

INITIAL SELECTION AND CAUSE OF DISABILITY FOR INDIVIDUAL PERMANENT HEALTH INSURANCE

BY

MARÍA CRISTINA GUTIÉRREZ-DELGADO AND ATHOL A. KORABINSKI

Department of Actuarial Mathematics and Statistics, Heriot-Watt University

ABSTRACT

We investigate the influence of initial selection (the impact of underwriting during the early years of a policy's life) on individual Permanent Health Insurance claim inceptions. In Gutiérrez-Delgado (1999) a decreasing trend was found. In this paper we include the effect of cause of disability and fit a generalized linear model in order to gain a greater understanding of the phenomenon. Both effects, policy duration and cause of disability, are found to have a significant effect on the number of claims. We describe their influence using factors that collect the information available through the fitted model. Results from both factors suggest that the grouping of diseases selected for the research helps to explain partially our earlier results. In addition there is some evidence of moral hazard in mental disorders and musculoskeletal diseases which also contributes to the understanding of the negative trend found.

KEYWORDS

Permanent Health Insurance; Disability Income Insurance; Income Protection Insurance; Claim Inceptions; Initial Selection; Cause of Disability, Generalized Linear Models; Moral Hazard.

1. INTRODUCTION

Individual Permanent Health Insurance (PHI) is a product available in the U.K. market to cover income protection demands for long-term disability. It provides regular payments, subject to a ceiling based on a percentage of salary, to the policy-holder when he/she is in a disability episode that lasts beyond a pre-defined deferred period. Premiums are generally level and payable while benefit is not being paid. The contract usually ceases at normal retirement age: 65 for males and 60 for females.

The initial selection effect is related to the time a policy has been in force. It might be expected that under a robust underwriting process there will be lower

claim rates in the first few years of coverage due to medical checks when the policy is effected. Nevertheless after an initial analysis and modelling process of the individual PHI data available we found a negative trend (see Gutiérrez-Delgado (1999)). This trend, that was also reported in other experiences, without major analysis, by Lundberg (1969), Miller and Courant (1974), C.M.I.B. (1996) and the A.D.C. (1998), may indicate an important degree of moral hazard in some diseases, as suspected by underwriters in the market.

We decided to add at this stage the cause of disability, after a careful analysis of the possible influences relating to the negative trend and the data at hand. We expect that the resulting analysis will allow us to gain a more in-depth knowledge of the phenomenon.

Therefore our main aim in this paper is to examine the influence of initial selection and cause of disability on the actual claim inceptions in the presence of the effects from deferred period, sex, year and age.

We consider that the results of this research will be of interest in other disability income insurance markets (also called income protection insurance) outside the U.K. although our numerical results come from U.K. data.

2. THE CAUSE OF DISABILITY

In our work we group the causes of disability available (see section 3.1) to respond to two main objectives. Firstly to gain statistical robustness by reducing the large number of causes to a manageable number. Secondly, and more importantly, to help us to answer the question about the extent of the effect of some of the causes on the negative trend found before.

The grouping of causes of disability we used was derived from the remarks of some underwriters and consultants with extensive experience in the market. Comments related to a possible moral hazard effect in some diseases and the impossibility of detecting certain others during the underwriting process given their sudden aleatory nature were incorporated in the definition of the groups. The five groups structure we selected is presented in Table 1.

TABLE I
GROUP OF DISEASES

<i>No.</i>	<i>Group of Diseases</i>	<i>No. of Causes from ICD8</i>
1	Musculoskeletal Diseases (excluding arthritis-spondilitis)	1
2	Mental Disorders	2
3	Infectious Diseases	27
4	All Other Diseases	36
5	Accidents and Injuries	5

Musculoskeletal diseases (back pains) and mental disorders, the first and second groups respectively, are considered as subjective kinds of diseases which allow a possibility of moral hazard. Moral hazard in this case is the risk that policy-holders commit fraud concerning their symptoms to make a claim and prolong the disability period because there is no definitive medical test that can reliably prove non-illness or recovery at present. This phenomenon is called the "will to work". According to underwriters' experience, the will to work is highly influenced by economic conditions, as widely commented in Soule (1994). Therefore these two groups of diseases are of particular interest to actuaries and underwriters. In addition both groups are very frequent causes of disability among workers as reported by Riihimaki, Kasl and Amick in McDonald (1995). We confirm this situation with the statistics presented in section 3.2.

The fourth group includes all other diseases that can be identified through medical history and medical examination during the underwriting process. This group incorporates half of the diseases including a high proportion of the chronic-degenerative illnesses. Some examples of these diseases are neoplasms, arthritis-pondylitis, diabetes mellitus and heart diseases. Because of its definition we expect that group four will be of importance for the older ages and longer policy durations among all deferred periods.

The third group (infectious diseases) and the fifth group (accidents and injuries) involve causes of disability that can not be detected during the underwriting process and because of their sudden aleatory nature are not associated with moral hazard. Some characteristics of the diseases in the third group are the short term period for reaching recovery, the seasonality of most of them and the possibility of epidemics. Therefore we expect that group three will be important in shorter deferred periods. Group five is expected to be important among young ages specially because of road traffic accidents. It could be argued that infectious diseases can be detected through medical tests. Nevertheless in practice policy-holders who are in an episode of any infectious disease delay their application to avoid being rejected or requested to present extra medical tests or be subject to extra-premiums and/or exclusions.

Figure 1 presents the mapping of the ICD8 code to the grouping we use.

