åldern 25-54 år är $\hat{\varphi}_M^{acc} = 0{,}0053857 * \exp(210{,}3 * \hat{\varphi}_N^{acc})$. Efter 54 års ålder är denna funktion inte lämplig då sannolikheten för skada är för hög. I intervallet 20-24 ger modellen en högre sannolikhet än det empiriska datat som i detta intervall är mycket osäkert. Grafen 8.3.2 nedan visar den skattande sannolikheten för män att drabbas av cancer eller dö en förtidig död i den försäkrade populationen, befolkningen samt den alternativa modellen.

Graf 8.3.2: Sannolikheten för skada $\hat{\varphi}^{acc}$ för nationell och försäkrad befolkning samt den alternativa modellen



8.3.2 Kvinnor

I tabellen 8.3.2 nedan kan vi konstatera att antalet skador i den lägsta ålderklassen 20-24 år är noll samt att antalet försäkringar är få. I ålderklasserna 50 år och äldre är det också få försäkringar och vi väljer därför att exkludera dessa data i regressionen.

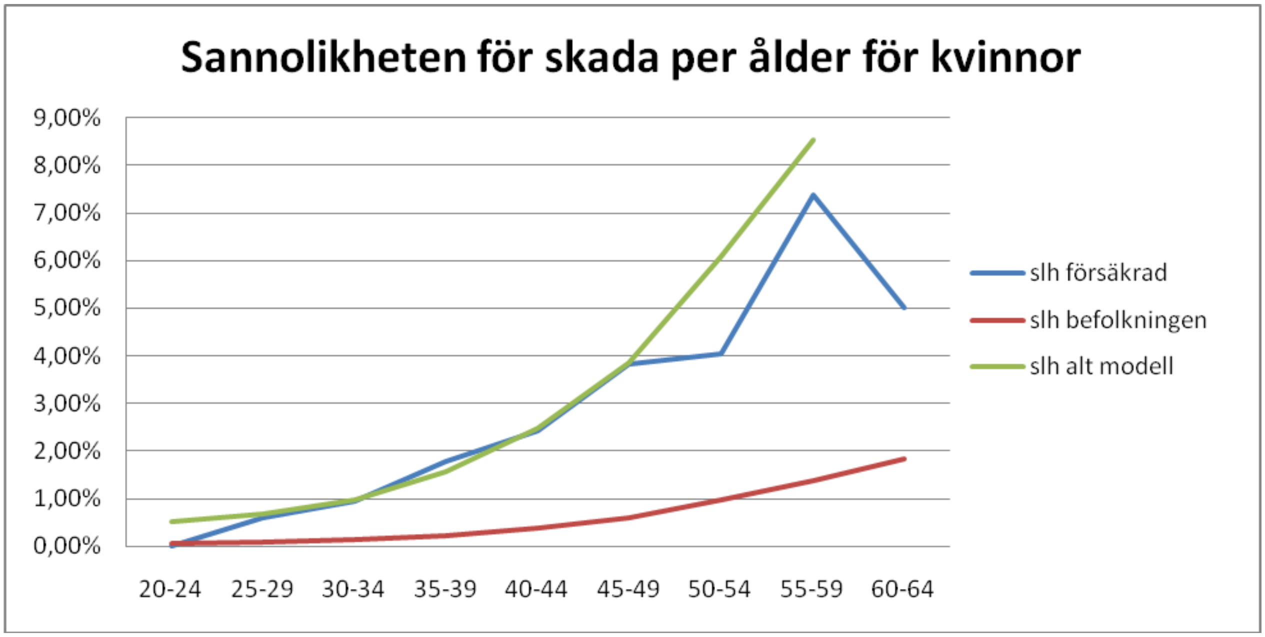
Tabell 8.3.2: Kvinnor

ålder	antal skador	exponering	slh försäkrad $\hat{\varphi}_I^{acc}$	slh befolkningen $\hat{\varphi}_N^{acc}$	slh alt modell $\hat{\varphi}_M^{acc}$
20-24	0	254	0,00 %	0,06 %	0,53 %
25-29	10	1681	0,59 %	0,08 %	**0,69 %**
30-34	26	2762	0,94 %	0,13 %	**0,97 %**
35-39	49	2759	1,78 %	0,23 %	**1,56 %**
40-44	50	2059	2,43 %	0,38 %	**2,47 %**
45-49	57	1488	3,83 %	0,61 %	**3,85 %**
50-54	40	992	4,03 %	0,98 %	6,09 %
55-59	22	298	7,38 %	1,39 %	8,53 %
60-64	2	40	5,00 %	1,85 %	11,30 %

Vi försöker alltså att fina en linjär funktion som beskriver försäkrad data i intervallet 25 till 49 år med hjälp av våra datapunkter från det nationella datat.