<i>ICD8</i>	<i>Description</i>	<i>Group</i>	<i>ICD8</i>	<i>Description</i>	<i>Group</i>
1	Typhoid, other salmonella infections	3	37	Venous thrombosis and embolism	4
2	Bacillary dysentery and amoebiasis	3	38	Other diseases of the circulatory system	4
3	Enteritis and other diarrhoeal	3	39	Acute respiratory infections	3
4	Tuberculosis of respiratory system	3	40	Influenza	3
5	Other tuberculosis, including late effects	3	41	Pneumonia	3
6	Brucellosis	3	42	Bronchitis, emphysema and asthma	3
7	Diphtheria	3	43	Hypertrophy of tonsils and adenoids	3
8	Whooping Cough	3	44	Pneumoconioses and related diseases	3
9	Streptococcal sore throat and scarlet fever	3	45	Other diseases of the respiratory system	3
10	Smallpox	3	46	Diseases of teeth and supporting structures	4
11	Measles	3	47	Peptic ulcer	4
12	Viral Encephalitis	3	48	Appendicitis	4
13	Infectious hepatitis	3	49	Intestinal Obstruction and Hernia	4
14	Typhus and other rickettsioses	3	50	Cholelithiasis and cholecystitis	4
15	Malaria	3	51	Other diseases of the digestive system	4
16	Syphilis and its sequelae	3	52	Nephritis and Nephrosis	4
17	Gonococcal infections	3	53	Calculus of the urinary system	4
18	Helminthiasis	3	54	Hyperplasia of prostate	4
19	All other infective and parasitic diseases	3	55	Other diseases of the genito-urinary system	4
20	Malignant neoplasms	4	56	Abortion	4
21	Benign neoplasms	4	57	Other complications of pregnancy, childbirth	4
22	Thyrotoxicosis with or without goitre	4	58	Delivery without mention of complication	4
23	Diabetes mellitus	4	59	Infections of skin and subcutaneous tissue	3
24	Avitaminoses and other nutritional deficiency	4	60	Other diseases of skin and subcutaneous tissue	4
25	Other endocrine and metabolic diseases	4	61	Arthritis and spondylitis	4
26	Anaemias	4	62	Other diseases of the musculoskeletal system	1
27	Psychoses and non-psychotic mental disorders	2	63	Congenital anomalies	4
28	Inflammatory diseases of the eye	4	64	Certain causes of perinatal morbidity	4
29	Cataract	4	65	Other specified and ill-defined diseases	4
30	Otitis media and mastoiditis	4	66	Road transport accidents	5
31	Other diseases of nervous system	2	67	All other accidents	5
32	Active rheumatic fever	4	68	Attempted Suicide and self-inflicted injuries	5
33	Chronic rheumatic heart disease	4	69	Attempted homicide – injury by other persons	5
34	Hypertensive disease	4	70	All other external causes	5
35	Ischaemic heart disease	4	76	Myalgic Enceph. (Chronic Fatigue Syndrome)	2
36	Cerebrovascular disease	4	77	Aids and HIV	NA

The codes 76 and 77 were added later to the original list (published in 1967) because none of them were recognised by the WHO before the early 80's. In practice cause 77 is excluded of the coverage.

There are two other codes of diseases not included in this list. Both are related to unknown cause of disease (ICD8 codes 0 and 80). Actual inceptions ($A_{d,t,s,v,d,j}$) in these cases are proportionally distributed among the other categories in the grouping proposed by the underwriters.

FIGURE 1: List "C" of the International Statistical Classification of Diseases, Injuries and Causes of Death Version 8 and our Groups of Diseases.

3. INITIAL ANALYSIS OF DATA

3.1. The Data Available

The Continuous Mortality Investigation Bureau (CMIB) supplied us with their individual PHI standard experience data for 1987 to 1994. These data incorporate class I occupations only, excluding policies with occupational and/or medical ratings/exclusions¹. (Further details can be found in Kluwer (1998) and C.M.I.B. (1984).)

Our data were classified by deferred period ($d = 1, 4, 13, 26$ or 52 weeks); duration since the policy was effected ($t = 0, 1$ or 2 or more years); sex ($s = 1$ -Males, 2 -Females); age ($x = 17$ to 66 inclusive); calendar year ($= 1987$ to 1994 inclusive) and cause of disability ($i = 72$ possible causes from the international classification of diseases version 8 (ICD8)). Age (x) is classified by age nearest birthday at the start of the calendar year. Duration (t) is the curtate duration at the end of the calendar year.

For each combination of these effects we were provided with values for the actual number of claim inceptions ($A_{d,t,s,x,y,i}$) as well as the number of days spent claiming ($C_{d,t,s,x,y,i}$). In addition we were given the number of policies in force at the start ($SIF_{d,t,s,x,y}$) and at the end ($EIF_{d,t,s,x,y}$) of each calendar year.

Given the characteristics of the policy-holders and the exclusions imposed by the product, not all ages or causes of disability are present in some deferred periods and policy durations. Therefore our dataset is composed only of 44,342 actual claim inceptions for the analysis.

The scarcity or non-existence of data for several ages imposed on us the need to group our range of 50 individual ages into smaller groups to gain statistical robustness. We decided to work with a structure of seven groups of ages as shown in Table 2.

TABLE 2
GROUPS OF AGES

<i>Group</i>	<i>Ages</i>
1	18 to 29
2	30 to 34
3	35 to 39
4	40 to 44
5	45 to 49
6	50 to 54
7	55 to 66

¹ Occupations are white collar workers or other similar non-hazardous occupations mainly. Standard exclusions are AIDS, pregnancy, activities of war, self injuries or attempt of suicide, drugs abuse and alcoholism. Common ratings/exclusions are hazardous pastimes or sports, aerial activity other than as a fare paying passenger, criminal acts and failure to seek or follow medical advice.

3.2. Exploratory Data Analysis

We present basic statistics for the data described in section 3.1. These numbers will provide some evidence of how the effects under analysis may influence $A_{d,t,s,x,y,t}$. The first statistics we report in Table 3 are the totals summed over the eight years period of study. Notice that the values for the expected inceptions included in Tables 3, 4 and 5 were calculated following the methodology presented in section 4.1.

TABLE 3
TOTALS FOR THE EIGHT YEARS PERIOD OF STUDY

<i>Concept</i>	<i>Total</i>
Years exposed to risk	2,396.959
Actual inceptions	44,342
Musculoskeletal diseases	15%
Mental disorders	12%
Infectious diseases	28%
All other diseases	32%
Accidents and injuries	13%
Expected inceptions	42,001
A/E%	105.5%

It is evident from the statistics in Table 3 that musculoskeletal diseases and mental disorders are very important causes of disability. They represent 27% of the total actual inceptions despite the fact that they correspond to only three causes of disability in the ICD8. Accidents and injuries are also an important source of claims with 13% of the total for five causes in the ICD8. (See Table 1)

More detailed statistics are shown in Tables 4 and 5. From these we can observe that:

- Males contribute around 90% of the total data in all deferred periods and policy durations although female participation has increased in recent years. (See C.M.I.B. (1996) p. 166-169)
- Deferred period one week dominates the claim inceptions (with about 70% of the total) followed by deferred period four weeks (with 14%), even though the exposed to risk in these cases are lower than for the longer deferred periods.
- The lack of inceptions in the deferred periods 13, 26 and 52 weeks (with 8.5%, 6.4% and 2.2% of the total inceptions, respectively) may affect any modelling process, as happened in the studies developed by the C.M.I.B. (1991) and Gutiérrez-Delgado (1999) where the deferred period 52 weeks was excluded from the analysis due to its distorting effect on the results.

TABLE 4
EXPOSED TO RISK AND CLAIM INCEPTIONS BY DEFERRED PERIOD,
POLICY DURATION AND CAUSE OF DISABILITY. MALES

<i>Policy Duration</i>	<i>Exposed in Years</i>	<i>Claim Inceptions</i>			<i>Exposed in Years</i>	<i>Claim Inceptions</i>		
		<i>Actual</i>	<i>Expected</i>	<i>A E%</i>		<i>Actual</i>	<i>Expected</i>	<i>A E%</i>
		Deferred period 1 Week			Deferred period 4 Weeks			
0 Years	1,516	243	189	128.7	9,044	147	100	146.9
Musculoskeletal		26				24		
Mental Disorders		11				23		
Infectious Diseases		120				15		
All Other Diseases		43				34		
Accidents and Injuries		43				51		
1 Year	3,853	601	484	124.2	22,360	307	266	115.6
Musculoskeletal		78				51		
Mental Disorders		40				33		
Infectious Diseases		264				36		
All Other Diseases		92				84		
Accidents and Injuries		127				103		
2+ Years	184,821	26,737	26,161	102.2	255,619	4,540	6,018	75.4
Musculoskeletal		4,425				703		
Mental Disorders		2,352				664		
Infectious Diseases		10,172				293		
All Other Diseases		6,503				2,141		
Accidents and Injuries		3,285				739		
		Deferred period 13 Weeks			Deferred period 26 Weeks			
0 Years	11,886	40	31	128.4	5,028	17	4	423.6
Musculoskeletal		8				3		
Mental Disorders		8				8		
Infectious Diseases		1				0		
All Other Diseases		14				4		
Accidents and Injuries		9				2		
1 Year	45,699	122	132	92.7	36,952	68	32	212.5
Musculoskeletal		11				8		
Mental Disorders		18				26		
Infectious Diseases		5				6		
All Other Diseases		54				20		
Accidents and Injuries		34				8		
2+ Years	566,336	3,073	3,067	100.2	657,059	2,251	1,600	140.7
Musculoskeletal		382				195		
Mental Disorders		600				609		
Infectious Diseases		131				92		
All Other Diseases		1,595				1,234		
Accidents and Injuries		365				121		

TABLE 4 (Continued)

<i>Policy Duration</i>	<i>Exposed in Years</i>	<i>Claim Inceptions</i>			<i>Exposed in Years</i>	<i>Claim Inceptions</i>		
		<i>Actual</i>	<i>Expected</i>	<i>A/E%</i>		<i>Actual</i>	<i>Expected</i>	<i>A/E%</i>
Deferred period 52 Weeks								
1 Year	11,311	12	5	234.4				
Musculoskeletal		2						
Mental Disorders		4						
Infectious Diseases		0						
All Other Diseases		6						
Accidents and Injuries		0						
2+ Years	279,292	794	425	186.9				
Musculoskeletal		87						
Mental Disorders		249						
Infectious Diseases		33						
All Other Diseases		390						
Accidents and Injuries		35						

TABLE 5

EXPOSED TO RISK AND CLAIM INCEPTIONS BY DEFERRED PERIOD,
POLICY DURATION AND CAUSE OF DISABILITY. FEMALES

<i>Policy Duration</i>	<i>Exposed in Years</i>	<i>Claim Inceptions</i>			<i>Exposed in Years</i>	<i>Claim Inceptions</i>		
		<i>Actual</i>	<i>Expected</i>	<i>A/E%</i>		<i>Actual</i>	<i>Expected</i>	<i>A/E%</i>
Deferred period 1 Week					Deferred period 4 Weeks			
0 Years	518	82	63	130.3	4,187	67	38	176.0
Musculoskeletal		10				10		
Mental Disorders		5				8		
Infectious Diseases		42				7		
All Other Diseases		14				26		
Accidents and Injuries		11				16		
1 Year	1,222	166	150	110.9	9,529	172	93	185.8
Musculoskeletal		8				20		
Mental Disorders		8				30		
Infectious Diseases		73				24		
All Other Diseases		39				58		
Accidents and Injuries		38				40		
2+ Years	14,825	2,607	1,969	132.4	46,105	1,011	695	145.5
Musculoskeletal		324				144		
Mental Disorders		188				165		
Infectious Diseases		1,073				85		
All Other Diseases		723				473		
Accidents and Injuries		299				144		

TABLE 5 (Continued)

Policy Duration	Exposed in Years	Claim Inceptions			Exposed in Years	Claim Inceptions		
		Actual	Expected	A/E%		Actual	Expected	A/E%
		Deferred period 13 Weeks			Deferred period 26 Weeks			
0 Years	3,567	21	8	269.0	2,082	3	1	247.9
Musculoskeletal		4				0		
Mental Disorders		5				3		
Infectious Diseases		1				0		
All Other Diseases		8				0		
Accidents and Injuries		3				0		
1 Year	12,418	59	30	198.3	13,962	33	9	380.4
Musculoskeletal		6				4		
Mental Disorders		16				12		
Infectious Diseases		4				1		
All Other Diseases		23				14		
Accidents and Injuries		10				2		
2+ Years	71,084	487	270	180.6	82,253	501	124	403.8
Musculoskeletal		70				61		
Mental Disorders		100				156		
Infectious Diseases		30				15		
All Other Diseases		247				250		
Accidents and Injuries		40				19		
		Deferred period 52 Weeks						
1 Year	4,748	6	2	347.8				
Musculoskeletal		2						
Mental Disorders		3						
Infectious Diseases		0						
All Other Diseases		1						
Accidents and Injuries		0						
2+ Years	39,684	161	33	483.8				
Musculoskeletal		20						
Mental Disorders		51						
Infectious Diseases		8						
All Other Diseases		73						
Accidents and Injuries		9						

4. THE MODELLING PROCESS

4.1. Generalized Linear Models

In order to model the effect of cause of disability and initial selection on claim inceptions, we fitted generalized linear models (GLM) to our data. (See Dobson (1990) as well as McCullagh and Nelder (1983) for an extended background.)

The model follows the basic form:

$$A_{d,t,s,x,y,i} \sim \text{Poisson}(\mu_{d,t,s,x,y,i})$$

with

$$\mu_{d,t,s,x,y,i} = E_{d,t,s,x,y} f_{d,t,s,x,y,i}$$

where $E_{d,t,s,x,y}$ is the expected number of claim inceptions for all causes for each combination of factors (d, t, s, x, y) . $E_{d,t,s,x,y}$ is calculated using the standard basis described in C.M.I.B. (1991). This basis was created using data from males 1975-1978. Therefore an adjustment is required since the basis in C.M.I.B. (1991) is almost certainly not appropriated for the data we analysed. In addition the standard basis does not allow us to evaluate the influence of the initial selection and cause of disability in our data.