Funktionen $\hat{\varphi}_M^{acc} = 0{,}0019 + 6{,}0231\hat{\varphi}_N^{acc}$ ger en god anpassning. Däremot är funktionen inte lämplig att använda som modell för kvinnor över 50 år då sannolikheten blir alldeles för hög. Grafen 8.3.3 nedan visar den skattande sannolikheten för kvinnor att drabbas av cancer eller dö en förtidig död i den försäkrade populationen, befolkningen samt den alternativa modellen.

Graf 8.3.3: Sannolikheten för skada $\hat{\varphi}^{acc}$ för nationell och försäkrad befolkning samt den alternativa modellen



9 Slutsats

Syftet med detta arbete var att ta fram en modell för att bestämma priset för en Critical Illness försäkring i Storbritannien och Sverige. Metoden som prövats i arbetet för att uppnå detta, dvs. att kombinera osäkra skattningar baserat på ett begränsat skadedata, med mer stabila skattningar från hela nationens population inte är tillämpbar. Det visade sig att sannolikheten att drabbas av skada i den försäkrade populationen är mellan 5 till 9 gånger högre än motsvarande sannolikheten i den nationella populationen. Produkten har ett tydligt moturval. För cancer är det svårt att genom livsstilsfaktorer säga vilka individer som löper en ökad risk att diagnostiseras med cancer och därför är det svårt att i en modell särskilja dessa individer. En annan orsak till att den föreslagna modellen inte är lämplig är på grund att kredibilitetsteorin. Kredibilitetsteorin tar hänsyn till spridningen både inom en observation och mellan observationer. Den ena av våra två observationer består av observationer från en hel nation vilket gör att variansen inom stickprovet är mycket liten. Efter att ha konstaterat att den prövade modellen inte är ett bra alternativ till prissättning av Critical Illness försäkringen så kvarstår problemet. Befintlig rate och raten baserad på nationsdata är för låg och raten baserad på befintlig portfölj är för osäker för att tillämpa. För

Storbritannien prövar vi därför en alternativ modell som är en funktion av nationsdatat anpassat till portföljnivån för att få en jämnare och stabilare rate. För män finner vi en funktion som ger ett bra estimat i intervallet 25 till 54 år och för kvinnor finner vi motsvarande funktion som är tillämpningsbar i åldersintervallet 25 till 50 år. Vi nöjer oss med dessa resultat då det är i dessa intervall som det tecknas flest försäkringar. Vi kan avrunda med några ord om raten för Sverige. Vi kan konstatera att i graf 7.2 skiljer sig sannolikheten för skada per kön i åldersintervallet 25 till 50 år inte nämnvärt mellan Storbritannien och Sverige. Det som är avgörande är hur moturvalet är i Sverige. Detta går inte att mäta innan det finns någon skadedata att jämför med. Därför är det rimligt att som en utgångspunkt använda samma pris i Sverige som i Storbritannien och följa skadeutvecklingen i Sverige noggrant och se om den avviker från utvecklingen i Storbritannien.

10 Referenslista

Dash, Alison and Grimshaw, David: Dread Disease cover An Actuarial Perspective
Dean, Curtis Gary: Topics in Credibility Theory
Goovearts, M.J and Hoogstad W.J: Credibility Theory

Eurostat http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database
Statistiska centralbyrån http://www.scb.se
UK National statistics http://www.statistics.gov.uk