In consequence the calculation of $f_{d,t,s,x,y,i}$ is our main objective in this stage because through it we can analyse the influence of the initial selection (t) and cause of disability (i) effects. We fulfil our aim by developing the modelling process described below.

Using logarithms, we obtain

$$\log(\mu_{d,t,s,x,y,i}) = \log(E_{d,t,s,x,y}) + \log(f_{d,t,s,x,y,i}) \quad (1)$$

with $\log(f_{d,t,s,x,y,i})$ being modelled through a linear expression involving the factors and possibly interactions.

The modelling process assumes an over-dispersed Poisson distribution for $A_{d,t,s,x,y,i}$ in this case. The over-dispersion derives mainly from duplicate policies among the data. Therefore, variance inflation factors¹ (V_d) provided by the C.M.I.B. were used to calculate the weights ($1/V_d$). These are incorporated in the GLM process by putting more or less weight on the observations from the different deferred periods, similar to the basic idea in weighted least squares. (See Venables and Ripley (1994) for further details of the mathematical model.) Values of V_d are: 3.890 for deferred period (DP) 1 week, 1.320 for DP 4 weeks, 1.210 for DP 13 weeks, 1.244 for DP 26 weeks and 1.0006 for DP 52 weeks. We also insert additional rows with zeros in the $A_{d,t,s,x,y,i}$ to complete the sets of five groups of causes in cases where there is a non-zero value for the overall $E_{d,t,s,x,y}$ but not for all $A_{d,t,s,x,y,i}$. The inclusion of these rows fulfils statistical conditions

¹ Factors by which the variance of the actual number of claim inceptions exceeds the mean due to the presence of duplicated policies. The mathematical procedure to obtain these factors is provided in C.M.I.B. (1991) [Part A: Appendix A.]

of consistency for the modelling process. Not doing this would ignore cases where there happens to be no actual inceptions when there are expected inceptions. A consequence of this inclusion is a higher over-dispersion than allowed for by the variance inflation factors.

The modelling was performed using the statistical package S-plus with the stepwise selection process (see Chambers and Hastie (1992) and Venables and Ripley (1994) for further details). We incorporate in the seed model all the effects and define the offset $E_{d,t,s,x,y}$ for inclusion in the model. In addition we include an interaction between policy duration and group of causes of disability because it is of primary interest. The parametrisation that considers the sum of the parameters for a factor or interaction to be equal to zero is used in order to facilitate the analysis of results.

The main criterion to assess the goodness of fit for the models in our work is the value of the residual deviance (RD), referred to a χ^2 distribution on the corresponding degrees of freedom (DF). Large RD, relative to the appropriate χ^2 distribution, implies a poor fit.

The addition of terms to the model is assessed by using the difference in RD, referred to a χ^2 distribution on the corresponding difference in DF. A significant difference implies that the added terms do improve the fit of the model. However if we have a model which is already complex enough and gives a good fit according to the main criterion, we may compromise and choose not to add the terms. This is associated with the concept of parsimony.

In addition we use the value of the residual mean square (RSM) to measure the incorporation of over-dispersion. Under the Poisson distribution the RMS should be close to the ideal of 1. For this work we were satisfied with values less than 1.20.

The first results from the stepwise process showed a big distortion caused by the scarcity of data for $A_{d,t,s,x,y,t}$ for deferred period 52 weeks, as happened during the previous stage of the research. Therefore we decided to exclude this category from the analysis. Excluding the data from deferred period 52 weeks resulted in a decrease of only 974 (2.25%) in the number of claim inceptions. In consequence the loss was not significant in terms of the quantity of data while it improved to a high degree the goodness of fit.

Two models were considered at the fitting stage. The first includes all the effect mentioned in section 3.1 and several interactions. The year effect was included both as a factor and as covariate (linear and quadratic). Although technically significant, year and the interaction of year with cause of disability, were included at a late stage of the modelling procedure. The inclusion of year in the model makes the analysis of the cause of disability a very complex work that would involve the analysis of a six-way table.

Another approach we used to explain the year effect was to replace year by an economic index. We used both unemployment rates (separately for males and females) as well as the retail price index (to June and to December, including a lag of a year). However these did not provide any improvement over using year itself.

In addition eight calendar years are not enough to develop a time trend analysis. Nevertheless we are conscious that the year effect is significant for a more specific study as mentioned in Korabinski and Waters (1998), Gutiérrez-Delgado (1999) and Renshaw and Haberman (1999).

The second, and chosen, model excludes the year effect. Although results from this imply a slightly worse goodness of fit than the former model, the reduced complexity of the analysis makes this second option preferable.

The model selected from the stepwise process fits very well although it does not completely incorporate the over-dispersion effect (one of our criteria for evaluating the goodness of fit), at least to the levels reported in previous analysis (see Gutiérrez-Delgado (1999)). The reason for the presence of over-dispersion is the values of the variance inflation factors (V_d) which are only an approximation to the real over-dispersion from duplicated policies and thus they may underestimate it. In addition the rows with $A_{d,t,s,x,y,i} = 0$ we included to complete the set of five groups of causes may contribute to increase the phenomenon. Nevertheless this issue is of secondary importance to our main purpose. In consequence we decided to continue our analysis with the model selected. The model produced a residual deviance (RD) of 5,395 on 6,462 degrees of freedom (DF) and a residual mean square (RMS) of 1.259. Details of this model are provided in section 4.2.

4.2. The Model Fitted

The model fitted has the following components:

Main effects:

- Deferred period (d): factor with 4 levels
- Policy duration (t): factor with 3 levels
- Sex (s): factor with 2 levels
- Group of age (x): factor with 7 levels
- Group of cause of disability (i): factor with 5 levels

Interactions:

- d by i : 20 levels
- x by i : 35 levels
- s by i : 10 levels
- t by i : 15 levels
- d by t : 12 levels
- s by x : 14 levels
- d by x : 28 levels
- d by s : 8 levels

Using the notation of generalized linear model theory we can express our model as:

$$\log \mu_{d,t,s,x,y} = \log E_{d,t,s,x,y} + \phi_0 + \alpha_d + \beta_t + \gamma_s + \tau_x + \eta_i + (\alpha\eta)_{d,i} + (\tau\eta)_{x,i} + (\gamma\eta)_{s,i} + (\beta\eta)_{t,i} + (\alpha\beta)_{d,t} + (\gamma\tau)_{s,x} + (\alpha\tau)_{d,x} + (\alpha\gamma)_{d,s} \quad (2)$$

where the terms of the linear model are as follows:

ϕ_0	constant term
α_d	deferred period (1 = 1 week, ..., 4 = 26 weeks)
β_t	policy duration (0 = 0 years, ..., 2 = 2 or more years)
γ_s	sex (1 = males, 2 = females)
τ_x	age (1 = 18 to 29, 2 = 30 to 34, ..., 7 = 55 to 66)
η_i	cause of disability (See Table 2)
$(\alpha\eta)_{d,i}$	deferred period by cause of disability
$(\tau\eta)_{x,i}$	group of ages by cause of disability
$(\gamma\eta)_{s,i}$	sex by cause of disability
$(\beta\eta)_{t,i}$	policy duration by cause of disability
$(\alpha\beta)_{d,t}$	deferred period by policy duration
$(\gamma\tau)_{s,x}$	sex by group of ages
$(\alpha\tau)_{d,x}$	deferred period by group of ages
$(\alpha\gamma)_{d,s}$	deferred period by sex

The presence of the interaction terms in the model implies that the description of the effects on the response $A_{d,t,s,x,y,i}$ is complicated in that we can not summarise in a simple way neither the effect of initial selection nor cause of disability. We can only do this by presenting appropriate multi-way tables and graphs of the different relations. We calculate a set of factors to measure the influence of the effects of interest. The factors are calculated from the series of coefficients related to policy duration and cause of disability that are included in the fitted model described in equation (2). The factors are defined in equations (3) and (4) below.

Initial selection factor:

$$F_{i,d,i} = \exp \left[\beta_t + (\beta\eta)_{t,i} + (\alpha\beta)_{d,t} \right] \quad (3)$$

Cause of disability factor:

$$F_{i,d,t,s,x} = \exp \left[\eta_i + (\alpha\eta)_{d,i} + (\tau\eta)_{x,i} + (\gamma\eta)_{s,i} + (\beta\eta)_{t,i} \right] \quad (4)$$

Note that, for example, for the initial selection factor:

$$F_{2;1,1} = \exp \left[-(\beta_0 + \beta_1) - \left[(\beta\eta)_{0,1} + (\beta\eta)_{1,1} \right] - \left[(\alpha\beta)_{1,0} + (\alpha\beta)_{1,1} \right] \right]$$

which exploits facts such as $\beta_0 + \beta_1 + \beta_2 = 0$ using our parametrisation.

A more extended explanation about the influence of the several effects as well as of both factors is presented in section 5.

5. RESULTS

The estimated coefficients from the fitted GLM for our model defined in equation (2) are given in Table 6.

TABLE 6
COEFFICIENTS FROM THE FITTED MODEL

<i>Term</i>	<i>Value</i>	<i>s.e.</i>	<i>t value</i>	<i>Term</i>	<i>Value</i>	<i>s.e.</i>	<i>t value</i>
α_1	-0.164	0.045	-3.631	$(\pi\eta)_{1,1}$	-0.158	0.063	-2.518
α_2	-0.150	0.044	-3.398	$(\pi\eta)_{1,2}$	-0.252	0.081	-3.129
α_3	-0.179	0.059	-3.048	$(\pi\eta)_{1,3}$	0.279	0.050	5.549
β_0	0.188	0.063	2.994	$(\pi\eta)_{1,4}$	-0.570	0.055	-10.405
β_1	-0.017	0.043	-0.405	$(\pi\eta)_{2,1}$	-0.050	0.044	-1.158
γ_1	-0.240	0.013	-18.303	$(\pi\eta)_{2,2}$	-0.078	0.059	-1.318
η_1	-0.155	0.053	-2.897	$(\pi\eta)_{2,3}$	0.186	0.034	5.402
η_2	-0.161	0.062	-2.617	$(\pi\eta)_{2,4}$	-0.286	0.039	-7.430
η_3	-0.420	0.053	-7.901	$(\pi\eta)_{3,1}$	0.127	0.034	3.713
η_4	0.960	0.043	22.578	$(\pi\eta)_{3,2}$	-0.060	0.049	-1.230
τ_1	0.102	0.051	2.012	$(\pi\eta)_{3,3}$	0.199	0.029	6.951
τ_2	0.196	0.040	4.896	$(\pi\eta)_{3,4}$	-0.348	0.033	-10.706
τ_3	0.121	0.033	3.656	$(\pi\eta)_{4,1}$	0.126	0.031	4.128
τ_4	0.011	0.028	0.373	$(\pi\eta)_{4,2}$	0.005	0.042	0.111
τ_5	0.019	0.027	0.716	$(\pi\eta)_{4,3}$	0.093	0.026	3.577
τ_6	-0.038	0.028	-1.353	$(\pi\eta)_{4,4}$	-0.117	0.027	-4.312
ϕ_0	-1.378	0.035	-38.873	$(\pi\eta)_{5,1}$	-0.049	0.031	-1.565
$(\beta\eta)_{0,1}$	0.063	0.086	0.732	$(\pi\eta)_{5,2}$	0.173	0.039	4.446
$(\beta\eta)_{1,1}$	-0.129	0.068	-1.899	$(\pi\eta)_{5,3}$	-0.228	0.027	-8.384
$(\beta\eta)_{0,2}$	-0.048	0.106	-0.455	$(\pi\eta)_{5,4}$	0.197	0.025	7.920
$(\beta\eta)_{1,2}$	0.052	0.079	0.665	$(\pi\eta)_{6,1}$	0.076	0.033	2.292
$(\beta\eta)_{0,3}$	0.055	0.076	0.720	$(\pi\eta)_{6,2}$	0.157	0.042	3.778
$(\beta\eta)_{1,3}$	-0.033	0.059	-0.558	$(\pi\eta)_{6,3}$	-0.190	0.030	-6.358
$(\beta\eta)_{0,4}$	-0.058	0.072	-0.806	$(\pi\eta)_{6,4}$	0.370	0.026	14.249
$(\beta\eta)_{1,4}$	0.016	0.054	0.287	$(\alpha\tau)_{1,1}$	-0.657	0.059	-11.184
$(\alpha\eta)_{1,1}$	0.068	0.026	2.638	$(\alpha\tau)_{2,1}$	0.021	0.060	0.355
$(\alpha\eta)_{2,1}$	0.081	0.034	2.363	$(\alpha\tau)_{3,1}$	0.274	0.089	3.098
$(\alpha\eta)_{3,1}$	0.020	0.042	0.465	$(\alpha\tau)_{1,2}$	-0.140	0.043	-3.223
$(\alpha\eta)_{1,2}$	-0.900	0.028	-32.065	$(\alpha\tau)_{2,2}$	0.035	0.051	0.684
$(\alpha\eta)_{2,2}$	-0.089	0.035	-2.507	$(\alpha\tau)_{3,2}$	-0.043	0.067	-0.639
$(\alpha\eta)_{3,2}$	0.272	0.039	6.892	$(\alpha\tau)_{1,3}$	0.003	0.035	0.093
$(\alpha\eta)_{1,3}$	1.389	0.034	40.660	$(\alpha\tau)_{2,3}$	0.009	0.043	0.214