11 Appendix

Bühlmanns kredibilitetsskattning

Antag att vi har N stycken upprepade observationer av en risks j skadedata $X_1, X_2, ..., X_N$. Varje utfall har samma väntevärde $E_{X|\vartheta}[X_t|\vartheta = \theta] = \mu(\theta)$ där θ är riskparametern till den valda risken. Enligt förutsättningarna är de stokastiska variablerna oberoende och lika fördelade och väntevärdet för en risk är $\bar{X} = \left(\frac{1}{N}\right)\sum_{t=1}^{N} X_t$ och väntevärdet för hela populationen som kan ha olika riskparametrar är
$E(\bar{X}) = E_\vartheta\left[E_{X|\vartheta}[X|\vartheta]\right] = E_\vartheta[\mu(\theta)] = \mu$

Vi söker ett linjärt estimat av $\mu(\theta)$ och använder utfallet av $\bar{X}$: $\hat{\mu}(\theta) = a + b\bar{X}$
Vi använder minsta kvadraten metoden för att finna ett estimat av $\hat{\mu}(\theta)$

Minimera $l(g) = E\left[(\hat{\mu}(\theta) - \mu(\theta))^2 = E[(a + b\bar{X} - \mu(\theta))^2]\right]$

$$\begin{aligned}
E[(a + b\bar{X} - \mu(\theta))^2] &= E[(a + b\bar{X} - b\mu(\theta) + b\mu(\theta) - \mu(\theta))^2] \\
&= E\left[\left(b\left(\bar{X} - \mu(\theta)\right) + a - (1-b)\mu(\theta)\right)^2\right] \\
&= E\left[b^2\left(\bar{X} - \mu(\theta)\right)^2 + 2b\left(\bar{X} - \mu(\theta)\right)(a - (1-b)\mu(\theta)) + (a - (1-b)\mu(\theta))^2\right] \\
&= E\left[b^2\left(\bar{X} - \mu(\theta)\right)^2\right] + 2E\left[b\left(\bar{X} - \mu(\theta)\right)(a - (1-b)\mu(\theta))\right] \\
&+ E[(a - (1-b)\mu(\theta))^2]
\end{aligned}$$

Termen i mitten är lika med noll:
$E\left[b\left(\bar{X} - \mu(\theta)\right)(a - (1-b)\mu(\theta))\right] = E_\vartheta\left[E_{X|\vartheta}\left[b(a - (1-b)\mu(\theta))\left(\bar{X} - \mu(\theta)\right)|\vartheta\right]\right] =$
$E\left[b(a - (1-b)\mu(\theta))E_{X|\vartheta}\left[\left(\bar{X} - \mu(\theta)\right)|\vartheta\right]\right] = 0$

Eftersom $E_{X|\vartheta}[X_t|\vartheta] = \mu(\theta)$

Kvar är:
$E[(a + b\bar{X} - \mu(\theta))^2] = b^2 E\left[\left(\bar{X} - \mu(\theta)\right)^2\right] + E[(a - (1-b)\mu(\theta))^2]$

Där endast den andra termen innehåller a. Vilket värde på a minimerar termen?

$$\begin{aligned}
E[(a - (1-b)\mu(\theta))^2] &= E[a^2 - 2a(1-b)\mu(\theta) + (1-b)^2\mu^2(\theta)] \\
&= a^2 - 2a(1-b)E\left[\mu(\theta) + (1-b)^2 E[\mu^2(\theta)]\right]
\end{aligned}$$

Derivera med avseende på a och sätt uttrycket lika med noll och ersätt $E[\mu(\theta)]$ med μ vilket ger $a = (1-b)\mu$

Ersätt a med ovanstående uttryck vilket ger:

$$E[(a+b\bar{X}-\mu(\theta))^2] = b^2 E\left[\left(\bar{X}-\mu(\theta)\right)^2\right] + (1-b)^2 E[(\mu(\theta)-\mu)^2]$$
$$= b^2\frac{EPV}{N} + (1-b)^2 VHM$$

Derivera med avseende på b och sätt uttrycket lika med noll:

$$2b\frac{EPV}{N} - 2(1-b)VHM = 0$$

Vilket ger $b = \frac{VHM}{VHM+\frac{EPV}{N}} = \frac{N}{N+\frac{EPV}{VHM}} = \frac{N}{N+K}$ där $K = \frac{EPV}{VHM}$

Sätt in uttrycken i estimatet av $\mu(\theta)$

$$\hat{\mu}(\theta) = a + b\bar{X} = \left(\frac{N}{N+K}\right)\bar{X} + \left(1-\frac{N}{N+K}\right)\mu$$

Vilket ger den sökta formen.