TABLE 6 (Continued)

<i>Term</i>	<i>Value</i>	<i>s.e.</i>	<i>t value</i>	<i>Term</i>	<i>Value</i>	<i>s.e.</i>	<i>t value</i>
$(\alpha\eta)_{2,3}$	-0.310	0.048	-6.457	$(\alpha\tau)_{3,3}$	-0.003	0.053	-0.059
$(\alpha\eta)_{3,3}$	-0.598	0.064	-9.313	$(\alpha\tau)_{1,4}$	0.147	0.029	5.064
$(\alpha\eta)_{1,4}$	-0.537	0.019	-28.444	$(\alpha\tau)_{2,4}$	0.001	0.037	0.018
$(\alpha\eta)_{2,4}$	0.016	0.025	0.631	$(\alpha\tau)_{3,4}$	-0.077	0.045	-1.711
$(\alpha\eta)_{3,4}$	0.185	0.029	6.274	$(\alpha\tau)_{1,5}$	0.210	0.027	7.917
$(\gamma\tau)_{1,1}$	0.123	0.027	4.560	$(\alpha\tau)_{2,5}$	0.038	0.034	1.122
$(\gamma\tau)_{1,2}$	0.062	0.025	2.437	$(\alpha\tau)_{3,5}$	-0.125	0.040	-3.092
$(\gamma\tau)_{1,3}$	0.007	0.022	0.308	$(\alpha\tau)_{1,6}$	0.187	0.027	6.970
$(\gamma\tau)_{1,4}$	-0.047	0.021	-2.211	$(\alpha\tau)_{2,6}$	-0.027	0.035	-0.763
$(\gamma\tau)_{1,5}$	-0.083	0.021	-3.959	$(\alpha\tau)_{3,6}$	-0.017	0.040	-0.431
$(\gamma\tau)_{1,6}$	-0.046	0.023	-1.977	$(\gamma\eta)_{1,1}$	0.056	0.022	2.518
$(\alpha\beta)_{1,0}$	-0.053	0.079	-0.671	$(\gamma\eta)_{1,2}$	-0.041	0.026	-1.572
$(\alpha\beta)_{2,0}$	0.012	0.076	0.161	$(\gamma\eta)_{1,3}$	-0.030	0.019	-1.569
$(\alpha\beta)_{3,0}$	0.012	0.101	0.122	$(\gamma\eta)_{1,4}$	-0.129	0.017	-7.735
$(\alpha\beta)_{1,1}$	0.106	0.056	1.891	$(\alpha\gamma)_{1,1}$	0.134	0.016	8.200
$(\alpha\beta)_{2,1}$	0.065	0.054	1.191	$(\alpha\gamma)_{2,1}$	-0.006	0.019	-0.325
$(\alpha\beta)_{3,1}$	-0.155	0.073	-2.110	$(\alpha\gamma)_{3,1}$	0.033	0.023	1.420

These results confirm that all main effects and interactions are significant. Note that the t values should be considered in groups rather than individually. For instance the interaction $(\alpha\gamma)_{d,s}$, shows a t value for the second coefficient of -0.325 which individually is not significant. However the group of three coefficients includes one which is 8.20. This value makes the interaction significant.

It should be noticed here that our main purpose is to describe the extent of the effects of interest. Therefore the fitting provided by the GLM should be taken with reservations if an extrapolation is desired because projections are a secondary purpose of the GLM. In addition predictions using a GLM can provide unreliable values when they are used under lack of data.

The factors ($F_{r,d,i}$ and $F_{i,d,t,s,x}$) described in section 4.2 were calculated together with their standard errors. The resulting values for $F_{r,d,i}$ are given later in Table 7. Results of $F_{i,d,t,s,x}$ for males in group of ages 40 to 44 are presented in Table 8 as an example of the outcomes. The full set of results for $F_{i,d,t,s,x}$ is omitted but can be requested from the authors. We also illustrate graphically in Figure 2 results for $F_{r,d,i}$ for deferred period 4 weeks being the case where the initial selection effect is strongest. The graphs include upper and lower two-standard-error limits that are approximate 95% confidence limits and also the horizontal line $F_{r,d,i} = 1$ which represents the stage where the effects analysed have no influence.

The results from the initial selection factor should be interpreted in the following way:

- If there is no influence due to the initial selection effect, then for the combination of t, d, i in $F_{t,d,i}$ we have that $F_{t,d,i} = \exp[\beta_t + (\eta\beta)_{i,t} + (\alpha\beta)_{d,t}] = \exp[0] = 1$
 - If there is an influence due to the initial selection effect, then $F_{t,d,i} = \exp[\beta_t + (\eta\beta)_{i,t} + (\alpha\beta)_{d,t}] \neq 1$, where $F_{t,d,i} > 1$ implies more inceptions and $F_{t,d,i} < 1$ indicates fewer inceptions.
- The interpretation for the cause of disability factor is similar.

5.1. The Initial Selection Factor

Results for the initial selection factor are presented in Table 7 which is a three-way table. They show a decreasing trend in general except for deferred period 13 weeks where there is a decrease-increase in four of the five groups of causes. The most significant decreasing trends are found for deferred period 4 weeks with infectious diseases being the best example, followed by accidents and injuries and musculoskeletal disorders. There is evidence of a decreasing influence in accidents and injuries over all deferred periods as happens for infectious diseases and all other diseases. Musculoskeletal diseases show a decreasing pattern for deferred periods 1 and 4 weeks. Mental disorders presents this trend in deferred periods 4 and 26 weeks. (See Table 7 and Figure 2.) In consequence although the large standard errors are such that individual cases are not necessarily significant, a negative pattern is prevalent, suggesting the same trend found in the previous stage of analysis.

Notice that the standard errors for both factors are generally higher for policy duration zero years than for the other durations. This is because the amount of data for the first category of policy duration is considerably smaller than for the other categories as we can confirm from Tables 4 and 5. Nevertheless there are cases where the standard errors increase with increasing policy duration. We have to remember for these situations that the standard errors increase/decrease with the size of the factor due to the underlying Poisson distribution. This fact opposes and can surpass the influence of the increase in the amount of data available generating a bigger standard error than expected. (See Tables 7 and 8.)

TABLE 7

INITIAL SELECTION FACTOR BY DEFERRED PERIOD AND GROUP OF CAUSE OF DISABILITY

<i>Policy Duration</i>	<i>F_{t,d,i}</i>	<i>s.e.</i>	<i>Policy Duration</i>	<i>F_{t,d,i}</i>	<i>s.e.</i>
<i>DP 1 Week</i>			<i>DP 4 Weeks</i>		
Musculoskeletal Diseases					
0 years	1.219	0.365	0 years	1.301	0.398
1 year	0.961	0.271	1 year	0.922	0.223
2+ years	0.854	0.240	2+ years	0.834	0.260
Mental Disorders					
0 years	1.091	0.271	0 years	1.164	0.294
1 year	1.151	0.179	1 year	1.105	0.245
2+ years	0.796	0.200	2+ years	0.777	0.197
Infectious Diseases					
0 years	1.209	0.472	0 years	1.290	0.481
1 year	1.058	0.248	1 year	1.015	0.260
2+ years	0.834	0.271	2+ years	0.763	0.253
All Other Diseases					
0 years	1.080	0.254	0 years	1.153	0.302
1 year	1.111	0.220	1 year	1.066	0.193
2+ years	0.834	0.206	2+ years	0.814	0.206
Accidents and Injuries					
0 years	1.131	0.249	0 years	1.207	0.291
1 year	1.201	0.247	1 year	1.153	0.243
2+ years	0.736	0.123	2+ years	0.719	0.119
<i>DP 13 Weeks</i>			<i>DP 26 Weeks</i>		
Musculoskeletal Diseases					
0 years	1.301	0.535	0 years	1.323	0.840
1 year	0.740	0.232	1 year	0.850	0.342
2+ years	1.039	0.360	2+ years	0.889	0.391
Mental Disorders					
0 years	1.164	0.422	0 years	1.184	0.689
1 year	0.887	0.265	1 year	1.019	0.404
2+ years	0.969	0.279	2+ years	0.829	0.316
Infectious Diseases					
0 years	1.290	0.576	0 year	1.313	0.836
1 year	0.815	0.270	1 year	0.936	0.394
2+ years	0.951	0.327	2+ years	0.814	0.340
All Other Diseases					
0 years	1.153	0.438	0 year	1.172	0.710
1 year	0.856	0.221	1 year	0.983	0.356
2+ years	1.014	0.295	2+ years	0.868	0.331
Accidents and Injuries					
0 years	1.207	0.438	0 years	1.228	0.728
1 year	0.925	0.262	1 year	1.063	0.410
2+ years	0.896	0.203	2+ years	0.766	0.266

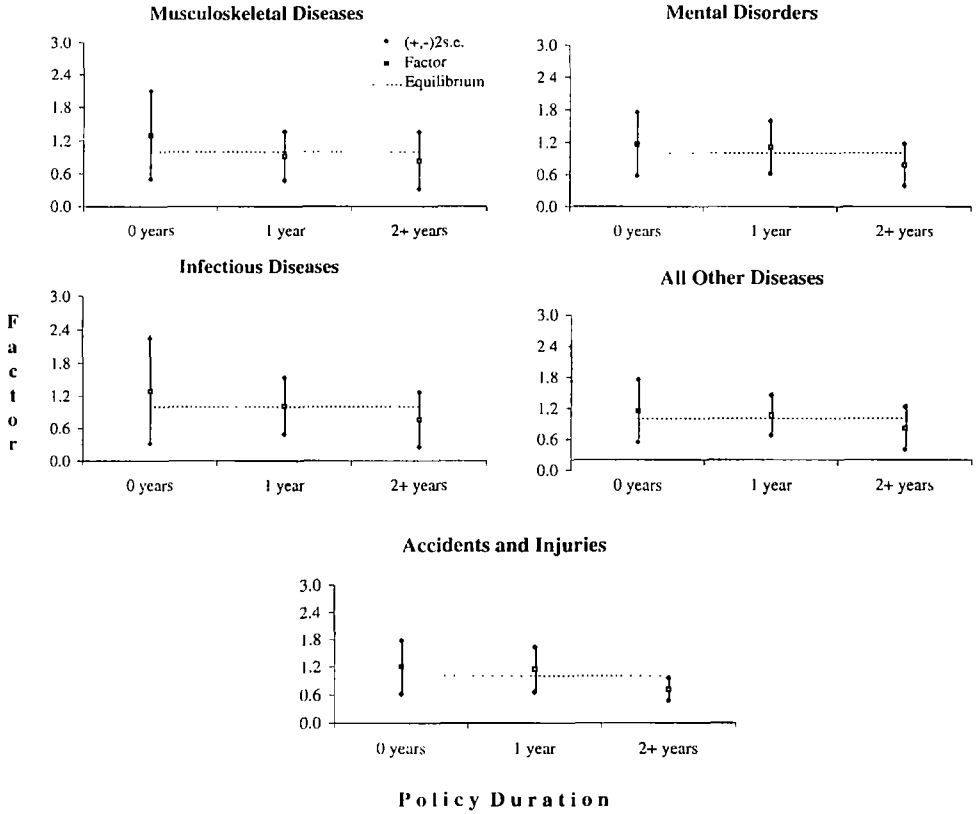


FIGURE 2: Initial Selection Factor.

5.2. The Cause of Disability Factor

Our analysis is based on the full set of results from $F_{i,d,t,s,x}$ although we just present values for males in group of ages 40 to 44 in Table 8 which including all the results form a five-way table. The results focus on the changes within the groups of causes of disability. The main points to observe are:

- Values of $F_{i,d,t,s,x}$ are in general slightly higher for males than for females denoting that the influence of the cause of disability is stronger among the male population
- Values of $F_{i,d,t,s,x}$ are very high for infectious diseases in deferred period 1 week as well as for all other diseases in deferred period 26 weeks. The situation could be a consequence of the characteristics of the kind of diseases involved. (See section 2)
- Results from factors for musculoskeletal diseases (back pains) are in general below but close to one. Nevertheless there are values above one for males, ages 35 to 44 as well as ages 50 to 54 in deferred periods 1, 4 and 13 weeks.

- Mental disorders factors increase their significance with deferred period and age. Therefore the biggest values are found in deferred period 26 weeks for both sexes.
- Results for infectious diseases are very high in deferred period 1 week as expected. In contrast the values are quite low in the other deferred periods.
- All other diseases increase their importance with deferred period and age as anticipated. Results may show the influence from the chronic-degenerative diseases in the group.
- Results from accidents and injuries are high for young ages in most deferred periods except for females deferred period 26 weeks. They decrease with age as predicted.

TABLE 8

CAUSE OF DISABILITY FACTOR BY DEFERRED PERIOD, GROUP OF AGES AND POLICY DURATION.
MALES GROUP OF AGES 40 TO 44

<i>Cause of Disability</i>	<i>Policy Duration</i>					
	<i>0 Years</i>		<i>1 Year</i>		<i>2+ Years</i>	
	<i>Factor</i>	<i>s.e.</i>	<i>Factor</i>	<i>s.e.</i>	<i>Factor</i>	<i>s.e.</i>
	Deferred Period 1 Week					
Musculoskeletal	1.171	0.413	0.967	0.239	1.175	0.103
Mental	0.318	0.223	0.352	0.149	0.333	0.064
Infectious	2.967	2.220	2.716	1.172	2.746	0.415
All Other Dis.	1.126	0.337	1.213	0.239	1.245	0.101
Accidents	0.803	0.260	0.893	0.195	0.749	0.078
	Deferred period 4 Weeks					
Musculoskeletal	1.186	0.425	0.979	0.250	1.190	0.154
Mental	0.716	0.333	0.791	0.235	0.748	0.122
Infectious	0.542	0.209	0.496	0.146	0.502	0.092
All Other Dis.	1.958	0.781	2.108	0.600	2.164	0.332
Accidents	1.108	0.358	1.232	0.286	1.033	0.140
	Deferred Period 13 Weeks					
Musculoskeletal	1.116	0.414	0.921	0.253	1.120	0.177
Mental	1.026	0.464	1.135	0.343	1.074	0.184
Infectious	0.407	0.205	0.372	0.157	0.376	0.119
All Other Dis.	2.319	1.151	2.497	0.933	2.563	0.573
Accidents-Inj.	0.925	0.317	1.028	0.260	0.862	0.150
	Deferred Period 26 Weeks					
Musculoskeletal	0.924	0.360	0.763	0.235	0.927	0.182
Mental	1.603	0.838	1.772	0.662	1.675	0.355
Infectious	0.457	0.231	0.419	0.180	0.423	0.145
All Other Dis.	2.697	1.725	2.904	1.468	2.980	0.986
Accidents-Inj.	0.548	0.244	0.609	0.206	0.511	0.149

6. SUMMARY AND CONCLUSIONS

The initial selection and the cause of disability effects proved to be significant in the modelling process of the actual claim inceptions. The extent of the influence from the former, measured through $F_{t,d,i}$ appeared to be negative in several deferred periods and causes of disability. The strongest example of the trend in this factor was found in deferred period 4 weeks. The impact from the latter factor, quantified by $F_{i,d,t,s,x}$, showed that musculoskeletal diseases were significant only in very specific groups of ages. Mental disorders increased their importance by deferred period and age, as happened also for the group of All other diseases. Infectious diseases were highly significant only in deferred period 1 week. Accidents and injuries were quite important exclusively in young ages.

Results from the factors showed that the grouping of causes of disability helped to partially explain the decreasing trend found in Gutiérrez-Delgado (1999) as a consequence of some moral hazard. We concluded that there might be some other effects, not available for this research, contributing to the negative trend on actual claim inceptions.

The model selected for the analysis provided a very good fit although it did not incorporate the over-dispersion at the level obtained in previous analysis.

ACKNOWLEDGEMENTS

The authors are grateful to the C.M.I.B. for supplying the data, as the basis of this analysis; to the Faculty of Sciences, Universidad Nacional Autónoma de México and the Department of Actuarial Mathematics and Statistics, Heriot-Watt University for their financial support as well as to Prof. Howard R. Waters, Mr. Hugh James, Ms. Karin Lloyd, and Mr. Tony Bartholomew for their invaluable comments and suggestions. All findings and observations are exclusive responsibility of the authors.

REFERENCES

- A.D.C. (1998) *1997 Report of the Disability Committee*. Australian Disability Committee.
- CHAMBERS, J. and HASTIE, T. (1992) *Statistical Models in S*. Wadsworth & Brooks, New York.
- C.M.I.B. (1984) *CMI Report No. 7*. The Faculty of Actuaries and the Institute of Actuaries, United Kingdom.
- C.M.I.B. (1991) *CMI Report No. 12*. The Faculty of Actuaries and the Institute of Actuaries, United Kingdom.
- C.M.I.B. (1996) *CMI Report No. 15*. The Faculty of Actuaries and the Institute of Actuaries, United Kingdom.
- DOBSON, A. (1990) *An Introduction to Generalized Linear Models*. Chapman and Hall, New York.
- GUTIÉRREZ-DELGADO, C. (1999) Initial selection for Permanent Health Insurance. In *Insurance: Mathematics and Economics*, **25** (5).
- KLUWER (1998) *Permanent Health Insurance 1997/98. The Unique Annual Review*. Croner Publications and Kluwer Publishing, United Kingdom.

- KORABINSKI, A.A. and WATERS, H.R. (1998) PHI claim inception rates: An analysis of individual company experience in the United Kingdom. In *Transactions of the 26th International Congress of Actuaries*, 6, 188-216.
- LUNDBERG, O. (1969) Methods of studying the risk process in disability insurance. *ASTIN Bulletin*, 5, 267-273.
- MCCULLAGH, P. and NELDER, J. (1983) *Generalized Linear Models*. Chapman and Hall, New York.
- MCDONALD, C. (1995) *Epidemiology of Work Related Diseases*. BMJ Publishing Group, United Kingdom.
- MILLER, J. and COURANT, S. (1974) A mathematical model of the incidence of disability. In *Transactions of the Society of Actuaries*, XXVI: 1-41.
- RENSHAW, D. and HABERMAN, S. (1999) Recent time trends in UK Permanent Health Insurance recovery, mortality and claim inception transition intensifies. In *Transactions of the Third International Congress Insurance: Mathematics and Economics*, 7.
- SOULE, C. (1994) *Disability Income Insurance. The Unique Risk*. Business One Irwin, New York.
- VENABLES, W. and RIPLEY, B. (1994) *Modern Applied Statistics with S-plus*. Springer, New York.

MARÍA CRISTINA GUTIÉRREZ-DELGADO AND

ATHOL A. KORABINSKI

Department of Actuarial Mathematics and Statistics

Heriot-Watt University

Edinburgh EH14 4AS

United Kingdom

E-mails: c.gutierrez@ma.hw.ac.uk

a.a.korabinski@ma.hw.ac.uk

